Bundesamt
für Sicherheit in der
Informationstechnik

BioKeyS Pilot-DB Teil 2 (Projekt Template Protection)

Abschlussbericht

Version 1.4.5 vom Mai 2011

Bundesamt für Sicherheit in der Informationstechnik
Postfach 20 03 63
53133 Bonn
Tel.: +49 228 99 9582-0
E-Mail: biometrics@bsi.bund.de
Internet: http://www.bsi.bund.de

Vorwort

Biometrische Verfahren spielen schon heute in hoheitlichen Anwendungen eine große Rolle. Ihre Verbreitung führt dazu, dass jeder Bürger mit der Erfassung und Verarbeitung seiner biometrischen Daten Erfahrung sammeln kann, sei es bei der Beantragung eines elektronischen Reisepasses oder auch bei der Ausstellung eines neuen Personalausweises. Genutzt werden die biometrischen Daten beispielsweise dann, wenn in der automatischen biometrischen Grenzkontrolle (easyPass) am Frankfurter Flughafen durch den biometrischen Vergleich die Bindung von Reisedokument an den Dokumenteninhaber geprüft wird.

Aus der Erfahrung im Umgang mit biometrischen Sensoren wächst unter den Bürgern auch die Bereitschaft und Motivation, ihre biometrischen Daten auch in kommerziellen Systemen zu hinterlegen, um so den größeren Komfort einer bestimmten Dienstleistung nutzen zu können. Das "Bezahlen mit dem Fingerabdruck" sei als Beispiel genannt, das in vielen Einzelhandel-Filialen das Bargeld oder token-basierte Verfahren abgelöst hat.

In all diesen Anwendungen, in denen die biometrischen Daten unserer Bürger erfasst und verarbeitet werden, ist es unerlässlich, den Anforderungen des Datenschutzes zu genügen. Bei der Konzeption und Umsetzung biometrischer Systeme sollte die Grundsatz-Anforderung "Privacy by Design" berücksichtigt werden. Daher wurden in der biometrischen Forschung in den letzten Jahren Schutzmöglichkeiten von biometrischen Referenzdaten (Biometric Template Protection) entwickelt, die die Vorteile der Biometrie mit bewährten Verfahren der Kryptographie effektiv kombinieren, so dass das ursprüngliche biometrische Merkmal in Analogie zum Passwort-Verfahren nicht mehr im Klartext, sondern kryptiert als öffentlicher Referenzdatensatz gespeichert werden kann.

Das Bundesamt für Sicherheit in der Informationstechnik (BSI) widmet sich in der Projekt-Reihe BioKeyS diesem wichtigen Thema. Mit den aktuellen Ergebnissen der biometrischen Template-Protection-Verfahren ist es möglich, biometrische Referenzdatensätze zu erzeugen, aus denen das biometrische Merkmal nicht abgeleitet werden kann. Die Referenz bleibt für die betroffene Person eindeutig innerhalb einer Anwendung, wird jedoch durch eine andere Wahl von Parametern für eine andere Anwendung diversifizierbar. Der vorliegende Bericht dokumentiert den Stand der Technik und die Beiträge zu kryptographisch-biometrische Authentisierungssysteme mittels biometrischer Protection-Verfahren, die durch BioKeyS erstellt wurden.

Bonn, im Mai 2011

Michael Hange, Präsident des BSI

Danksagung

In diesem Bericht wird über die Ergebnisse im Projekt BioKeyS - Pilot-DB Teil 2 berichtet. In diesem Projekt wird eine wissenschaftliche Begleitstudie erstellt zum laufenden Projekt BioKeyS, das zwischen dem Bundesamt für Sicherheit in der Informationstechnik und dem Unternehmen secunet Security Networks AG (secunet) durchgeführt wird. Die Begleitstudie soll Möglichkeiten aufzeigen, die BioKey-Implementierung in einer nachfolgenden Projektphase um neue leistungsfähigere Algorithmen zum Schutz von biometrischen Templates zu erweitern.

Die Projektergebnisse wurden im Auftrag des BSI von der Hochschule Darmstadt als Generalunternehmer in enger Kooperation mit dem Fraunhofer IGD, der Ludwig-Maximilians-Universität München (LMU) sowie der Rheinisch-Westfälische Technische Hochschule Aachen (RWTH) bearbeitet. Ein Dank gilt allen Personen, die zum Erfolg des Projektes beigetragen haben, namentlich Sebastian Abt, Christian Böhm, Christoph Busch, Ines Färber, Sergej Fries, Claudia Nickel, Alexander Nouak, Alexander Opel, Annahita Oswald, Bianca Wackersreuther, Peter Wackersreuther, Thomas Seidl und Xuebing Zhou.

Ein besonderer Dank gilt dem BSI als Auftraggeber für die gute Kooperation, namentlich Frau Dr. Korte und Herr Heinrich Ihmor sowie den beteiligten Mitarbeitern der secunet als Projektpartner, namentlich Herrn Dr. Merkle, Herrn Niesing und Herrn Schwaiger.

Inhaltsverzeichnis

Abbildungsverzeichnis

Tabellenverzeichnis

1 Einleitung

Biometrische Authentisierung wird häufig zur Verbesserung der Identitätsverifikation in Betracht gezogen. Durch die Nutzung biometrischer Verfahren entstehen gleichermaßen neue Herausforderungen an den Schutz der Privatsphäre betroffener Personen, wie an die Sicherheit der Verifikations-Systeme. In biometrischen Systemen gespeicherte Referenzdaten enthalten Informationen, die aus den biometrischen Charakteristika einer Person abgeleitet wurden und damit eindeutige Merkmale dieser Person. Für den Fall dass gespeicherte Merkmale direkt eine biometrische Charakteristik abbilden (z.B. Fingerbilder) besteht ein hohes Risiko. Sind diese Daten erst einmal kompromittiert, können sie nicht so leicht ersetzt oder gelöscht werden wie in herkömmlichen Authentisierungs-Systemen. Des Weiteren ist die Anzahl der biometrischen Charakteristika eines Nutzers begrenzt. Eine Mehrfachnutzung von biometrischen Merkmalen in verschiedenen Anwendungen kann zu sog. Cross-Matching-Problemen führen, wenn Anwendungen miteinander verknüpft werden. Darüber hinaus können Referenzdaten für die Authentisierung irrelevante Informationen enthalten (z.B. ethnische Zugehörigkeit, Krankheiten).

Das Projekt adressiert den Schutz von biometrischen Daten exemplarisch im Kontext von Fingerbilderkennungsverfahren. Dieser Bericht führt mit Kapitel 2 in die Thematik ein, wobei die Motivation und der Schutzbedarf für biometrische Referenzdaten erläutert wird. Es wird die im Europäischen Integrierten Projekt TURBINE [TUR08] entwickelte Referenzarchitektur [Bre08] vorgestellt. Das Projekt wurde im Rahmen des siebten Forschungsrahmenprogramm von der Europäischen Union unterstützt (ICT-2007-216339). Das Kapitel 3 behandelt den Stand der Technik und stellt ausgewählte Template-Protection-Verfahren dar, die für Fingerbilderkennungs-Systeme geeignet sind. Ein wesentlicher Faktor für die biometrische Erkennungsleistung ist die Güte der Feature-Extraktions-Algorithmen. Beim Einsatz von Template-Protection-Verfahren muss mit einem Abfall der Erkennungsleistung gerechnet werden. Daher ist die Betrachtung von Qualität von Feature-Extraktions-Verfahren von besonderer Wichtigkeit und wird in Kapitel 4 untersucht. Zur Steigerung der Sicherheit von Template-Protection-Verfahren ist die Integration von Zusatzinformationen von Interesse. Die diesbezügliche Untersuchung wird in Kapitel 5 beschrieben. Im Kontext einer möglichen Anwendung von zentralen Datenbanken mit biometrischen Referenzdaten wird die Anwendung von Template-Protection-Verfahren in Identifikationssystemen in Kapitel 6 untersucht, wobei ein besonderes Augenmerk auf der schnellen Suche in großen Datenbanken liegt. Abschließend wird in Kapitel 7 die relevante internationale Standardisierung im Bereich Template-Protection betrachtet.

Dieser Bericht orientiert sich in der Verwendung der Fachbegriffe an dem ISO SC37 Harmonized Biometric Vocabulary (Standing Document 2 version 12), das mit einer deutschen Übertragung zu finden ist unter:
http://www.3dface.org/media/vocabulary.html

2 Eine Referenzarchitektur zum Schutz biometrischer Templates

Dieses Kapitel führt in die Thematik von Schutznotwendigkeiten und Schutzmöglichkeiten von biometrischen Referenzdaten ein und erläutert dazu die im EU Integrated Project TURBINE [TUR08] entwickelte Referenzarchitektur für Biometric-Template-Protection, die von Brebaart et al [Bre08] in 2008 publiziert wurde und in wesentlichen Teilen auch in den entsprechenden ISO Standard übernommen wurde, der in Kapitel 7 dieses Dokumentes behandelt wird. Das Kapitel orientiert sich bewusst an der Struktur der ursprünglichen Publikation. Es wird ein Anforderungskatalog an biometrische Verarbeitungstechniken vorgeschlagen, der der Wahrung von Privatsphäre und der Sicherheit von Templates in biometrischen Systemen dient. Abgeleitet von diesen Anforderungen wird eine Referenzarchitektur vorgestellt, die Prozesse und Schnittstellen im Rahmen des Schutzes biometrischer Templates auf einer hohen, technologieneutralen Ebene darstellt.

2.1 Motivation

Die steigende Nachfrage an Verbesserungen der Sicherheit in der Grenzkontrolle und die zunehmende Anzahl elektronischer Transaktionen, die über kabelgebundene und drahtlose Netzwerke getätigt werden, haben einen starken Bedarf an einem zuverlässigeren Identitätsmanagement erweckt. Existierende, besitzbasierte Identifikationsmethodiken, wie zum Beispiel eine ID Karte, ein Token oder ein Schlüssel, oder wissensbasierte Methodiken, wie zum Beispiel eine PIN oder ein Passwort, können vergessen, verloren, verteilt oder gestohlen werden, woraus möglicher Weise ein Identitätsdiebstahl oder Identitätsmissbrauch resultiert.

Dieser Identitätsdiebstahl kann zum Beispiel im Falle von Finanzangelegenheiten zum Missbrauch eines Bankkontos, zur Fälschung von Bankkarten oder Schecks oder zur Verwendung gestohlener Kreditkartennummern führen. Im Gesundheitswesen kann Identitätsdiebstahl zu unerlaubtem Zugriff auf Krankenblätter, unbefugtem Zutritt zu zutrittsbeschränkten Bereichen, unberechtigter Nutzung von Medikamenten oder ärztlicher Behandlungen sowie Krankenversicherungsbetrug führen. Auf staatlicher Ebene kann Identitätsdiebstahl in Fälschung oder Missbrauch von Personaldokumenten resultieren, was ernstzunehmende Konsequenzen nach sich ziehen kann, da staatlich ausgestellte Personaldokumente häufig zur Authentisierung von Identitäten in anderen Bereichen genutzt werden.

Nach Informationen der ID-theft Webseite [tw] fällt in den Vereinigten Staaten von Amerika alle drei Sekunden eine Person einem Identitätsdiebstahl zum Opfer und der durch Identitätsdiebstahl verursachte jährliche in den USA entstehende Gesamtschaden wird auf 53 Milliarden US Dollar geschätzt. Der daraus resultierende Bedarf an verlässlicheren Identitätsverifikationssystemen führt zu einem verstärkten Interesse an biometrischen Verfahren zur Ergänzung traditioneller besitz- und wissensbasierter Methodiken. Biometrie kann hierbei gleichermaßen zu einer höheren Zuverlässigkeit in der Verifikation von Identitätsbehauptungen beitragen, wie zu einem erhöhten Nutzerkomfort führen, da biometrische Charakteristika nur schwer vergessen werden oder verloren gehen können.

2.2 Herausforderungen

2.2.1 Privatsphäre

Die Nutzung von Biometrie zur Identitätsverifikation hat auch Bedenken aufgeworfen. Die enge Verbindung biometrischer Verifikationsmethoden zu physikalischen, anatomischen Eigenschaften der betroffenen Personen ermöglicht die Nutzung biometrischer Messdaten für andere als die angedachten Verwendungszwecke und stellt somit eine Gefährdung der Privatsphäre dar, die sich in die folgenden vier Kategorien unterteilen lässt:

- **Nichtauthorisierte Erfassung**: Erfassung biometrischer Samples ohne das Wissen der betroffenen Person, zum Beispiel durch Verwendung versteckter Kameras.
- **Unnötige Erfassung**: Anwendung biometrischer Methoden ohne oder mit nur wenig zusätzlichem Nutzen im Vergleich zu gewöhnlicher Nutzerverifikation.
- **Nichtauthorisierte Verwendung und Preisgabe**: Nutzung biometrischer Verfahren für andere als die von der betroffenen Person genehmigten Zwecke, wie zum Beispiel für forensische Untersuchungen, unerwünschte Verknüpfung von Datenbanken, Überwachung des Alltags eines Individuums und ähnliches.
- **Schleichende Erweiterung des Verwendungsrahmens**: Erweiterung eines Systems in Bereiche in denen die Verwendung ursprünglich nicht vorgesehen war, wie es zum Beispiel bei der Personenkennziffer geschehen ist.

Ungünstiger Weise können Mechanismen zur Minimierung eines Risikos zu einer Erhöhung eines anderen Risikos führen. Wenn zum Beispiel ein System Mechanismen zur Verbesserung der Überwindungssicherheit beinhaltet, indem Fingerabdruckmuster und Finger-Venen-Muster zur gleichen Zeit herangezogen werden, so erhöht sich das Potenzial einer schleichenden Erweiterung des Verwendungsrahmens, in dem die zusätzlichen, in den Venenbildern enthaltenen, gesundheitsrelevanten Informationen für andere Zwecke verwendet werden.

Die für das Verarbeiten biometrischer Daten zu beachtende ist Direktive 95/46/EC [EE95]. Sie behandelt den Schutz personenbezogener Daten und über die freie Verfügbarkeit derartiger Daten. Die Direktive gibt jedoch keine eindeutige Antwort auf Fragen dieser und anderer die Privatsphäre betreffender Risiken, die durch den Einsatz von Biometrie entstehen. Der Artikel 29 EU Beratungsausschuss für Datenschutz und Privatsphäre hat daher in seinem 2003 veröffentlichten Arbeitspapier über Biometrie [Par03] die Bedeutung von den Schutz der Privatsphäre verbessernden Technologien hervorgehoben, um hierdurch biometrische Systeme zu fördern, die eine dem Schutz der Privatsphäre und dem Schutz der Daten freundliche Architektur aufweisen und übermäßiges Sammeln von Daten und einen ungesetzmäßigen Umgang mit diesen Daten erschweren bzw. verhindern.

2.2.2 Sicherheit

Biometrie wird oft zur Steigerung der Sicherheit einer Anwendung in Form einer Erhöhung der Genauigkeit und Zuverlässigkeit einer Identifikation eingesetzt. Ein möglicher Vorbehalt gegen die

Verwendung biometrischer Verfahren ist, dass die erreichte erhöhte Sicherheit mit einem verminderten Schutz der Privatsphäre einhergehen kann [CS07]. Darüber hinaus ist es möglich, dass die Einbeziehung biometrischer Verfahren auf Grund von im biometrischen Subsystem vorherrschender Schwachstellen in neuen Sicherheitsrisiken resultiert. Nach Jain [Jain08] kann das Sicherheitsrisiko eines biometrischen Verifikationssystems in vier Kategorien unterteilt werden:

- Immanente biometrische Fehler auf Grund von vom biometrischen Verifikationssystem getroffener falscher Entscheidungen, die häufig durch Wahrscheinlichkeitswerte für Falsch-Akzeptanz und/oder Falsch-Rückweisung ausgedrückt werden.
- Angriff auf die Systemverwaltung auf Grund unzulänglicher Verwaltungsrichtlinien.
- Unzulänglich geschützte Infrastruktur, resultierend in Schwachstellen im Zusammenhang mit nicht hinreichend gesicherter Hardware, Software oder Kommunikationskanälen.
- Öffentlichkeit von biometrischen Charakteristika, die die versteckte Gewinnung biometrischer Samples erleichtert und die Erzeugung von Plagiaten oder anderer Mittel zur Beeinflussung des Ergebnisses eines Identitätsverifikationssystems ermöglicht.

2.2.3 Vertrauen

Ein dritter Faktor der für die Akzeptanz biometrischer Verifikationssysteme von Bedeutung ist, ist Vertrauen. Vertrauen unterscheidet sich von objektiven Messwerten wie zum Beispiel Falsch-Akzeptanz-Raten in dem Sinne, dass es ein subjektives Gut ist. Vertrauen ist eine Vorhersage auf eine Aktion und die daraus resultierenden Konsequenzen, basierend auf dem Wissen, das ein Subjekt über eine Anwendung oder eine Technologie verfügt. Beispiele für Vertrauen oder Ermangelung von Vertrauen sind Sorgen um gesundheitliche Auswirkungen, die von biometrischen Messungen (zum Beispiel die regelmäßige Ablichtung der Retina) ausgehen, um Hygieneprobleme (bei der Benutzung von Fingerabdrucksensoren) oder um das Risiko gestohlener Körperteile, die biometrische Charakteristika enthalten (wie zum Beispiel Finger). Darüber hinaus können einige biometrische Modalitäten negative Assoziation hervorrufen, wie zum Beispiel die Assoziation zwischen Fingerabdrücken und Verbrechen.

2.2.4 Risikominderung

Die Beständigkeit biometrischer Charakteristika ist einerseits eine für die Erkennungsleistung erstrebenswerte Eigenschaft hat aber auch bedeutende Auswirkungen auf die eingeschränkten Möglichkeiten der Risikominimierung im Bezug auf Identitätsdiebstahl. Sobald ein biometrisches Charakteristikum einem Diebstahl oder Missbrauch zu Opfer gefallen ist und einem potentiellen Angreifer in Form eines Plagiates zur Verfügung steht, ist es so gut wie unmöglich dieses Charakteristikum zu erneuern. Für das biometrische Charakteristikum selbst ist dieses Problem nicht ohne chirurgischen Aufwand zu lösen. Eine signifikante Verminderung der mit gestohlenen biometrischen Charakteristika einhergehenden Risiken kann jedoch dadurch erreicht werden, dass die Erneuerbarkeit biometrischer Templates, das heißt die Repräsentation eines biometrischen Charakteristikums in einem Identitätsverifikationssystem, sichergestellt wird.

2.2.5 Interoperabilität

Schlussendlich sei angemerkt, dass es bei der gegebenen großen Bandbreite biometrischer Modalitäten, Sensortypen, Merkmalsextraktionsverfahren und Templateformate, schwierig ist, ein vollständig kompatibles und alle Technologiepermutationen von Sensoren, Merkmalen, und Templatetypen unterstützendes Schema zu realisieren. Jedoch ist Interoperabilität für große offene Anwendungen (wie zum Beispiel biometrische Reisepässe, Bürgerkarten oder biometrische Bankkarten) zwingend erforderlich. Aus diesem Grund gibt es Standardisierungsbemühungen [37], [tcM], [AI], [ISOa], [ISOb], [17], deren Wirkbereich derzeit jedoch noch kein vollständiges, Ende-zu-Ende interoperables biometrisches Verifikationssystem umfasst, das Techniken zum Schutz der Privatsphäre betroffener Personen bietet.

2.2.6 Hersteller- und Produktunabhängigkeit

Ein in diesem Kontext auftretendes Problem ist das Risiko einer sich einstellenden Herstellerabhängigkeit. Viele zur Zeit am Markt existierende biometrische Verifikationslösungen basieren auf proprietären Sensoren, Templateformaten und Vergleichsalgorithmen. Der Wechsel eines Herstellers führt für den Betreiber eines biometrischen Systems daher zu beträchtlichen Kosten, wenn etablierte Standards bei der Einrichtung des Systems nicht berücksichtigt wurden.

2.3 Anforderungen an biometrische Templates

Basierend auf den im vorherigen Kapitel beschriebenen Herausforderungen können wir einen Anforderungskatalog an ein biometrisches Verifikationssystem ableiten, der sicherstellt, dass Privatsphäre und Sicherheit gleichermaßen adressiert werden [CS07] und das Risikomaß gemindert wird.

2.3.1 Geschützte Templates

Die Repräsentation biometrischer Templates, die in Verifikationssystemen eingesetzt werden und die Privatsphäre schützen, sollten als biometrische Referenz den folgenden Bedingungen genügen:

- Es ist unmöglich die ursprünglichen biometrischen Samples (z.B. Fingerbilder), die biometrischen Attribute, die (ungeschützten) Templates oder irgendeine aus dem biometrischen Sample hergeleitete Information, die Aufschluss über die Privatsphäre (wie zum Beispiel Gesundheitsinformationen, Informationen über ethnischen Ursprung, etc.) erteilt, unmittelbar aus dem geschützten Template zu entnehmen oder das geschützte Template entsprechend zu dekodieren.
- Es ist unmöglich eindeutige Verbindungen zwischen Personen innerhalb einer Datenbank oder datenbankübergreifend durch das Vergleichen von Templates herzustellen.
- Ein biometrisches Template repräsentiert ausschließlich Daten für einen spezifischen, im Voraus definierten Verwendungszweck oder eine Anwendung.

Diese Rahmenbedingungen sollten bei allen Verarbeitungsschritten wie der Speicherung, der Übertragung und beim Vergleichen biometrischer Templates eingehalten werden. Wenn eine Repräsentation eines Templates diesen Bedingungen genügt, wird es als *geschütztes biometrisches Template* bezeichnet[1].

2.3.2 Widerrufbare, erneuerbare und diversifizierbare geschützte Templates

Geschützte biometrische Templates sollten Mechanismen für deren Widerruf (zum Beispiel durch Verwendung von Zertifikaten einer Zertifizierungsstelle) unterstützen. Weiterhin sollte der Enkodierungsprozess über Möglichkeiten zur Generierung mehrerer unabhängiger geschützter Templates auf Basis der selben oder sehr ähnlicher biometrischer Charakteristika verfügen. Dieser Prozess der Generierung mehrerer unabhängiger geschützter Templates von den selben biometrischen Charakteristika bezeichnet man als Diversifikation. Die Diversifikationseigenschaft wird unter anderem zur Vermeidung unerwünschter Verknüpfungen von Personen zwischen Datenbanken und zur Vermeidung der Suche nach Personen mit sehr ähnlichen biometrischen Charakteristika benötigt.

2.3.3 Universeller Ansatz

Der Ansatz geschützter biometrischer Templates sollte generell auf jede biometrische Modalität anwendbar sein und die Kombination biometrischer Modalitäten (Fusion) zur Erreichung einer höheren Erkennungsleistung unterstützen. Vorzugsweise sollten biometrische Modalitäten für jeden Enrollee innerhalb der selben Anwendung individuell selektiert und/oder kombiniert werden können, um potenziellen Problemen mit schwachen biometrischen Charakteristika bestimmter Enrollee-Gruppen entgegenwirken zu können.

2.3.4 Interoperabilität

Obwohl Interoperabilität im Allgemeinen als zunehmende Bedrohung der Privatsphäre angesehen wird, wird das geschützte biometrische Template keine Erstellung anwendungs- oder datenbankübergreifender Verbindungen zwischen Personen erlauben (siehe Ausführungen weiter oben).

Interoperabilität schreibt vor, dass ein biometrisches Verifikationssystem auf einem im Voraus definierten Format und einer im Voraus definierten Methode basieren sollte, die die oben genannten Bedingungen erfüllen. Dieses Format sollte mit einer breiten Masse von Sensortypen und Merkmalsextraktionstypen kompatibel sein. Das Erreichen einer solchen Interoperabilität ist durch einen zweiphasigen Ansatz vorgesehen:

1 In der Publikation von Breebaart et al. wurde ein geschütztes biometrisches Template als Kombination von „pseudo identity“ (deutsch: Pseudoidentität) und unterstützenden Daten (AD) bezeichnet. Der Begriff „pseudo identity“ wird zur Harmonisierung der Sprache innerhalb dieses Dokuments und zur Anpassung an den von ISO gebrauchten Terminus „pseudonymous identifier“ abweichend von Publikation Breebaart im Weiteren durch den Begriff „pseudonymer Identifikator“ ersetzt.

- Konvertierung eines biometrischen Samples in ein modalitätsabhängiges, vordefiniertes Merkmalsdatenformat, das vorzugsweise an existierende (und/oder standardisierte) Templateformate angepasst ist;
- Konvertierung der modalitätsabhängigen, vordefinierten biometrischen Merkmalsdaten in ein geschütztes Template unter der Verwendung im Voraus definierter Formate und Methoden.

Bei diesem zweiphasigen Ansatz erlaubt das Konvertieren in ein intermediäres, vordefiniertes biometrisches Merkmalsdatenformat die Nutzung von Teilsystemen unterschiedlicher Hersteller (eingeschlossen Sensoren und Merkmalsextraktionsalgorithmen) innerhalb eines Gesamtsystems. Ein gutes Beispiel eines solchen intermediären Merkmalsdatenformats ist die Nutzung von Fingerabdruck-Minutien-Daten [ISOc].

Das gleiche Argument gilt für den zweiten Schritt. Wenn es sich bei den zur Erzeugung geschützter Templates genutzten Formaten um standardisierte Ein- und Ausgabeformate handelt und der Prozess zur Erzeugung der geschützten Templates wohldefiniert ist (entweder durch Beschreibung des gesamten Prozesses oder durch Verwendung von Konformitätskriterien), könnte eine völlig kompatible Kette beginnend beim biometrischen Sensor bis hin zum geschützten biometrischen Template erreicht werden.

2.3.5 Datenminimierung

Für effizientes Speichern, effiziente Übertragung und effizientes Abgleichen geschützter Templates und zur Gewährleistung maximalen Schutzes der Privatsphäre, sollte der mit dem Template assoziierte Binärdatenanteil minimiert werden. Gleichzeitig sollten jedoch durch diese Minimierung entstehende negative Auswirkungen auf die Erkennungsleistung ebenfalls möglichst klein gehalten werden.

2.3.6 Intrinsische Sicherheit

Die Erkennungsleistung auf Basis geschützter biometrischer Templates sollte bevorzugt vergleichbar mit der Erkennungsleistung auf aktuellem Stand der Technik bestehender konventioneller biometrischer Systeme sein. Eine geringfügige Degradierung der Erkennungsleistung ist jedoch zu erwarten und akzeptierbar, insofern sie im Gleichgewicht mit dem hierdurch zusätzlich gewonnenen Schutz der Privatsphäre steht. Weiterhin sollte eine Justierung zwischen Falsch-Akzeptanz- und Falsch-Rückweisungs-Raten auf Anwendungs- und bevorzugt auch Personenebene vornehmbar sein. Letzteres ist insbesondere zur Vermeidung von Unannehmlichkeiten durch sich wiederholende Rückweisungen bei Personen mit schwachen oder verrauschten biometrischen Charakteristika von Bedeutung. Der während eines Vergleiches erreichte Grad an Ähnlichkeit (ausgedrückt in einem Vergleichswert) kann bestimmt und, unter der Prämisse des Vorliegens einer Übereinstimmung, einer Anwendung kommuniziert werden, vorausgesetzt es existiert ein dringender Bedarf oder es entsteht durch den Erhalt einer derartigen Information ein Vorteil. Sollte keine Übereinstimmung vorliegen, ist die Ableitung eines Vergleichswertes bevorzugt intrinsisch unmöglich, um etwaigen die Sicherheit und Privatsphäre gefährdenden Attacken (wie zum Beispiel Hill-Climbing-Attacken) entgegenzuwirken.

2.3.7 Nahtlose Integration in existierende Verifikationsmethoden

Die biometrische Architektur sollte sich nahtlos an existierende zwei oder drei Faktor Verifikationsmethoden (d.h. besitz- und wissensbasierte Authentisierung) anpassen lassen und die sich hierdurch ermöglichende Verwendung mehrerer verschiedenartiger Verifikationsmethoden sollte einen multiplikativen Effekt auf den Aufwand zur Durchführung so genannter Zero-Effort-Attacken haben. Die Nutzung von sowohl anwendungs-, als auch personenspezifischer Geheimnisse sollte unterstützt werden. Zur Gewährleistung maximaler Flexibilität und maximalen Nutzerkomforts sollte die Balance zwischen besitzbasierter, wissensbasierter und biometrischer Sicherheit auf Anwendungs- oder Personenebene einstellbar sein.

Da eine betroffene Person unter bestimmten Umständen das Recht hat, der Verarbeitung biometrischer Daten auf Grund zwingend einzuhaltender, die Situation der Person betreffender gesetzlicher Grundlagen [EE95] , wie zum Beispiel Besorgnis um den Schutz der eigenen Privatsphäre, Schwierigkeiten mit dem Enrolment oder Falsch-Rückweisung, zu widersprechen, sollte ein Authentisierungssystem alternative, nicht auf biometrischen Verfahren basierende Authentisierungsmöglichkeiten unterstützen. Solche Verfahren sollten bevorzugt für den Fall des Auftretens von Fehlern beim Enrolment einer Person, bei Akquirierungsfehlfunktionen und bei Falsch-Rückweisung spezifiziert werden.

2.3.8 Architekturflexibilität

Das Template-Protection-Schema sollte sowohl online Verifikation (unter Verwendung einer zentralen Datenbank), als auch offline Verifikation (unter Nutzung einer lokalen Datenbank) unterstützen. Die Architektur sollte einen Modus bieten, bei dem sowohl zentral, als auch lokal gespeicherte Templateinformationen zur erfolgreichen Verifikation vorausgesetzt werden.

Der in Breebaarts Artikel beschriebene Ansatz beabsichtigt das Gewährleisten zusätzlicher Sicherheit bei gleichzeitig zusätzlichem Schutz der Privatsphäre. Es soll darauf hingewiesen werden, dass manche Anwendungen das sichere Speichern und Verarbeiten unter der Verwendung persönlicher Token zum Speichern der Daten ermöglichen. Derartige tokenbasierte Systeme speichern biometrische Templates in einem sicheren Bereich von Smartcards und vergleichen diese auf der Karte selbst [Ber08]. Diese Art der Anwendungen werden derzeit im ISO/IEC JTC1 SC17 standardisiert [ISOd].

2.4 Referenzarchitektur und Konzept der pseudonymen Identifikatoren

Mit der Absicht, die Verwendung eines einheitlichen Vokabulars[2] zu fördern und die zur Einhaltung der in den vorherigen Abschnitten beschriebenen Anforderungen notwendigen Architekturaspekte darzustellen, wird im Folgenden die auf sogenannten pseudonymen Identifikatoren (PIs) basierende Referenzarchitektur für geschützte Templates vorgestellt, wie sie von Breebaart eingeführt

2 Die Begrifflichkeiten befinden sich im Einklang mit dem ISO/IEC JTC1 SC37 Harmonized Biometric Vocabulary (SC37 SD2 v12 – siehe auch http://www.3dface.org/media/vocabulary.html)

wurde [Bre08]. Diese Referenzarchitektur liefert einen Rahmen für die Erzeugung und die Speicherung von geschützten biometrischen Referenzen und fügt sich ein in die Referenzarchitektur eines generischen biometrischen Systems, wie sie in Kapitel 7 zitiert wird.

Der Lebenszyklus eines pseudonymen Identifikators, seine Einbindung in die Referenzarchitektur sowie die damit verbundenen Schnittstellen und Prozesse werden in den kommenden Abschnitten erläutert. Diese Referenzarchitektur erhebt den Anspruch einer Technologieneutralität, d.h. sie soll ein generisches Framework für viele zur Zeit existierende Techniken zum Schutz von Templates bieten und darüber hinaus zukunftssicher sein. Abhängig von einer konkreten biometrischen Anwendung können unterschiedliche technische Anforderungen an die Architektur gestellt werden, weshalb für eine spezifische Implementierung einzelne funktionale Komponenten der Referenzarchitektur möglicher Weise weg gelassen oder zusätzlich hinzugefügt werden müssen.

2.4.1 Pseudonyme Identifikatoren

Pseudonyme Identifikatoren (PI – vgl. [DCB+08]) sind diversifizierbare, geschützte Identitätsverifikationsstrings (Binärer String) innerhalb eines vordefinierten Kontexts (d.h. z.B. innerhalb des geschützten biometrischen Ökosystems). Ein pseudonymer Identifikator gibt keine Informationen preis, die Aufschluss über die ursprünglich erhobenen Daten, das zu Grunde liegende biometrische Template oder die wahre Identität dessen Besitzers geben. Innerhalb eines geschützten biometrischen Ökosystems durchleben pseudonyme Identifikatoren die folgenden, in Abbildung 1 dargestellten, vier unterschiedlichen Phasen:

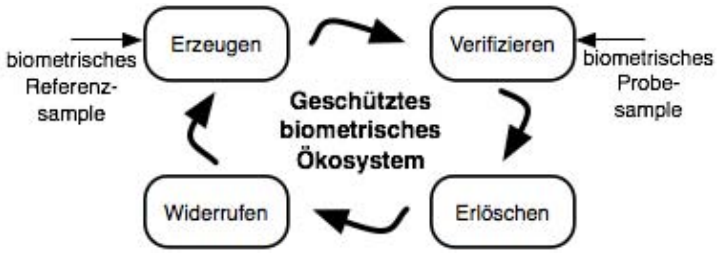

Abbildung 1: Lebenszyklus einer PI innerhalb eines geschützten biometrischen Ökosystems.

1. Erzeugung (oder Erneuerung) von pseudonymen Identifikatoren auf Basis von biometrischen Referenzdaten während einer Enrolmentphase.
2. Verifikation eines pseudonymen Identifikators auf Basis einer biometrischen Probe.
3. Ablauf der Gültigkeit eines pseudonymen Identifikators.
4. Widerruf eines pseudonymen Identifikators nach Ablauf seiner Gültigkeit.

2.4.2 Erzeugen von pseudonymen Identifikatoren

Der Prozess zur Erzeugung von pseudonymen Identifikatoren wird in Abbildung 2 dargestellt. Während einer Enrolmentphase wird für ein Individuum eine biometrische Referenz generiert.

Innerhalb dieses Prozesses werden von einem biometrischen Datenerfassungsgerät ein oder mehrere biometrische Samples, wie zum Beispiel ein Bild eines Fingerabdruckes oder ein Gesichtsphoto, erzeugt und im Anschluss von einem Feature-Extraktor zur Erzeugung biometrischer Merkmale verarbeitet, die, wenngleich nicht zwangsläufig erforderlich, vorzugsweise in einem existierenden (standardisierten) Templateformat gespeichert werden. Abschließend werden von einem Pseudonymer-Identifikator-Encoder (PIE)[3] ein pseudonymer Identifikator (PI) sowie möglicher Weise benötigte unterstützenden Daten (Auxiliary Data - AD) erzeugt, die, abhängig von den eingesetzten Methoden und Algorithmen, den folgenden Verwendungszwecken dienen können:

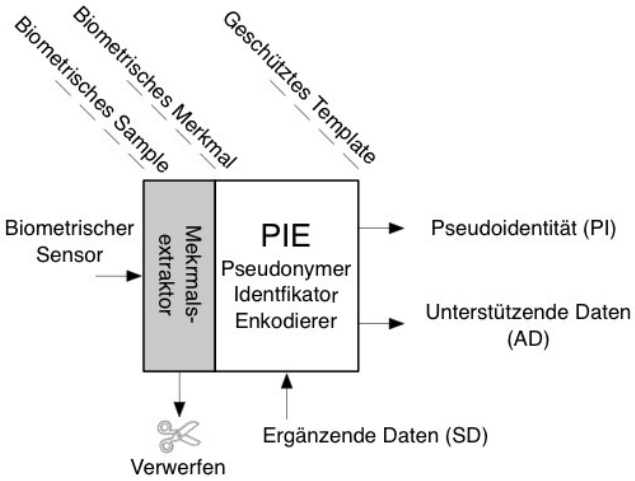

Abbildung 2: Erzeugung geschützter Templates

- Dem Erzeugen mehrerer von einander unabhängiger pseudonymer Identifikatoren für ein und dasselbe Individuum innerhalb einer Anwendung mit dem Ziel, die Erneuerung von Templates zu ermöglichen.
- Dem Erzeugen von einander unabhängiger pseudonymer Identifikatoren über Anwendungsgrenzen hinweg, um unerwünschte, Datenbanken übergreifende Verknüpfungen zu verhindern.
- Dem Erzeugen von voneinander unabhängigen pseudonymen Identifikatoren für Personen, die über sehr ähnliche biometrische Charakteristika verfügen, um Vortäuschungen durch Identifizieren biometrischer Doppelgänger zu verhindern.
- Dem Separieren von Templatedaten, um Sicherheit und Schutz der Privatsphäre zu erhöhen.
- Dem Bestimmen individueller Vergleichsparameter zur Optimierung der Erkennungsleistung.

Wenn die unterstützenden Daten vom Diversifikationsprozess genutzte Datenelemente enthalten, so werden diese Elemente entsprechend auch Diversifikationsdaten genannt. Die unterstützenden Daten können von verschiedenen Ansätzen stammen, die erneuerbare und geschützte Templates

3 Dieser Begriff bleibt eng am Englischen Original. Allerdings sollte betont werden, dass der erzeugte Identifikator pseudonym ist – der Encoder nach dem Kerckhoffs'schen Prinzip jedoch publiziert sein sollte.

erzeugen[4]. Tabelle 1 bietet einen Überblick über einige existierenden Methoden zum Schutz biometrischer Templates sowie die dazugehörigen pseudonymen Identifikatoren und unterstützenden Daten.

Methode	*Pseudonymer Identifikator*	*Unterstützende Daten*
Fuzzy commitment [JW99], [Kor08]	Hash einer geheimen Zeichenkette	Offset, Hilfsdaten
Cancelable biometrics [RCCB07]	Transformiertes Template	Transformationsparameter
Feature transformation [YB09]	Transformiertes Template	Transformationsparameter
Helper data systems [TAK05]	Hash einer geheimen Zeichenkette	Hilfsdaten
Biometric encryption [SRS98]	Kryptografischer Schlüssel	Filter und Verknüpfung zum Schlüssel
Fuzzy vault [JS02], [NJP07]	Hash einer geheimen Zeichenkette	Punktmenge P
Shielding functions [LT03]	Hash einer geheimen Zeichenkette	Authentication Challenge W
Fuzzy extractors [DRS04]	Hash einer geheimen Zeichenkette	Öffentliche Zeichenkette P
Extended PIR [BCPT07]	Verschlüsseltes Template	nicht vorhanden

Tabelle 1: Überblick über Methoden zum Schutz biometrischer Templates und deren dazugehöriger pseudonymer Identifikatoren und unterstützenden Daten.

Die Kombination von pseudonymen Identifikatoren (PI) und unterstützenden Daten (Auxiliary Data - AD) wird als ein geschütztes Template bezeichnet und verfügt über die in Kapitel 2.3 beschriebenen Anforderungen. Sowohl der pseudonyme Identifikator, als auch die unterstützenden Daten werden nach deren Erzeugung gespeichert, wohingegen die erfassten biometrischen Merkmale zerstört werden. Das Speichern von PI und AD kann auf unterschiedlichen Wegen stattfinden, die sich wie folgt in drei Kategorien einteilen lassen: zentrales Speichern (sowohl PI, als auch AD werden in einer Datenbank gespeichert), lokales Speichern (PI und AD werden gemeinsam auf einem Token gespeichert) und hybrides Speichern, durch Separierung von PI und AD (zum Beispiel durch Speichern der unterstützenden Daten auf einem Token und des pseudonymen Identifikators in einer Datenbank). Die verschiedenen Speichermodelle werden in Kapitel 7 im Detail beschrieben. Vorteile des zentralen Speicherns zumindest einer der beiden Datenelemente liegen in der Möglichkeit des Erstellens einer schwarzen Liste, des Realisierens von Prüf-Funktionalitäten (Audits) und des Ermöglichens eines simplen Widerruf-Prozesses. Die Vorteile des lokalen Speicherns sind das Nichtvorhandensein von für zentrale Datenbanken spezifischen Sicherheitsrisiken sowie der vollständige Besitz der Kontrolle über Referenzdaten bei der betroffenen Person. Das hybride Speichern zeichnet sich dadurch aus, dass sowohl die betroffene Person, als auch der Anbieter Kontrolle über die Nutzung der Templatedaten besitzt und

4 Es kann Methoden geben, die nicht zwingend unterstützende Daten verwenden. In diesem Fall werden die unterstützenden Daten als leere Zeichenkette betrachtet.

die hierdurch potenziell, durch eine zentrale Datenspeicherung entstehenden Sicherheitsrisiken reduziert werden können.

Der Pseudonyme-Identifikator-Encodierer (PIE) verwendet als Eingabe des Weiteren einige ergänzende Daten (Supplementary Data - SD), die zum Beispiel für folgende Zwecke verwendet werden können:

- Sicherheitsverbesserungen durch besitz- oder wissensbasierte Schlüssel, die vom Enrollee eingegeben werden müssen (vgl. biometrisch gehärtete Kennwörter [MRW99]);
- Verbesserungen durch anwendungs- oder systemspezifische Schlüssel oder Signaturen;
- Verbesserungen durch Limitierung des Wirkbereiches eines pseudonymen Identifikators durch Einbeziehen von digitalen Signaturen oder Zertifikaten oder von zeit- oder ortsabhängigen Informationen für die eine PI Gültigkeit besitzen soll.

Diese verwendeten ergänzenden Daten werden jedoch nicht zusammen mit den Templates gespeichert, sondern nach dem Generieren des pseudonymen Identifikators verworfen.

Einige dieser geschützten Templates könnten darüber hinaus auch zur sicheren Identifizierung genutzt werden und hierdurch zum Beispiel bei der Prüfung auf Duplikate während des Enrolments helfen.

2.4.3 Verifikation von pseudonymen Identifikatoren: PI-Recoder (PIR) Ansatz

Der Verifikationsprozess kann in zwei unterschiedliche Klassen unterteilt werden. Die erste Klasse der Verifikationsprozesse basiert auf einem PI-Recoder Ansatz. Der PI-Recoder Ansatz basiert auf dem Neuerzeugen (Recoding) eines pseudonymen Identifikators während des Verifikationsprozesses. Dieser neu erzeugte pseudonyme Identifikator wird mit dem während des Enrolments erzeugten PI verglichen (z.B. [SRS98], [JW99], [LT03], [DRS04], [TAK05], [ST06], [NJP07], [RCCB07]). Die Verifikation wird hierbei durch die Transformation eines erfassten biometrischen Erkennungssamples in einen neuen pseudonymen Identifikator PI* unter Verwendung der bereitgestellten unterstützenden Daten (vgl. Abbildung 3(a)) erreicht. Sollten dem PIE während der Enrolmentphase darüber hinaus ergänzende Daten zur Verfügung gestellt worden sein, so müssen diese Daten auch dem PIR während des Erzeugens des neuen pseudonymen Identifikators zur Verfügung gestellt werden. Nach dem Erzeugen des neuen pseudonymen Identifikators PI* durch den PIR werden alle Eingabedaten, d.h. das biometrische Sample, die Merkmalsdaten und die ergänzenden Daten gelöscht und PI* wird an einen Pseudonymen Identifikator-Comparator (PIC)[5], der PI mit PI* vergleicht, übergeben. Ausschließlich bei Identität der beiden pseudonymen Identifikatoren ist die Verifikation erfolgreich. Der Vorteil dieses Ansatzes liegt darin, dass der zwischen PIR (der zum Beispiel in einem biometrischen Sensor oder lokalen Terminal integriert sein könnte) und PIC (der sich auf Anwendungs- oder Dienstanbieterebene befinden könnte) notwendige Informationsaustausch in geschützter Form stattfindet (vgl. [CS07]).

5 Auch dieser Begriff ist eng an den Englische Orignal-Begriff „pseudonymous identifier comparator“ (PIC) angelegt, der die biometrische Probe mit der biometrischen Referenz vergleicht.

2.4.4 Verifikation von pseudonymen Identifikatoren: PI-Verification (PIV) Ansatz

Die zweite Klasse der Verifikationsprozesse arbeitet nicht mit Neuerzeugung eines pseudonymen Identifikators PI* während der Verifikationsphase, sondern verifiziert einen pseudonymen Identifikator unmittelbar auf Basis des zur Verfügung gestellten Erkennungssamples (vgl. [DKM+07], [BCI+07], [BCPT07]) und ist in Abbildung 3(b) dargestellt. Gegeben ist in diesem Fall ein geschütztes Template, bestehend aus pseudonymem Identifikator und unterstützenden Daten, ein mittels eines biometrischen Sensors erfassten Samples und, sofern während des Enrolmentprozesses verwendet, möglicherweise ergänzenden Daten. Auch wenn die Verifikation mittels eines PI-Verifiers (PIV) durchgeführt wird, werden alle verwendeten Eingabedaten nach Bekanntgabe des Verifikationsergebnisses verworfen. Der Vorteil dieses Ansatzes liegt darin, dass zur Verifikation keinerlei Austausch bzw. Übertragung von Template-Informationen benötigt wird, wenn sowohl das PIV Modul als auch das geschützte Template auf dem gleichen Gerät, zum Beispiel in Form einer On-Card-Comparison-Lösung [17], [ISOd], implementiert ist.

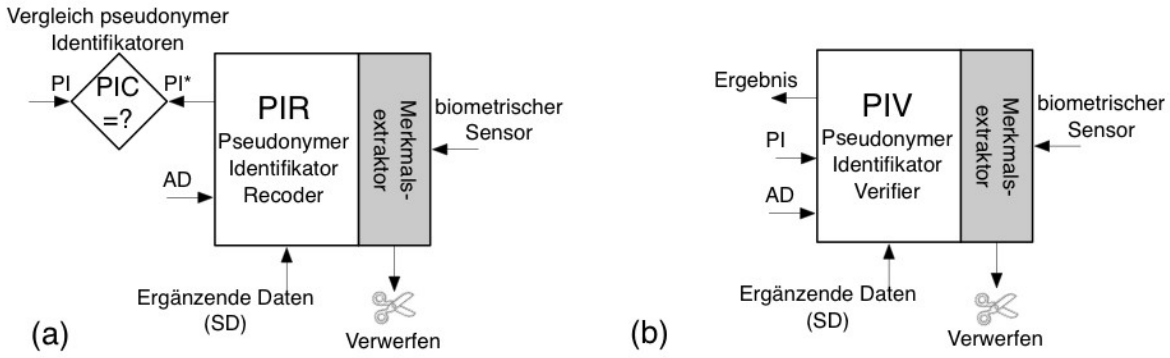

Abbildung 3: (a) Verifikation geschützter Templates via PI-Recoding und Vergleichen; (b) Verifikation mittels unmittelbarer PIV

Die Unterscheidung von PIR und PIV-Konzept zur Verifikation ist rein theoretischer Natur und historisch darin begründet, das die bei der Entwicklung der Referenzarchitektur beteiligten Unternehmen an dieser Stelle kein Einvernehmen erzielen konnten. Der Unterschied hat für Implementierungen keine tatsächliche Relevanz, da in einer Komponenten-Beschreibung ein Pseuodnymer-Identifikator-Verifier(PIV) als Komposition von PI-Recoder(PIR) und Pseudonymen Identifikator-Comparator (PIC) dargestellt und entsprechend umgesetzt werden kann.

2.4.5 Erlöschen von pseudonymen Identifikatoren

Die Gültigkeit pseudonymer Identifikatoren kann aus verschiedenen Gründen erlöschen. Zum Beispiel kann ein PI lediglich für eine nur begrenzte Zeitperiode ausgestellt worden sein oder einer Erneuerung auf Grund einer Kompromittierung bedürfen. Weiterhin können Alterungseffekte des biometrischen Charakteristikums, wie zum Beispiel im Falle des menschlichen Gesichts, die Erkennungsleistung negativ beeinflussen, sodass die Erneuerung der biometrischen Referenz erforderlich ist. Validitätsprüfungen und Löschungen können mit Hilfe von Überwachungslisten kontrolliert werden. Alternativ kann in einigen Fällen die Gültigkeitsdauer dem PIE in Form von ergänzenden Daten zur Verfügung gestellt werden, woraus eine PI-intrinsische Gültigkeitsprüfung resultiert.

2.4.6 Widerrufen von Pseudonymen Identifikatoren

Abhängig von der Implementation eines Verifikationssystems können pseudonyme Identifikatoren entweder durch Löschen aus einer Datenbank und/oder durch Entfernen der Nutzungsautorisierung eines pseudonymen Identifikators widerrufen werden. Der Widerruf ist ein aktiver Vorgang, der durch eine der beteiligten Personen initiiert wird.

Da für eine erfolgreiche Verifikation sowohl der pseudonyme Identifikator (PI) als auch die unterstützenden Daten (AD) vorliegen müssen, kann die Möglichkeit zum Widerruf einer Registrierung sowohl vom Systembetreiber als auch von der betroffenen Person ausgeübt werden, wenn PI und AD wie oben beschrieben separiert gespeichert werden (siehe auch Beschreibung der Modelle G und H in Kapitel 7). Erfolgt beispielsweise die Speicherung von PI in der Datenbank des Systembetreibers und parallel die Speicherung von AD im Token der Betroffenen Person, so kann der Betreiber den Widerruf durch Löschung des PI in der Datenbankreferenz erzielen. Die betroffene Person kann durch Zerstörung des Tokens den Wiederruf erzielen, da ein PI alleine nicht ausreichen ist.

Im Anschluss an das Widerrufen kann ein neues geschütztes Template durch ein neues Enrolment der betroffenen Person erzeugt werden, wozu, in Abhängigkeit von der gegebenen Architektur, das Erfassen eines neuen biometrische Samples und das Erzeugen einer neuen biometrischen Referenz notwendig werden kann.

2.5 Architekturübersicht

Abbildungen 4 und 5 geben einen Überblick über die Gesamtarchitektur zum Erstellen, Speichern und Verifizieren von pseudonymen Identifikatoren. Abbildung 4 beschreibt hierbei die Architektur unter Verwendung des PI-Recoder Ansatzes, wohingegen Abbildung 5 die Architektur des PI-Verification Ansatzes darstellt. Während einer Enrolmentphase werden pseudonyme Identifikatoren erzeugt. Nachdem die pseudonymen Identifikatoren erzeugt wurden, werden das biometrische Sample, die biometrischen Merkmale und die ergänzenden Daten, sofern diese von der gegebenen Anwendung erfasst und verarbeitet wurden, verworfen (oder an einem sicheren Ort zur späteren Wiederverwendung gespeichert, um zum Beispiel einen pseudonymen Identifikator erneuern zu können, ohne hierfür die physische Präsenz der betroffenen Person zu benötigen). Der pseudonyme Identifikator (PI) und die unterstützenden Daten (AD) werden ausgegeben und auf einem passenden Medium oder auf unterschiedlichen Medien, wie zum Beispiel Datenbanken, Smartcards, Barcodes,

etc., gespeichert. Während der Verifikationsphase wird im Falle des PI-Recoder Ansatzes basierend auf den übergebenen unterstützenden Daten, einem biometrischen Sample und, sofern im Anwendungsfall vorgesehen, ergänzenden Daten ein neuer pseudonymer Identifikator PI* erzeugt (Recoding). Dieser neu erzeugte pseudonyme Identifikator PI* wird zusammen mit dem zu verifizierenden pseudonymen Identifikator PI einer Anwendung übergeben, die die endgültige Verifikation durchführt. Im Falle des PI-Verification Ansatzes liefert der PIV unmittelbar ein Verifikationsergebnis, ohne vorher einen neuen pseudonymen Identifikator PI* zu bestimmen.

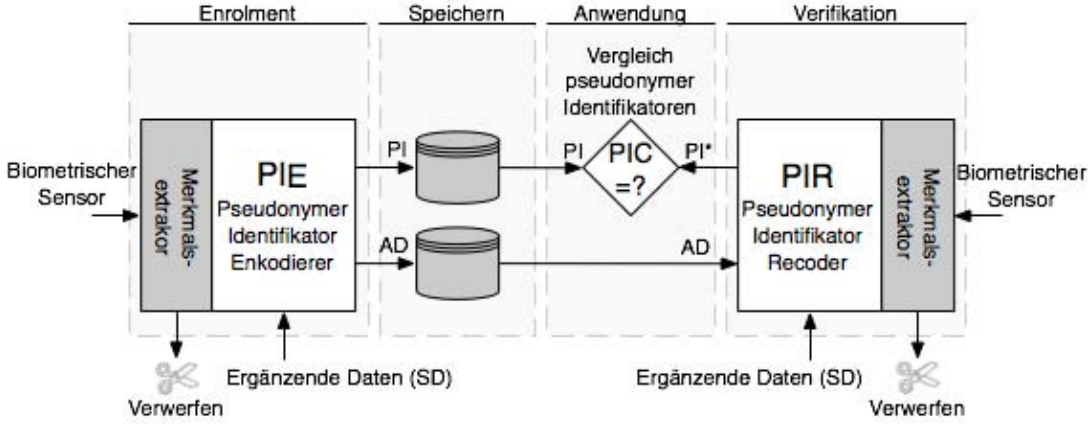

Abbildung 4: Referenzarchitektur eines Systems zum Schutz biometrischer Templates auf Basis von pseudonymer Identifikator Recoding und Vergleich.

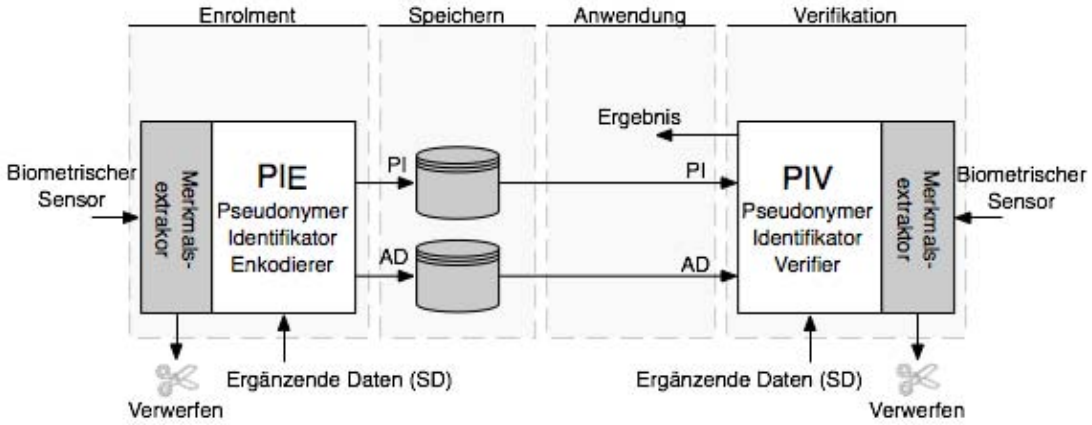

Abbildung 5: Referenzarchitektur eines Systems zum Schutz biometrischer Templates auf Basis direkter Verifikation pseudonymer Identifikatoren.

2.6 Abgleich des BioKeyS-Ansatzes mit der Referenzarchitektur

Das Bundesamt für Sicherheit in der Informationstechnik hatte die Herausforderung *Schutz der biometrischen Referenzdaten* zunächst exemplarisch an Hand von DNA-Templates untersucht und dazu das Fuzzy Commitment Verfahren nach Juels und Wattenberg angewandt [JW99]. Die Ergebnisse dieser ersten Untersuchung BioKeyS wurden im Frühjahr 2008 publiziert [Kor08].

Dieses Template Protection Verfahren basiert auf einem binären Error Correcting Coding Verfahren, dessen Encoder Nachrichten mit k Symbolen in ein Codewort mit n Symbolen ($n>k$) überführt. Der Decoder kann eine verrauschte Nachricht rekonstruieren, wenn nicht mehr als t Symbole verändert wurden. Für das Enrolment wird ein Geheimnis s zufällig erzeugt. Aus s wird einerseits ein Codewort generiert und mit dem binarisierten Merkmalsvektor verknüpft und andererseits ein Hashwert $h(s)$ über das Geheimnis berechnet (siehe auch die Beschreibung in Kapitel 3).

Das Fuzzy-Commitment-Verfahren lässt sich auf die Referenz-Architektur abbilden, wenn folgenden Zuordnungen getroffen werden.

i) PI – pseudonymer Identifikator: Hashwert über Geheimnis s.

ii) AD – unterstützende Daten: Hilfsdaten, die sich aus Verknüpfung von Codewort und binärem Merkmalvektor ergeben. (Zusätzlich – offset der selektierten Bits)

iii) SD – ergänzende Daten: Hintergrundwissen der betroffenen Person, das zur Authentisierung erforderlich ist wie zum Inputwerte zur Reduktion des binären Vektors vor der Verknüpfung (diese Information könnte auch als Teil von AD gespeichert werden)

Mit dem Projekt BioKeyS_Multi und dem Kooperationspartner secunet wurde in einer zweiten Projektphase eine praktisch einsetzbare Implementierung für Fingerbilderkennungssysteme auf Basis von mehreren Fingern realisiert, die in der Publikation von Korte et al. dargestellt ist [KMN09]. Die Umsetzung des BioKeyS-Ansatzes basiert auf dem Fuzzy-Vault-Verfahren [JS02], das unter anderem von Nandakumar et al. [NJP07] auf Fingerbildererkennung angewandt wurde. Das Verfahren wird im Kapitel 3 detailliert beschrieben. Der Grundgedanke ist, die durch die Feature-Extraction ermittelten Minutien-Punkte der biometrischen Referenz zur Berechnung eines Geheimnisses K zu nutzen, das durch ein Polynom P kodiert wird, wobei die Minutien als Stützstellen des Polynoms verwendet werden. Die Anzahl und Ordnung der Minutien-Punkte ist dabei unbestimmt, eine Mindestanzahl von Minutien ist jedoch zur Rekonstruktion des Polynoms notwendig. Um das Polynom und damit das Geheimnis einem Angreifer nicht zugänglich zu machen, werden die Minutien in einer Wolke von sogenannten Chaff-Points versteckt. Wie bei vielen Fingerbilderkennungsverfahren ist auch beim Fuzzy-Vault-Verfahren das Alignment wichtig, um mögliche Rotationen zwischen biometrischer Probe und biometrischer Referenz zu kompensieren. Nach [NJP07] kann diese wichtige Alignment-Information durch charakteristische Punkte zum Beispiel Linienabschnitte mit großer Krümmung berechnet werden und als Helper Data (unterstützende Daten / Auxiliarry Data – AD) gespeichert werden. Diese aus der Literatur bekannte Möglichkeit ist jedoch in der gegenwärtigen BioKeyS_Multi-Implementierung noch nicht realisiert, bedeutet aber eine interessante zukünftige Optimierungsmöglichkeit.

Der auf dem Fuzzy-Vault basierende BioKeyS_Multi-Ansatz lässt sich auf die Referenz-Architektur abbilden, wenn folgenden Zuordnungen getroffen werden.

iv) PI – pseudonymer Identifikator: Hashwert über Geheimnis K, wodurch Polynom P bestimmt ist.

v) AD – unterstützende Daten: Punktwolke mit Minutien-Punkten und Chaff-Points und optional zusätzlich die Alignment-Information zur Rotation der Fingerbilder in eine Normallage.

vi) SD – ergänzende Daten: Hintergrundwissen der betroffenen Person, das zur Authentisierung erforderlich ist wie zum Beispiel der Grad des berechneten Polynoms.

Dieser Abschnitt zeigt, dass das BioKeyS/BioKeyS_Multi-Verfahren grundsätzlich durch die Referenzarchitektur abgebildet werden kann. Damit ist jedoch keine Aussage darüber getroffen, ob die in Kapitel 2.3 formulierten Anforderungen unter anderem hinsichtlich der Sicherheit des Verfahrens erfüllt sind. Dazu sei auf Kapitel 3 und die Publikation von Schreier und Boult verwiesen [SB07].

2.7 Schlussfolgerung und Bedeutung der Architektur

Dieses Kapitel stellte Herausforderungen und daraus resultierende Anforderungen an Methoden zum Schutz biometrischer Templates dar. Basierend auf diesen Anforderungen wurde eine Referenzarchitektur beschrieben, die die relevanten Schnittstellen und Prozesse für den Schutz biometrischer Templates auf technologieneutrale Art erläutert. Diese Architektur wurde inzwischen Bestandteil des ISO/IEC CD 24745 Biometric Template Protection [ISOtp], dessen Genese und Status in Kapitel 7 näher beschrieben wird.

3 Verfahren zu Biometric Template Protection

3.1 Einführung

Biometrie hat viele Vorteile gegenüber Passwort- oder Token-basierten Authentifizierungsverfahren. Hierzu gehören die gute Benutzerfreundlichkeit und die eindeutige Verknüpfung zwischen dem Benutzer und seiner Identität. Biometrische Referenzdaten müssen lokal oder zentralisiert gespeichert werden, um einen Vergleich der biometrischen Daten während der Authentifizierung durchführen zu können. Die Vergleichsergebnisse basieren auf der Ähnlichkeit zwischen den Daten. Die Speicherung der biometrischen Referenzdaten führt leider zu vielen Problemen bezüglich der Sicherheit und der Privatsphäre. Die wichtigsten werden im Folgenden aufgelistet.

- **Verstärkter Identitätsdiebstahl:** Gespeicherte biometrische Referenzdaten können von externen Angreifern über Systemschwachstellen oder von internen Angreifern wie böswilligen Systemadministratoren entwendet werden. Mittels solcher biometrischer Referenzdaten können synthetische biometrische Merkmale (Attrappen) hergestellt werden. Zum Beispiel wurde in [CLM07] gezeigt, wie ein Fingerabdruckbild aus standardmäßigen Referenzen rekonstruiert werden kann. Dies ist auch möglich für andere biometrische Modalitäten wie Gesicht und Iris [AC09]. Deshalb ist es dringend notwendig, die gespeicherten biometrischen Referenzdaten zu schützen. Allerdings muss angemerkt werden, dass eine auf digital gesendeten Daten basierende Authentisierung auch ohne die Rekonstruktion der biometrischen Modalität angreifbar ist.
- **Unveränderlichkeit:** Beim Einsatz von Biometrie werden Personen und ihre Identität durch ihre einzigartigen biometrischen Charakteristiken verknüpft. Falls biometrische Daten kompromittiert werden, können sie nicht einfach als ungültig erklärt oder erneuert werden, wie es bei der gewöhnlichen Passwort- oder besitzbasierten Authentisierung möglich ist. Man kann lediglich eine andere biometrische Modalität wählen oder versuchen, die kompromittierten Daten zu modifizieren. Bedauerlicherweise sind beide Möglichkeiten keine angebrachten Lösungswege. Wir besitzen eine begrenzte Anzahl biometrischer Modalitäten (z. B. zehn Finger, ein Gesicht, zwei Iriden) und um sie zu verändern, sind nur komplizierte Methoden wie kosmetische Chirurgie oder Transplantationen möglich.
- **Verknüpfung:** Wenn dieselbe biometrische Modalität in mehreren Applikationen eingesetzt wird, sind dessen Datensätzen potenziell verknüpfbar. Ein unseriöser Datenverwalter kann Aktivitäten einer betroffenen Person in externen Applikationen verfolgen und diese Informationen zu unrechtmäßigen Zwecken ausnutzen. Zusätzlich sind, falls eine biometrische Identität in einer Applikation entschlüsselt ist, alle anderen Applikationen in ihrer Sicherheit gefährdet.
- **Privatsphäre:** Aus Sicht der Gesetzgebung sind die Erfassung und die Verwendung biometrischer Daten oftmals streng limitiert. Allerdings müssen in vielen Applikationen wie dem Automatisierte Fingerabdruckidentifizierungssystem (AFIS), dem US-VISIT-Programm oder dem ePassport aufgrund der Interoperabilität biometrische Samples gesammelt, speichert und übertragen werden. Systeme, die statistische Methoden verwenden, benötigen ebenso Samples, so dass Systemparameter aktualisiert und die

Systemleistung im Falle neu hinzugekommener Personen optimal gehalten werden kann. Es gibt Analysen, die den Einfluss von Krankheiten und sexueller Ausrichtung auf Fingerabdrücke untersuchen. Augenkrankheiten wie eine freischwebende Iriszyste oder das diffuse Iris-Melanom können das Irismuster verändern. Anhand eines Gesichtsfotos kann man ebenso Geschlecht und Rasse der Benutzer bestimmen. Die DNA enthält die genetischen Informationen des Menschen. Somit werden private Informationen, die nicht für die Authentisierung notwendig sind, mit gespeichert.

- **Zentrale Speicherung:** Die Verwendung zentraler Datenbanken ist ein kritischer Punkt im Umgang mit privaten Daten. Datenbanken sind die häufigsten Angriffsziele. Gespeicherte Informationen können abgefangen, kopiert oder verfälscht werden.
- **Hill-Climbing:** Die Entscheidung biometrischer Authentisierung beruht auf der Ähnlichkeit zwischen gespeicherter Referenz und Probe. Eine Rückmeldung über den Vergleichswert kann erhalten werden, indem ein trojanisches Pferd in den Komparator eingeschleust wird. Ein Angreifer kann diese Information auswerten, um ein biometrisches Sample zu rekonstruieren oder eine biometrische Charakteristik zu imitieren. Freie Software oder eine offene Beschreibung des biometrischen Algorithmus unterstützen diese Art des Angriffs.

In [CS07] wurde auch erläutert, dass einfaches Ersetzen traditioneller Authentifizierungsmethoden durch Biometrie zu „Zero-Gain“-Spielen in der Sicherheit führt (kein Sicherheitsgewinn). Dies zeigt, dass der Schutz der biometrischen Templates aus Gründen der Sicherheit und der Privatsphäre dringend notwendig ist.

Kryptographische Verschlüsselung ist eine weit verbreitete Methode, um digitale Daten zu sichern. Leider sind diese Verfahren sehr sensitiv gegenüber Änderungen der Eingabedaten. Eine Authentifizierung basierend auf verschlüsselten biometrischen Referenzdaten ist unmöglich, da biometrische Daten sich von Aufnahme zu Aufnahme aufgrund des Aufnahmerauschens, der Änderungen der Umgebungsbedingungen (Temperatur, Luftfeuchtigkeit), des Alterungseffekts usw. verändern. Infolgedessen wäre die Entschlüsselung während der Authentifizierung erforderlich. Jedoch hat dies den Nachteil, dass verschlüsselte Referenzdaten dann als Klartext wieder verfügbar wären und sich somit eine potenzielle Angriffsstelle ergibt. Außerdem ist ein Schlüsselverwaltungssystem notwendig, welches sehr aufwendig werden kann. Kryptographische Verschlüsselungen sind widerstandsfähig gegen externe Angriffe, jedoch können sie nicht gegen systeminterne Angriffe, zum Beispiel durch einen böswilligen Systemadministrator, schützen.

Als Alternative kann das sogenannte „Comparison-On-Card“-Verfahren die Privatsphäre schützen [Ber08]. Es wird zum Beispiel eine Smartcard verwendet, die die Funktionen des ganzen biometrischen Systems einschließlich der Feature-Extraktion und dem Vergleichsmodul vereint. Biometrische Referenzdaten sind auf der Smartcard gespeichert und gegen das Auslesen aus der Karte geschützt. Die Karte selbst berechnet das Authentifizierungsergebnis. Auf diese Art und Weise wird die Privatsphäre geschützt. Offensichtlich hat diese Methode die Schwäche, dass keine Identifikation möglich ist. Zurzeit ist dieses Verfahren nur für die Fingerabdruckerkennung verfügbar. Für andere Modalitäten ist es unmöglich, da die Erkennungs- und Vergleichsmodule zu komplex sind, um sie in eine Smartcard zu integrieren. Zusätzlich müssen Mechanismen verwendet werden, die die Echtheit der Karte prüfen.

In Abschnitt 2.3 wurden Anforderungen zum Schutz der biometrische Referenzdaten zusammengefasst. Weder normale kryptographische Verfahren noch die „Comparison-on-Card“-Methode entsprechen diesen Anforderungen. Deswegen sind Template-Protection-Verfahren

notwendig, die generalisierte Lösungen anbieten, um Sicherheit zu verbessern und Datenschutz zu garantieren. In diesem Kapitel werden die existierenden Methoden vorstellt und untersucht.

Template-Protection-Verfahren können unterschiedliche, unabhängige sichere Referenzen aus biometrischen Daten generieren. Die allgemeine Referenzarchitektur ist in Abschnitt 2.4 gezeigt. Sichere Referenzen bestehen aus sogenannten pseudonymen Identifikatoren (PI) und unterstützenden Daten (Auxiliary Data – AD). Pseudonyme Identifikatoren sind geheime Zeichenketten oder transformierte Daten, die zufällig im Pseudonymer-Identifikator-Encoder (PIE) erzeugt werden und unabhängig von den originalen biometrischen Daten sind. Mit der Hilfe der AD werden gleiche PI aus den biometrischen Daten des Genuine-Benutzer im PI-Recoder (PIR) wiederhergestellt. Sowohl PI als auch AD geben keine Information über die originalen biometrischen Daten preis. Im Folgenden werden biometrische Template-Protection-Verfahren vorgestellt und nach den verwendeten Schutzmechanismen klassifiziert. Die Sicherheitsanalyse der Verfahren wird in Abschnitt 3.3 dargestellt.

3.2 Existierende biometrische Template-Protection-Verfahren

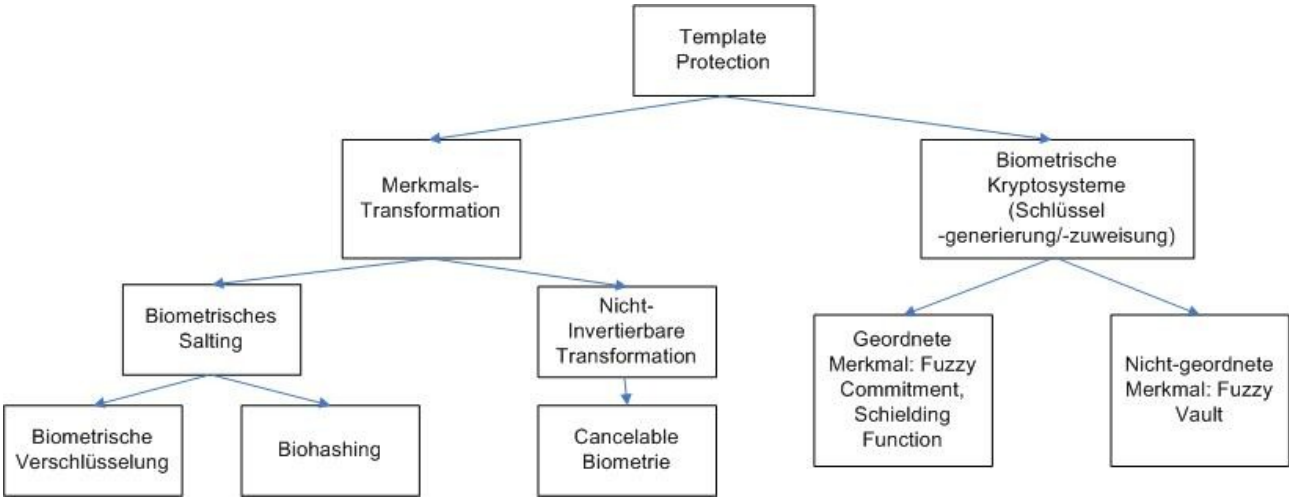

Abbildung 6: Überblick über Template-Protection-Verfahren

Seit biometrische Techniken mehr und mehr eingesetzt werden und deren Erkennungsleistungen sich weiter erhöhen, ist die Zuverlässigkeit aufgrund der potenziellen Sicherheitsrisiken und Datenschutzprobleme stark gefährdet. Daher wurden Template-Protection-Verfahren entwickelt, um biometrische Daten und Systeme effizient zu schützen. In [Jain08] wurde eine Übersicht über die Sicherheit biometrischer Systeme und der Stand der Technik im Bereich Template Protection dargestellt. In den existierenden Verfahren werden entweder eine Transformationsmethode oder Kryptografie-basierende Verfahren verwendet. In der Transformationsmethode werden Zufallswerte als Salz (Salt) oder Transformationsparameter verwendet, um biometrische Daten zu schützen. In biometrischen Kryptosystemen können unterschiedliche Schemata für geordnete und nicht-geordnete Merkmale verwendet werden. In Abhängigkeit der Anwendungen unterscheiden die Kryptoverfahren zwischen Schlüsselgenerierung und -zuweisung. Bei der Schlüsselgenerierung kann ein eindeutiger Schlüssel aus den biometrischen Daten erzeugt werden. Schlüsselzuweisungen verwenden unterschiedliche zufällige Schlüssel, um biometrische Daten zu schützen. In Abbildung

6 wird ein Überblick über Template-Protection-Verfahren und deren Abhängigkeiten dargestellt. Im Folgenden werden die Verfahren sowie deren Eigenschaften detailliert vorgestellt.

3.2.1 Transformationsmethoden

In den Transformationsmethoden werden die pseudonymen Identifikatoren (PI) aus biometrischen Merkmalen mit zufälligem Salt oder einer nicht-invertierbaren Funktion abgeleitet. Sowohl der zufällige Salt als auch die Parameter der nicht-invertierbaren Funktionen sind benutzer- und anwendungsspezifisch. Deshalb müssen diese unterstützenden Daten gespeichert und aus Sicherheitsgründen geheim gehalten werden.

3.2.1.1 Biometrisches Salting

Biometrisches Salting ist ähnlich dem Password-Salting in der Kryptografie. Lange Zufallszahlen werden als „Salt“ verwendet, um die originalen biometrischen Daten zu schützen. Biometrische Verschlüsselung und Biohashing sind zwei typische biometrische Salting-Algorithmen.

Biometrische Verschlüsselung (Biometric Encryption) verwendet biometrische Merkmale im Frequenzbereich. Merkmale können mithilfe von Zufallszahlen stark verrauscht werden. Vergleichbar ist dieses Verfahren mit einem Whitening-Prozess, nach dem die Daten wie ein Rauschenaussehen und die originalen Informationen nicht mehr zu erkennen sind. In [SRS98] wurde beispielsweise biometrische Verschlüsselung und deren Implementierung für Fingerabdruckverfahren vorgestellt. Zuerst werden mittels eines Korrelationsfilters Merkmale aus den Fingerabdruckbildern im Frequenzbereich berechnet. Diese Merkmale sind robust gegen Variationen der Fingerabdrücke. Der Korrelationsfilter und die Merkmale werden beide mit Zufallszahlen multipliziert. Die Ergebnisse werden in den Ortsbereich zurück transformiert und binarisiert. Die Ausgabe ist ein verrauschtes Bild. Anschließend wird eine zufällig generierte geheime Zeichenkette mit einer Lookup-Tabelle in das Bild virtuell eingebettet. Wenn das erste Bit der Zeichenkette Eins ist, werden beispielsweise 5 Pixel in dem Bild zufällig ausgewählt, deren

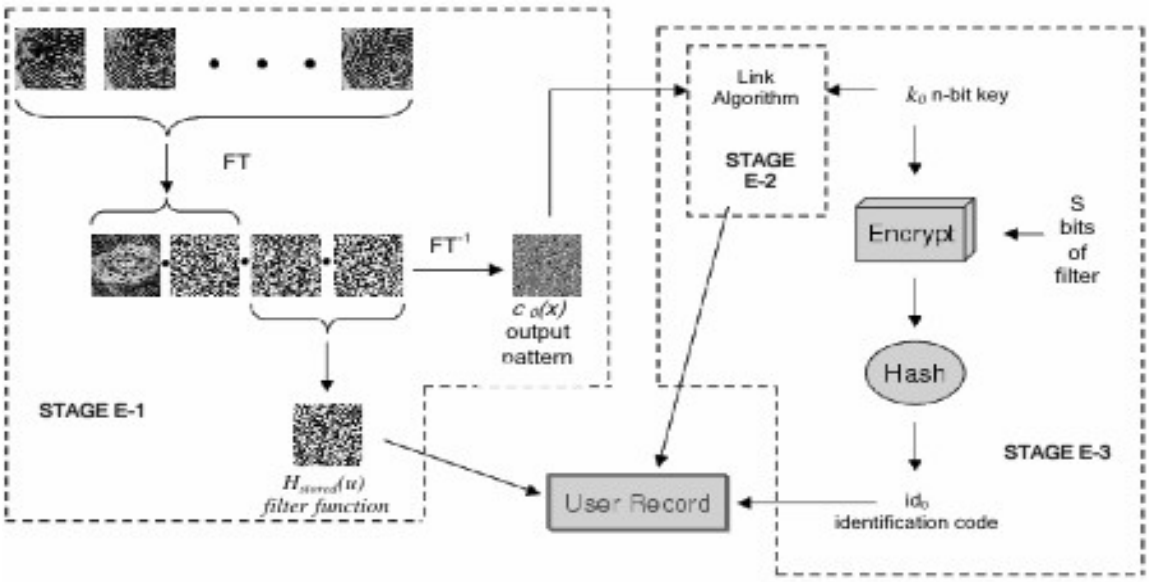

Abbildung 7: Enrolment bei biometrischer Verschlüsselung in [KYE+04]

Werte Eins entspricht. Die Position der Pixel werden in den Lookup-Tabelle notiert. Der verrauschte Korrelationsfilter, die Lookup-Tabelle und der Hashwert der geheimen Zeichenkette werden als sicheres Template gespeichert. Während der Verifikation kann das verrauschte Bild durch Multiplikation des neu gemessenen Fingerabdruckbildes und des gespeicherten Filters im Frequenzbereich rekonstruiert werden. Mithilfe der Lookup-Tabelle kann die geheime Zeichenkette nach Beschluss der Mehrheit geschätzt werden. Der gespeicherte und der neu berechnete Hashwert werden miteinander verglichen.

Statt eine geheime Zeichenkette zu benutzen, können die verrauschten Daten direkt als PI verwendet werden. Die Zufallszahlen, die die Merkmale verrauschen, werden als AD-Daten für die Verifikation gebraucht. Biometrische Verschlüsselung ist eine gut geeignete Methode für Merkmale, die im Frequenzbereich extrahiert werden. Dieses Verfahren kann auf die meisten biometrischen Modalitäten angewendet werden, wie z.B. Gesichter, Fingerabdrücke [SKK04] oder Venenmuster [SK09].

Biohashing-Verfahren konvertieren biometrische Merkmale zu binären Codes – so genannten Biohashes – mithilfe großer Mengen benutzerspezifischer Zufallszahlen. Allein aus den Biohashes können keine Informationen über die originalen Merkmale erhalten werden. Eine Rekonstruktion benötigt die benutzerspezifischen Zufallszahlen, die meistens in einem Token gespeichert sind. Deswegen ist Biohashing eine reine Verifikationslösung und nicht für eine Identifikation geeignet. Zum Vergleich zweier Biohashes kann der Hamming-Abstands-Komparator eingesetzt werden. Um die Sicherheit weiter zu erhöhen, ist es möglich, Biohashing mit Fuzzy Encryption zu kombinieren und Biohashes als Eingabe für Fuzzy Encryption zu benutzen. In Abbildung 8 wird der Biohash-Erstellungsprozess dargestellt.

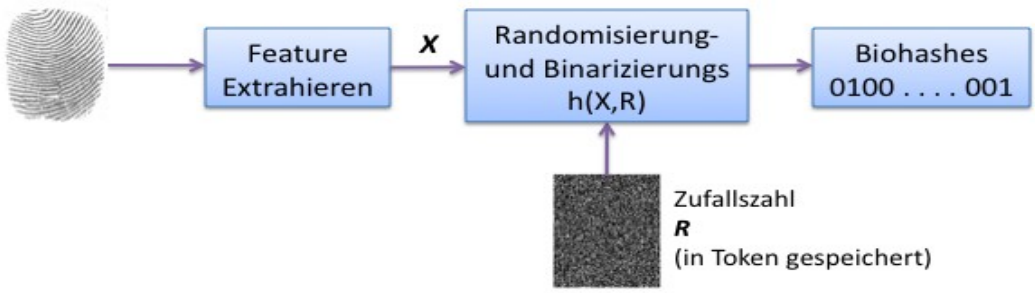

Abbildung 8: Generierung von Biohashes

Biohashing-Verfahren bestehen aus einem Randomisierungs- und einem Binarisierungsprozess. In [MS09] wurden drei unterschiedliche Biohashing-Algorithmen beschrieben. Ein Verfahren verwendet eine zufällige Skalarbinarisierung. Dabei wird ein einzelnes Element der Merkmalsvektoren mit zufällig generierten Schwellwerten verglichen, um einen binären Vektor zu erhalten. Die Schwellwerte müssen die gleiche Verteilung wie die biometrischen Merkmale aufweisen, so dass die Biohashes unterschiedlicher Benutzer gleichmäßig verteilt sind. Ein anderes Verfahren erstellt aus den originalen Merkmalsvektoren einen Vektor komplexer Zahlen. Der reale Teil enthält die originalen Merkmale und der imaginäre Teil enthält zufällig generierte Zahlen. Die Biohashes werden durch eine Binarisierung des Winkels der komplexen Zahlen errechnet. Das am häufigsten verwendete Biohashing-Verfahren basiert auf der Zufalls-Projektions-Methode (Random Mapping). Ein m-dimensionaler Merkmalsvektor wird auf n Basisvektoren projiziert, wobei die n Basisvektoren zufällig generiert und orthogonal zueinander sind. Die neuen Merkmale werden mit ihrem Interclass-Mittelwert binarisiert.

Theoretisch können diese Verfahren auf alle biometrischen Merkmale mit fester Länge angewendet werden. Die Verfahren wurden erfolgreich für Handfläche- [CTG05], Fingerabdruck- [TGG04] und Gesichts- [TGN06] [MS09]-Erkennungssysteme umgesetzt. Gute Erkennungsleistungen wurden erzielt. In [TGG04] wurde gezeigt, dass das Biohashing mit Zufalls-Projektion die optimale Erkennungsleistung für Fingerabdrücke erreichte und die Gleichfehlerrate (Equal-Error-Rate EER) gleich null ist. Wegen der benutzerspezifischen Projektion sind die Biohashes unterschiedlicher Benutzer gleichmäßig verteilt und unabhängig voneinander. Die Standardabweichung des Interclass-Abstand kann sehr klein sein, so dass der Interclass- und der Intraclass-Abstand nicht mehr überlappen. Aber wie in [KCZ06] gezeigt, kann eine Reduktion der Erkennungsleistung beobachtet werden, wenn eine nicht-authentische Person das Token einer authentischen Person verwendet (Token-Stolen Fall).

3.2.1.2 Cancelable Biometrie

Cancelable Biometrie basiert auf Transformationsfunktionen. Dabei werden die originalen Merkmale oder Proben mit einer nicht-invertierbaren oder Verzerrungsfunktion transformiert. Verschiedene sichere Templates können durch Änderung der Funktionsparameter erzeugt werden. Mit unterschiedlichen Transformationsparametern kann ebenso eine Unähnlichkeit zwischen den originalen und den transformierten Daten erreicht werden. Außerdem sollte die Transformation keinen großen Einfluss auf die Erkennungsleistung haben. Die Sicherheit dieses Verfahrens liegt in der Komplexität, die originalen Daten aus den transformierten Daten zu schätzen.

In [RCCB07] und [BCR04] werden unterschiedliche Transformationsfunktionen für Gesicht und Fingerabdruck vorgestellt. [RCB01] zeigt eine Morphing-Funktion, die 2-D-Gesichtsbilder verändert. Nach dem Morphing können normale Gesichtserkennungsverfahren benutzt werden, um eine Authentifizierung durchzuführen. Leider wurde die Erkennungsleistung dieses Verfahrens nicht veröffentlicht. Für die Minutien-basierte Fingerabdruckerkennung können kartesische und polare Transformationen oder Transformationen mittels Oberflächenfaltung (surface folding) eingesetzt werden [RCCB07] [BCR04]. Die kartesische Transformation bildet die Minutien, die in einer gleichverteilten Zelle liegen sind, zufällig auf eine neue Zelle ab. In gleiche Weise unterteilt und verwürfelt die Polar-Transformation die Minutien-Unterteilung in einem Polar-

Koordinatensystem. Die Transformation mittels Oberflächenfaltung kann die Position der Minutien bspw. mit einer Funktion mit Gauß-Kernel verändern. Die Simulationsergebnisse in [RCCB07]

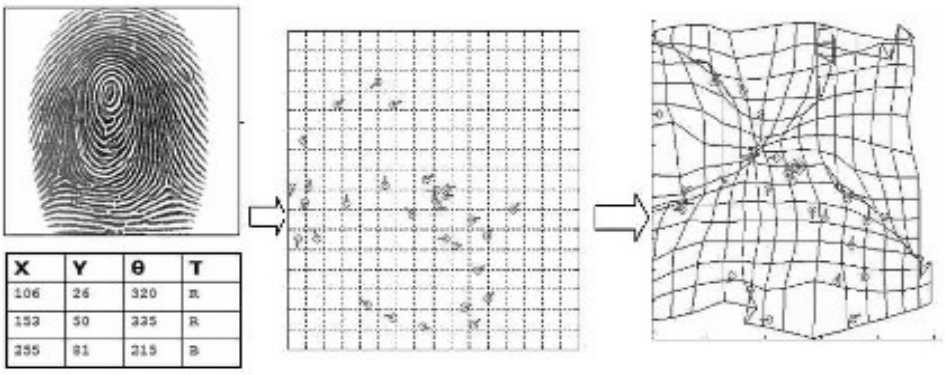

Abbildung 9: Oberflächenfalten-Transform von Minutien in [RCCB07]

zeigen, dass die Erkennungsleistung nach der Transformation nachlässt und die Oberflächenfaltungstransformation die besten Erkennungsergebnisse liefert.

Im Prinzip sind die Verfahren für alle biometrischen Systeme anwendbar. Sie müssen jedoch durch Verwendung unterschiedlicher Transformationen an die Eigenschaften der biometrischen Daten angepasst werden, um die Sicherheit zu garantieren.

3.2.2 Biometrische Kryptosysteme

Biometrische Kryptosysteme kombinieren Kryptografie mit Fehlerkorrekturverfahren, um biometrische Daten zu schützen. Kryptografie ist ein weit verbreiteter Schutzmechanismus für digitale Daten. Aber Kryptografie ist empfindlich gegenüber den Variationen biometrischer Daten, die beispielweise aufgrund der Alterung, der Änderung der Beleuchtung, der Luftfeuchtigkeit oder durch Sensorrauschen entstehen. Deswegen werden Fehlerkorrekturverfahren zur Kompensation dieser Variationen verwendet. Biometrische Kryptosysteme unterscheiden sich nach den eingesetzten kryptografischen Protokollen. Fuzzy Commitment ist für geordnete Merkmale geeignet, die als Vektor beschrieben werden können. Der Fuzzy Vault ist dagegen für ungeordnete Merkmale wie Minutien von Fingerabdrücken konzipiert.

3.2.2.1 Fuzzy-Commitment-Verfahren

In [JW99] haben Juel et al. das Fuzzy-Commitment-Verfahren beschrieben, das biometrische Daten mit existierenden kryptographischen Hash-Funktionen schützt. Ähnlich wie bei der Passwort-basierten Authentifizierung auf Unix-Computern, werden die originalen Daten nie im Klartext gespeichert oder verglichen. Eine Authentifizierung ist möglich, ohne die originalen Daten zu kennen. Aufgrund der Verwendung von Fehlerkorrekturverfahren ist Fuzzy-Commitment tolerant gegenüber Änderungen der biometrischen Daten.

In [LT03] werden praktische Umsetzungen des Fuzzy-Commitments gezeigt. Zu den Verfahren gehören das *Helper-Data-Schema* sowie neue Shielding-Funktionen. In [TG04] wurden die erforderlichen Eigenschaften der Schemas mathematisch formuliert sowie die möglichen Kapazitäten der Geheimhaltung und Identifikation (secrecy capacity und identification capacity) beschrieben. Es wurde auch gezeigt, dass das Schema vergleichbar mit der Geheimnisextraktion des als „Common Randomness“ [AC93] vorgestellten Verfahrens ist. Aus zwei korrelierten Daten, zum Beispiel biometrischen Daten eines Benutzers im Enrolment- und im Verifikationprozess, kann eine geheime Zeichenkette mithilfe der öffentlichen Hilfsdaten (Helper Data) extrahiert werden. Die Hilfsdaten kompensieren die Unterschiede zwischen den Enrolment- und Verifikationsdaten. Außerdem geben diese keine Informationen über die geheime Zeichenkette preis.

Dieser Ansatz konvertiert die biometrischen Daten idealerweise in einen gleichförmig unabhängig verteilten binären Vektor. Eine geheime Zeichenkette wird zufällig generiert, deren Hashwert die PI ist. Die Zeichenkette wird um ein Codewort mit einem Fehlerkorrekturverfahren abgebildet, das die gleiche Länge wie der binäre Merkmalsvektor hat. Das Codewort wird mit dem binären Vektor mittels XOR-Operation verknüpft. Die resultierende Bitfolge offenbart kaum noch Informationen über seine biometrische Herkunft und kann als AD gespeichert werden. Während der Verifikation kann ein „beschädigtes Codewort“ aus der biometrischen Probe und den Hilfsdaten erhalten werden. Wenn die Abweichung zwischen den Enrolment- und Verifikationsdaten nicht zu groß ist, kann der entsprechende Fehlerkorrektur-Dekodierer die Fehler korrigieren und die geheime

Zeichenkette rekonstruieren. Ein exakter Vergleich wird zwischen dem gespeicherten PI und dem Hashwert der rekonstruierten Zeichenkette gemacht, um ein Verifikationsergebnis zu liefern. Ein Blockdiagramm ist in Abbildung 10 dargestellt.

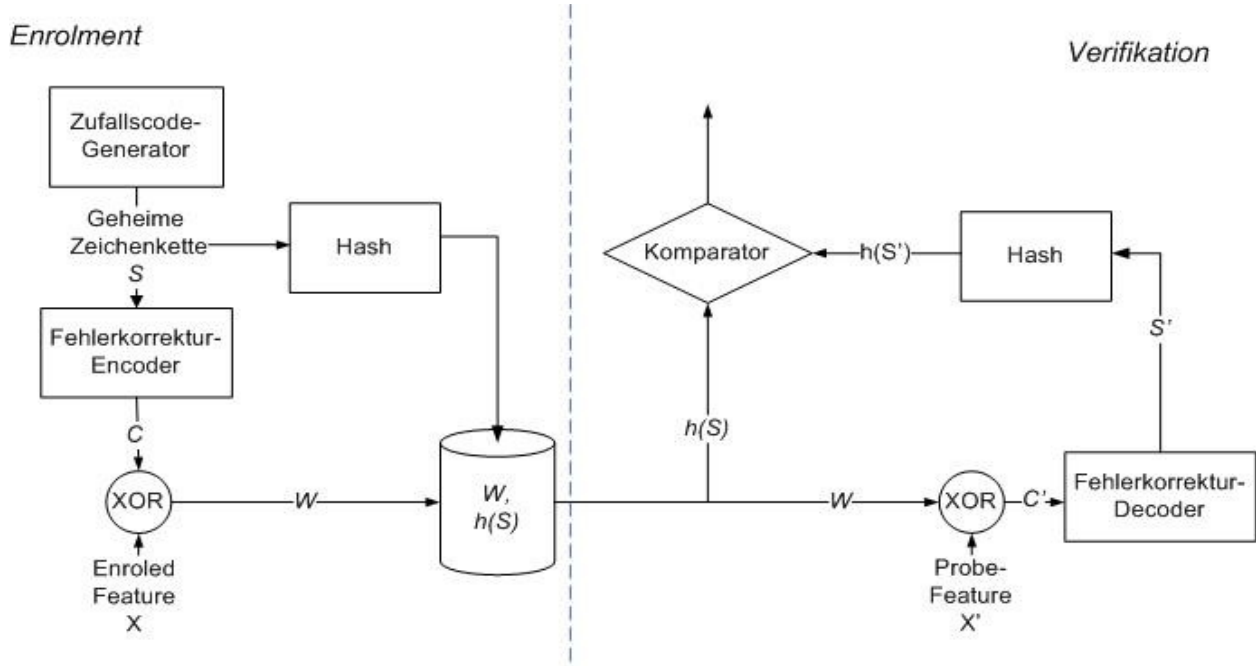

Abbildung 10: Blockdiagramm einer Implementierung des Fuzzy-Commitment-Verfahrens

Das Fuzzy-Commitment-Verfahren ist eins der meist verwendeten Template-Protection-Verfahren. Es wurde erfolgreich für 2-D- und 3-D-Gesichtserkennung [VK06][Zhou07], Iriserkennung [HD05], Fingerabdruckerkennung mit Fingerlinienrichtung [VT03], Handflächenvenenerkennung [HB09], Erkennung genetischer Fingerabdrücke [KKM08][BK06] und Ohrerkennung [TVI04] implementiert. Die experimentellen Ergebnisse zeigen, dass die Implementierungen des Verfahrens keinen großen Einfluss auf die Erkennungsleistung haben. Die Erkennungsleistung ist von der Länge der geheimen Zeichenkette abhängig, welche durch den verwendeten Fehlerkorrektur-Code limitiert ist. Wenn biometrische Daten relativ robust sind und eine hohe Entropie haben, können längere Zeichenketten extrahiert werden. Wenn die biometrische Daten jedoch sehr stark verrauscht sind, kann es passieren, dass bei der Verifikation entstandene Fehler nicht mit dem Fehlerkorrekturverfahren korrigiert werden können. Standardmäßige Fehlerkorrekturverfahren, wie BCH-Code oder RS-Code, können maximal 25% Bitfehler korrigieren. Die Rekonstruktion der Zeichenkette ist somit nicht immer machbar. Ein gutes Kodierungsverfahren kann eine bessere Kodierungsrate ermöglichen und damit die Länge der geheimen Zeichen vergrößern.

In dem Fuzzy-Commitment-Verfahren müssen biometrische Merkmale als diskrete Vektoren dargestellt werden. Da viele Merkmale kontinuierliche Daten sind, ist ein Quantisierungsprozess (Binarisierungsprozess) notwendig. Buhan et al. stellen in [BD08] ein Fuzzy-Embedder-Verfahren vor dass eine geheime Zeichenkette direkt in die kontinuierlichen Daten einbetten kann. In diesem Verfahren wird eine Quantisierungs-Index-Modulation (QIM) verwendet, welche ursprünglich in digitalen Wasserzeichen verwendet wurde. Ein biometrisches Merkmal kann zu einem der 2^l Quantisierer zugeordnet werden (l ist die Länge der geheimen Zeichenkette in Bits). Der Abstand zwischen den Merkmalen und dem Quantisierer wird als AD gespeichert. Mit diesen ADkann dieselbe Zeichenkette für die Verifikation wieder hergestellt werden. Einen Kompromiss zwischen

eingebetteter Informationsrate und Zuverlässigkeit des Einbettens kann gemacht werden. Außerdem spielt die räumliche Verteilung der Quantisierer eine wichtige Rolle. Es wurde in dem oben genannten Paper gezeigt, dass sechsfache Unterteilungen hexagonaler Kacheln eine bessere Robustheit und Informationsrate im 2-dimensionalen Merkmalsraum erreichen kann im Vergleich zu vierfachen Unterteilung quadratischer Kacheln.

3.2.2.2 Fuzzy Vault

Der Fuzzy Vault [JS02] ist für ungeordnete biometrische Merkmale mit unterschiedlicher Länge entwickelt worden, wie sie zum Beispiel bei der Verwendung von Fingerabdruckminutien auftreten. Weil sich die Anzahl und die Position der detektierten Minutien verändert, wird Shamir's Secret-Sharing-Protokoll [S79] statt des kryptographischen Hash-Werts verwendet. In Shamir's-Secret-Sharing können mehre Stützstellen für geheime Zeichenketten produziert werden und eine Teilmenge der Stützstellen ist ausreichend, um die Zeichenkette zu rekonstruieren.

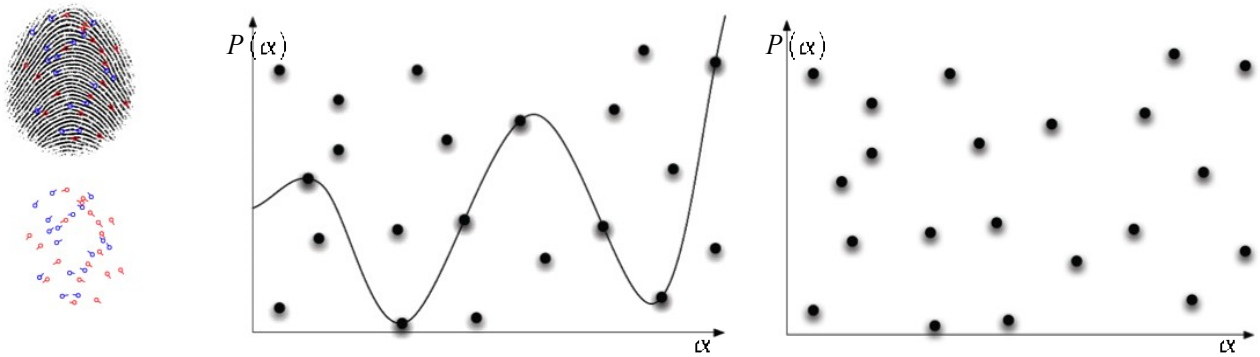

Abbildung 11: Stützpunkte und Steupunkte des Vault Sets

Im Enrolment des Fuzzy-Vault-Verfahrens wird ein Polynom $P(\alpha)$ bestimmter Ordnung *d* generiert. Dessen Koeffizienten entsprechen der geheimen Zeichenkette $S=[s_0, s_1, s_2, \cdots, s_d]$, nämlich $P(\alpha)=s_0+s_1\cdot\alpha+s_2\cdot\alpha^2+\cdots+s_d\cdot\alpha^d$. Die Informationen jeder Minutie des Enrolment-Fingerabdrucks werden als ein Zahl dargestellt. Beispielsweise können die x- und y-Position einer Minutie zu einer 16-Bit langen Zeichenfolgen konvertiert werden, die zu einer Zahl in einem endlichen Körper GF(2^4) korrespondiert. Eine Stützstelle des Polynoms besteht aus der Minutien-Zahl α_M und deren Projektion $P(\alpha_M)$ auf dieses Polynom. Eine große Anzahl von Streupunkten wird zufällig generiert. Diese dienen zur Verschleierung der Stützpunkte. Die Streupunkte und Stützpunkte bilden das „Vault“. Der Hash-Wert der geheimen Zeichenkette ist der pseudonyme Identifikator und wird mit dem Vault Set zusammen gespeichert.

Die linke Kurve in Abbildung 11 stellt ein Polynom dar, deren Form von der geheime Zeichenkette bestimmt wird. Durch die Minutieninformationen, wie in der Kurve gezeigt, werden die α - Werte der Stützpunkte festgelegt. Zahlreiche Streupunkte liegen außerhalb des Polynoms. Ohne Informationen über die echten Minutien ist es kaum möglich, die Stützpunkte zu finden, wie in der rechten Kurve zu sehen ist. Es ist rechnerisch sehr schwer, alle Kombinationen zu testen, um das korrekte Polynom zu finden.

Bei der Verifikation werden die Kandidatenpunkte für Stützpunkte anhand der Minutien aus einem Abfrage-Fingerabdruck gesucht. Da die Ordnung des Polynoms kleiner ist als die Anzahl der

abgebildeten Minutien-Punkte, reicht eine Teilmenge der Minutien-Punkte aus, um das Polynom und damit auch die geheime Zeichenkette zu rekonstruieren. Wenn eine ausreichende Anzahl der echten Stützpunkte gefunden wurde, kann das korrekte Polynom sowie die geheime Zeichenkette rekonstruiert werden. Die Korrektheit der geheimen Zeichenkette kann durch einen Vergleich des gespeicherten Hash-Wertes und des Hash-Wertes der neu berechneten Zeichenkette geprüft werden.

In [JS02] wurde erwähnt, dass die Robustheit des Verfahrens verbessert werden kann, wenn zusätzlich ein Fehlerkorrektur-Code für Polynome in Betracht gezogen wird. In [CKL03] und [YV04] wurden die Parameter und die Erkennungsleistung des Fuzzy-Vault-Verfahrens analysiert. In [UPJ05] wurde eine Lagrange-Interpolation verwendet, um das Polynom zu rekonstruieren. Weiterhin wurde eine CRC-Kodierung benutzt, um die Korrektheit der berechneten geheimen Zeichenkette zu überprüfen. Eine 128-Bit-AES-Verschlüsselung wurde verwendet, um die geheime Zeichenkette zu schützen. In [NJP07] wird gezeigt, dass die Erkennungsleistung dieses Verfahrens mit „Helper Data“ verbessert werden kann. Als „Helper Data“ werden die Punkte auf den Fingerlinien verwendet, die an Stellen hoher Krümmung liegen. Dabei können Sie genutzt werden, um die Ausrichtung der Minutien-Punkte zu verbessern. Zusätzlich wird beim Vergleich der Minutien die nicht-lineare Verzerrung der Minutien berücksichtigt.

3.2.2.3 Minutien als Merkmalsvektor

Die Anzahl und die Position der Minutien ändert sich zwischen unterschiedlichen Aufnahmen. In der letzten Zeit wurden Verfahren entwickelt, die eine Repräsentation der Minutien als Merkmalsvektoren mit fester Länge ermöglichen. Dadurch kann auch das Fuzzy-Commitment für Minutien-basierte Fingerabdruckerkennung verwendet werden. In [XV08] wurde eine revidierte Version der Fourier-Mellin-Transformation benutzt, die die Minutien im Frequenzbereich darstellt. Jede Minutie ist als ein Dirak-Impuls im Spektrumbereich repräsentiert. Das resultierende Frequenzspektrum wird in einem Polar-Logarithmischen Gitter abgetastet. Die neuen Merkmale sind dadurch invariant gegen Rotation, Skalierung und Translation der Minutien. Ein korrelationsbasierter Vergleich kann somit angewendet werden. Zusätzlich zur Position der Minutien können ihre Orientierungen verwendet werden, um die Erkennungsleistung zu verbessern [XR09].

In [VDR09] wurden zwei Verfahren zur Darstellung der Minutien als Vektor vorgestellt. Ein Fingerabdruckbild wird dabei in kleine Gitterbereiche unterteilt. Die Bereiche, die Minutien enthalten, entsprechen einer Eins, sonst besitzt der Bereich den Wert Null. Die Minutieninformationen werden als ein 2-D-Binärbild mit fixer Größe dargestellt. Um mögliches Entfernen, Einfügen und Verschieben der Minutien zu simulieren, werden in den Trainingsphasen Änderungen der Minutien modelliert und die entsprechenden Wahrscheinlichkeit geschätzt. Diese Informationen und das Modell werden in einem LDPC-Kodierungsverfahren verwendet. In der Veröffentlichung wurde gezeigt, dass eine Sicherheit von 17 Bits erreicht werden kann. Bei einem Schwellwert von 35 Minutien wurde eine FRR von 32% und eine FAR von 0,03% erzielt. Die Sicherheit ist aufgrund der hohen Redundanz im konvertierten binären Vektor nicht gut. Im zweiten Verfahren, welches in [VDR09] vorgestellt wurde, wird die folgende Merkmalstransformation verwendet. Die Idee ist, einen gleichförmigen und statistisch unabhängigen binären Merkmalsvektor zu bekommen, welcher mit einem Binary-Symmetric-Channel (BSC) simuliert werden kann. Der BSC hat den Vorteil, dass mit verschiedenen existierenden Kodierungsverfahren eine gute Leistungsfähigkeit erreicht werden kann. Bei der Transformation wird der Minutienmerkmalsraum nach Position und Winkel in zufällig überlappende Würfel unterteilt. Die

Anzahl der Minutien in den unterschiedlichen Würfeln werden gezählt und mittels eines Schwellwerts binarisiert. Durch Überlappungen der Würfel kann die Robustheit verbessert werden, jedoch erhöht dies auch die Anzahl der korrelierten Bits. Eine Ausrichtung der Fingerabdrücke ist zuvor notwendig. Alle Benutzer können die gleiche Würfeleinteilung verwenden. Um die Leistungsfähigkeit weiter zu verbessern, kann jeder Benutzer eigene Würfeleinteilungen verwenden. Die Einteilung kann dann so erfolgen, dass sehr robuste Bits erzeugt werden. In diesem Fall konnte eine Gleichfehlerrate von Null mit der Mitsubishi Electric (MELCO) Datenbank (579 Benutzer und 14 Sample pro Benutzer) erhalten werden, bei einer geheimen Zeichenkette mit der Länge nicht größer als 150 Bit.

3.3 Sicherheitsanalyse

In dem letzten Abschnitt wurden die existierenden Template-Protection-Verfahren vorgestellt. Viele Verfahren wurden bereits erfolgreich in biometrische Systeme integriert. Gute Umsetzbarkeit und Erkennungsleistung wurden gezeigt. Verschiedene Template-Protection-Produkte, beispielsweise von GenKey[6], Securics[7] und PrivID[8], sind auf dem Markt verfügbar. Template Protection dient als Technik zur Erhöhung der Sicherheit biometrischer Systeme. Außer der Erkennungsleistung ist die Sicherheit ein entscheidendes Evaluierungskriterium. Im Folgenden werden wir mögliche Angriffe auf Template-Protection-Verfahren zeigen. Nicht alle Template-Protection-Techniken weisen gute Resistenz gegen diese Angriffe auf. Außerdem werden wir die Sicherheitsanalyse basierend auf rechnerischer Komplexität und Informationstheorie einführen, die sowohl für die Entwicklung als auch für die Evaluierung der Verfahren sehr wichtig ist.

3.3.1 Angriffe gegen Template Protection

Die Angriffe auf Template Protection sind abhängig von den Informationen, die für Angreifer verfügbar sind. Ein gutes Verfahren sollte die potenziellen Sicherheitsrisiken minimieren. Im Folgenden werden die vier wichtigsten Angriffe vorgestellt. Dies sind Brute-Force-Angriff, Verknüpfungs-Angriff, Hill-Climbing-Angriff und Falsch-Akzeptanz-Angriff.

3.3.1.1 Brute-Force-Angriff

In vielen Template Protection-Verfahren sind Pseudonymer Identifikator Hashwert einer geheimer Zeichenkette wie in Tabelle 4 gezeigt. Eine gründliche Suche nach der originalen Zeichenkette, ein sogenannter Brute-Force-Angriff ist möglich. Die Komplexität des Angriff ist 2^l, wobei l die Länge der Zeichenkette darstellt.

In [MMT09] haben Mihailescu et al. einen Brute-Force-Angriff auf Fuzzy-Vault vorgestellt, der auf Polynom-Rekonstruktion basiert. Eine „Vault"-Menge enthält r Streupunkte und t Stützpunkte. Es sind k Stützpunkte erforderlich, um das echte Polynom zu berechnen ($k<t$). Es kann bewiesen werden: Aus allen r zufällig erzeugten Punkten in einem endlichem Körper (Galois-Field) F_q,

6 http://genkeycorp.com/

7 http://www.securics.com/

8 http://www.priv-id.com/

können mindestens $\mu/3 \cdot q^{t-k} \cdot (r/t)^t$ zufällige Polynome gefunden werden, bei der *t* Punkte in der Punktmenge auf dem Polynom liegt. Dabei ist *q* die Anzahl der Elemente in dem endlichen Körper und μ ist ein Faktor zwischen 0 und 1. Ein Angreifer kann alle solche Polynome ausprobieren. Anschließend können die falschen Polynome nach dem Kriterium ausgefiltert werden, ob weitere Punkte in der „Vault"-Menge darauf liegen oder nicht. Die Zeit des Angriff ist durchschnittlich kleiner als $8.0 \cdot (r/t)^k \cdot k \cdot \log^2(k)$. Die Komplexität des Angriff ist vom Verhältnis zwischen der Anzahl der Streupunkte und Stützpunkte abhängig.

3.3.1.2 Verknüpfungs-Angriff

Ähnlich wie beim Verknüpfungsproblem der ungeschützten biometrischen Templates, können bei Template-Protection-Verfahren die geschützte Templates unterschiedlicher Anwendungen kombiniert werden. Durch einen Verknüpfungs-Angriff können mehr Informationen über geheime Zeichenketten oder über die biometrischen Daten gewonnenen werden. Biometrischen Kryptosysteme sind dafür anfällig. In dem Fuzzy-Vault-Verfahren [SB07] kann dies ein großes Sicherheitsproblem verursachen. In einem Vault werden die echten Stützpunkte, die Informationen über die Minutien enthalten, in zahlreichen Streupunkten versteckt. Bedauerlicherweise können die echten Punkte einfach durch die Überlappung von zwei Vault-Mengen wieder aufgefunden werden, wenn ein Angreifer Zugriff zu den sicheren Templates eines Benutzers aus verschiedenen Enrolments hat. Es ist möglich, Stützpunkte des gleichen Benutzers zu erkennen und die geheime Zeichenkette zu berechnen. Zusätzlich kann ein Angreifer mit seinen eigenen biometrischen Daten gültige Punkte konstruieren und so den Eintrag in der Datenbank manipulieren. In diesem Fall wäre die Authentisierung sowohl für den authentischen Benutzer als auch für den Angreifer zulässig. Diese Angriffe zu detektieren ist nicht möglich.

Das Fuzzy-Commitment ist auch durch diese Angriffe gefährdet. Der Grund dafür ist, dass die Hilfsdaten (AD) des geschützten Templates benutzerspezifische Informationen enthalten. Zum Beispiel können BCH-Codes verwendet werden, wobei die mögliche Codewort-Länge 2^n-1 (*n* ist ein natürliche Zahl) ist. Für ein biometrische Merkmal mit 600 Bits, können maximal 511 Bits benutzt werden.. Nur die stabilsten Bits werden selektiert. Deren Position wird notiert und gespeichert. Weil die Merkmalsauswahl von den statistischen Eigenschaften des Individuums abhängig ist, gibt es eine Kopplung zwischen den gespeicherten Hilfsdaten in den unterschiedlichen Anwendungen [ZWBK09]. Zusätzlich können die Hilfsdaten wesentliche Informationen über die biometrischen Daten enthalten. Die Kombination von Hilfsdaten aus unterschiedlichen Anwendungen kann den Aufwand reduzieren, der notwendig ist, um die geheime Zeichenkette oder die biometrischen Daten zu schätzen [STP09]. Als ein Beispiel werden zwei geschützte Templates *[W1, h(S1)]* und *[W2, h(S2)]* aus den biometrischen Daten *X* von einem Benutzer generiert, wobei *W1* und *W2* die zwei Helper Data sind, *S1* und *S2* sind die geheime Zeichenketten, $W1 = S1 \bullet G1 + X$, $W2 = S2 \bullet G2 + X$, *G1* und *G2* die Generator-Matrizen des Fehlerkorrektur-Enkoders. $W1 + W2 = S1 \bullet G1 + S2 \bullet G2$. *W1* und *W2* sind bekannt. Wenn unterschiedliche Fehlerkorrekturkodierer verwendet werden und $G1 \neq G2$ ist, kann es vorkommen, dass diese Gleichungen genau eine Lösung haben. Im schlimmsten Fall können dadurch die geheimen Zeichenketten aus den beiden Anwendungen sowie die Merkmal-Vektor ermittelt werden. Meistens sind die Enrolment-Daten des gleichen Benutzers in unterschiedlichen Anwendungen nicht identisch. Deswegen können die geheimen Zeichenketten meist nicht genau ermittelt werden. Trotzdem können so über den Verknüpfungsangriff viele Informationen über die geheimen Zeichenketten und die biometrischen Daten gewonnen werden.

3.3.1.3 Hill-Climbing-Angriff

Der Hill-Climbing-Angriff ist eine Methode, die die Effizienz der Suche erhöht. In biometrischen Systemen wird häufig ein Vergleichswert der Ähnlichkeit von gespeicherter Referenz und Probe ausgegeben. Während des Hill-Climbing-Angriffs wird der Score-Wert ausgenutzt, um die Probe rekursiv zu verändern und die Ähnlichkeit zwischen Referenz und Probe zu erhöhen. Bei dem biometrischen Kryptosystem verwendet der Komparator die exakte Übereinstimmung des Hashwerts der geheimen Zeichenkette mit dem Hashwert des neu ermittelten Werts. Nur eine eindeutige Entscheidung kann getroffen werden und keine Ähnlichkeitswerte sind verfügbar. Ein Hill-Climbing-Angriff ist somit unmöglich.

Allerdings werden die biometrischen Merkmale bei der Biometric-Encryption-Methode durch die Multiplikation mit einem Zufallsmuster randomisiert, wodurch die originalen biometrischen Informationen immer noch im randomisierten Bild versteckt sind. Ein quantisiertes Hill-Climbing kann eingesetzt werden, um es anzugreifen. Bei Biometric Encryption ist kein Ähnlichkeitswert direkt vorhanden. Aber aus den Informationen, die in der Lookup-Tabelle gespeichert ist, kann ein quantisierter Ähnlichkeitswert erhalten werden. Während jeder Iteration werden Modifikationen nicht global sondern lokal angewendet, sodass die Änderungen ausreichende Verbesserungen des (quantisierten) Ähnlichkeitswertes bewirken. In [Adler05] ist ein Beispiel eines quantisierten Hill-Climbing-Angriffs für Gesichtsbilder gegeben. Eine kleine Gesichtsdatenbank wurde erstellt und Eigenfaces der Bilder wurden berechnet. Ein Initialgesicht wird ausgesucht und in 4 Quadranten unterteilt. Einem Quadranten wird Rauschen hinzugefügt, während der entgegengesetzte Quadrant im Eigenface-Raum leicht variiert wird, sodass sich der Ähnlichkeitswert mindestens um einen quantisierten Schritt erhöht. Die experimentellen Ergebnisse zeigen, dass für ein zufällig ausgewähltes Initialbild eine ausreichende Ähnlichkeit zum Zielbild erhalten werden kann.

Bei der Cancelable Biometrie basiert der Vergleich auch auf Ähnlichkeit. Theoretisch wäre ein Hill-Climbing-Angriff möglich. Allerdings könnte seine Realisierbarkeit durch die Verwendung der nicht-invertierbaren Funktion beschränkt werden.

3.3.1.4 Falsch-Akzeptanz-Angriff

Der Falsch-Akzeptanz-Angriff [ZWBK09] ist in allen biometrische Systeme mit oder ohne Template Protection durchführbar. Aufgrund der Variation der Aufnahmen eines Benutzers und der Ähnlichkeit zwischen Benutzern, treten falsche Akzeptanzen auf. Für ein biometrisches System mit einer Falsch-Akzeptanz-Rate von 0.01% bei gegebenen Systemeinstellungen, bedeutet dies, dass zwei unterschiedliche Benutzer mit „identischen“ biometrischen Merkmalen gefunden werden können, wenn 10^4 Interclass-Vergleiche durchgeführt werden. Ein Angreifer, der Zugang zu einer großen Datenbank hat, kann diesen Angriff durchführen. Die Ähnlichkeit der biometrischen Daten unterschiedlicher Benutzer ist eine inhärente Eigenschaft der Biometrie. Die Ähnlichkeit von genotypischen biometrischen Modalitäten wie Gesichtern kann bei Benutzern beobachtet werden, die zum Beispiel identische Gene (wie bei eineiigen Zwillingen) oder sehr ähnliche Gene (wie Kinder und Eltern) besitzen. Außerdem können auch genetisch bedingte Krankheiten die genotypischen Modalitäten beeinflussen und die Ähnlichkeit vergrößern. Personen mit Down Syndrom haben ähnliche Gesichtszüge wie hypoplastische Nasenknochen und flache Nasenrücken. Die Falsch-Akzeptanz-Rate entspricht der „Minimum-Entropie“ der Sicherheit biometrischer Systeme. Die Falsch-Akzeptanz-Angriffe zeigen die „Worst-Case“- Szenarios von Template-Protection-Verfahren. Kein Schutzmechanismus kann diese Angriffe verhindern. Jedoch sind diese

Angriffe nur mit sehr großen Datenbanken möglich, die Angreifern normalweise nicht zur Verfügung stehen.

3.3.2 Sicherheitsevaluierung

Template-Protection-Verfahren sollten möglichst wenige Sicherheitslücken haben. Sogar in den Fällen, in denen die Informationen über das System und das Verfahren teilweise oder vollständig für Angreifer zugänglich sind, sollten keine großen Sicherheitsprobleme existieren. Um die Sicherheit zu evaluieren, wird zuerst ein Angriffsmodell definiert, welche Informationen ein Angreifer aus den Schwachstellen der Systeme und Verfahren bekommen kann. Generell kann das Angriffsmodul nach Kerckhoffs' Prinzip definiert werden. Dieses besagt, dass die verwendeten Verfahren für die Öffentlichkeit und somit auch den Angreifer verfügbar sind. In den meisten existierenden Arbeiten sind die Sicherheitsevaluierungen der implementierten Template-Protection-Verfahren relativ oberflächlich und überschätzen die Sicherheit der Verfahren deutlich. Für viele vorgeschlagene Template-Protection-Verfahren sind bisher keine fundierten quantitativen Sicherheitsanalysen veröffentlicht worden. Dies gilt insbesondere für die Verfahren mit Merkmals-Transformation (biometrische Verschlüsselung, Biohashing, Cancelable Biometrie). Im Folgenden werden wir die Voraussetzungen für ein erfolgreiches Schutzverfahren sowie die Kriterien für einer quantitative Sicherheitsanalyse beschreiben.

Theoretisch können starke Schlüssel (geheime Zeichenketten) aus den reproduzierbar verrauschten Daten ausreichender Entropie extrahiert werden [DRS04]. Die Menge der extrahierbaren Informationen ist nicht größer als die Min-Entropie der biometrischen Daten. Min-Entropie ist definiert als Entropie der maximale Wahrscheinlichkeit der Zufallsvariable A, nämlich $\hat{H}(A) = H(max_a P(A=a))$. Die Average-Min-Entropie ist die Entropie des Erwartungswerts der maximalen bedingten Wahrscheinlichkeit der Zufallsvariable A bei gegebene Variable B $\tilde{H}_\infty(A|B) = H(E_b[max_a P(A=a|B=b)])$ [DORS07]. Min-Entropie und Average-Min-Entropie deuten die Worst-Case-Präditionsfähigkeit an. Wenn die Verteilung biometrischer Daten geschätzt werden kann, kann diese verwendet werden, um die Sicherheit eines Verfahrens zu bewerten

Mittelweile werden öffentliche unterstützende Daten (AD) erzeugt, die keine Informationen über den Schlüssel preisgeben. Die Sicherheit und die Machbarkeit der biometrischen Kryptosysteme bauen auf diesem Schema auf. Hinsichtlich des Datenschutzes ist jedoch eine teilweise Preisegabe der biometrischen Daten unvermeidbar, wie in den Doktorarbeiten von Smith und Ignatenko [Smith04][TI09] gezeigt wurde. In [TI09] wurde die Verlustrate biometrischer Daten als eine Funktion der Geheimschlüsselrate für generelle biometrische Kryptosysteme dargestellt. Die Verlustrate ist die Transinformation zwischen biometrischer Daten und AD pro Bit und die Geheimschlüsselrate ist die Informationsrate, die pro Bit übertragen werden kann. Es wurde bewiesen, dass die maximale Geheimschlüsselrate nicht größer als die Transsinformation zwischen den Enrolment- und Probe-Daten ist. Als eine praktische Realisierung dieses Schemas wird ein Exklusives Oder zwischen biometrischer Daten und Codewort im Fuzzy-Commitment verwendet. Ignatenko zeigt, dass der Abgleich zwischen Verlustrate biometrischer Daten und Geheimschlüsselrate nur optimal ist, wenn die maximale Geheimschlüsselrate verwendet wird. Erstaunlicherweise vergrößert sich die Verlustsrate biometrischer Daten mit reduzierter Geheimschlüsselrate. Die Autoren zeigten ebenfalls, dass die Offenlegungsrate biometrischer Daten Linkage-Angriffe ermöglicht. Im Bezug auf die Offenlegungsrate und angenommener Geheimschlüsselrate ist das Fuzzy-Commitment nur optimal, wenn die Geheimschlüsselrate maximal ist. Die Offenlegungsrate ist monoton wachsend mit abnehmender Geheimschlüsselrate.

Außerdem benötigt das Fuzzy-Commitment statistisch unabhängige, gleichmäßig verteilte Daten als Eingabedaten. Nur in diesem Fall kann das Exponieren des geheimen Schlüssels in den AD vernachlässigt werden [Smith04][TI09][LB08], da dies sonst zu Sicherheitsproblemen führen kann In vielen Publikationen über das Fuzzy-Commitment wurde die Länge der geheimen Zeichenkette als Sicherheitsmaß verwendet. Jedoch ist dieses Merkmal nur gültig für statistisch unabhängige, gleichmäßig verteilte Daten. Die statistische Verteilung der biometrischen Daten wird jedoch nur selten analysiert und kann für die meisten Merkmale ohnehin nur grob geschätzt werden. In [LB08] wurde auch gezeigt, dass die Länge der geheimen Zeichenkette zu einer Überschätzung der Sicherheit führen kann, wenn der Angreifer die statistischen Eigenschaften der biometrischen Daten kennt.

Die Sicherheit der Transformation-basierten Template-Protection-Verfahren kann nach ihrer Nicht-Invertierbarkeit und Diversität evaluiert werden. Die Nicht-Invertierbarkeit kann als Komplexität definiert werden, um die originalen Merkmale aus gegebenen sicheren Templates zu berechnen. Die Diversität zeigt die Korrelation der sicheren Templates, die vom selben Benutzer extrahiert werden. In solchen Template-Protection-Verfahren werden benutzerspezifische Parameter (Schlüssel) verwendet, um die Transformationsfunktion zu bilden. In dem Fall, dass der Schlüssel kompromittiert ist, sind die biometrische Merkmale und die Geheimzahl in Gefahr. Zum Beispiel können die charakteristischen Informationen der biometrische Merkmale mit den Biohashes und der entsprechenden Projektionsfunktion wiederhergestellt werden [ZK10].

Die Sicherheit der Cancelable Biometrie für Fingerabdrücke kann mit Coverage-Effort-Curve quantifiziert werden [NJ09]. Nagar u. a. [NJ09] gehen davon aus, dass Angreifer die verwendete Transformation sowie die Transformationsparameter kennen. Für die Minutienerkennungsverfahren werden alle möglichen Minutienpositionen aus dem originalen Merkmalsraum mit der bekannten Transformationsfunktion zurück transformiert und daraus die Kandidatenpositionen für die konvertierten Minutien berechnet. Um den Suchprozess zu beschleunigen, wird die Verteilung der Minutien und die Wahrscheinlichkeit der Transformation berücksichtigt. Anschließend wird die Wahrscheinlichkeit der Kandidatenpunkte berechnet. Eine Rangliste der Kandidatenpunkte kann nach deren Wahrscheinlichkeit sortiert und auf eine bestimmte Anzahl einschränkt werden. Der Aufwand (Effort) bezeichnet die durchschnittliche Schätzungskomplexität der Kandidaten (die Entropie der Likelihood-Wahrscheinlichkeit) in der Rangliste in Bits. Die Coverage bezeichnet die Wahrscheinlichkeit, dass die originalen Minutien innerhalb der Rangliste sind. Die experimentellen Ergebnis zeigt, dass die Parameter der Transformationsfunktion einen starken Einfluss auf die Nicht-Invertibarkeit haben. Bei allen im Paper untersuchten Einstellungen erreichte die 100% Coverage mit einem Effort kleiner als 1 Bits pro Minutie.

3.4 Zusammenfassung

In diesem Kapitel wurde ein Überblick über existierende Template-Protection-Verfahren gegeben. Nach den unterschiedlichen verwendeten Schutzmechanismen können diese Verfahren als Transformation oder biometrisches Kryptosystem klassifiziert werden. In den Transformation-Verfahren sowie der biometrischen Verschlüsselung, des Biohashings und der Cancelable Biometrie werden originale biometrische Daten mit benutzerspezifischen Parametern stark verzerrt oder zufällig geordnet. Die biometrischen Daten werden so stark verändert, dass die originalen Daten nicht mehr erkennbar sind. Im Gegensatz dazu, extrahieren biometrisch Kryptosysteme, wie Fuzzy-Commitment oder Fuzzy-Vault, eindeutige geheime Zeichenketten zuverlässig aus den korrelierten biometrische Daten. Eine eindeutige geheime Zeichenkette wird verschlüsselt und als

Pseudoidentität gespeichert. Die unterstützenden Daten (AD) sind notwendig, um einerseits keine Auskunft über die Pseudoidentität zu geben und andererseits die Variation der biometrischer Daten zu kompensieren. Eine Übersicht über die existierenden Verfahren wird in Tabelle 2 gegeben.

Um die Sicherheit zu analysieren, wurden mögliche Angriffe auf existierende Verfahren recherchiert. Es wurde gezeigt, dass manche Verfahren doch Sicherheitsschwächen haben und anfällig auf verschiedene Angriffe sind. Die am häufigsten vorkommenden Angriffe auf Template–Protection-Verfahren sind Brute-Force-, Verknüpfungs-, Hill-Climbing-, und Falsch-Akzeptanz-Angriff. Außerdem wurden die Risiken untersucht, welche Informationen der originalen biometrischen Daten zurück gewonnen werden können, wenn Details der Template-Protection-Verfahren für Angreifer verfügbar sind. Eine Preisgabe biometrischer Daten kann zusätzliche Sicherheitsprobleme durch unterstützende Daten verursachen. Um die Sicherheit der Template-Protection-Verfahren genau zu evaluieren, sind Angriffsmodelle notwendig sowie ein standardisiertes Evaluierungsframework.

Klasse	***Subklasse***	***Methode***	***Detail***
Transformation-verfahren	Nicht-invertierbare Funktion	Cancelable biometrics [RCCB07]	Transformation der Minutien von Fingerabdrücken
		Feature transformation [YB09]	Transformation der Minutien von Fingerabdrücken
	Biometrisches Salting	Biometric encryption [SRS98]	Änderung des Fingerabdruckbilds
Biometrische Kryptosystem	ECC+XOR	Fuzzy commitment [JW99], [Kor08]	Schemabeschreibung
		Fuzzy extractors [DRS04]	Mathematische Definition für Geheimnisgeneration aus rauschenden Daten
		Shielding functions [LT03]	Schemabeschreibung+Demo für Gauß-Modell
		Helper data systems [TAK05]	Detaillierte Beschreibung+Implementierung der Fingerabdruckerkennung (Fingerlinien)
	ECC+Secret-Sharing	Fuzzy vault [JS02], [NJP07]	Minutien von Fingerabdrücken
	Verschlüsselte Templates	Extended PIR [BCPT07]	Keine Schutz der Templates, aber Verifikation durch Privacy-Protected- Protokoll

Tabelle 2: Überblick der existierenden Template-Protection-Klasse und -Verfahren

4 Qualität der Feature-Extraktions-Algorithmen

4.1 Motivation

Viele der betriebenen biometrischen Systeme erfordern ein kompaktes Speichern biometrischer Referenzen. Eine biometrische Referenz sollte eine getreue Repräsentation eines biometrisches Charakteristikums (z.B. Fingerabdruck) darstellen und einem standardisierten interoperablen Format entsprechen, um eine Offenheit des Systems zu realisieren und einen Austausch von Referenzdaten aber auch den Austausch von Software-Komponenten zu gewährleisten.

Somit ist zu berücksichtigen, dass gegebenenfalls während der biometrischen Verifikation ein anderer Feature-Extraktions-Algorithmus zum Einsatz kommt, als bei der Enrolmentphase verwendet wurde. Im Falle von Fingerabdruck-Erkennungssystemen bietet kompaktes Kodieren von Minutien-Daten Interoperabilität zwischen Systemen unabhängig von der Frage, ob die biometrische Referenz auf dezentralen Servern oder auf einem Token mit limitiertem Speicherplatz gespeichert wird [bus09]. Beispiele solcher Systeme stellen die europäische Identitätskarte „European Citizen Card“ [ecc07] oder die US-amerikanische „Personal Identitiy Verification Card“ [nist07] dar. Für beinahe alle derzeit existierenden Fingerabdruck-Vergleichssubsysteme stellen Lokation, Typ (Papillarleistenende bzw. Gabelung einer Papillarleiste) und Richtung von Fingerabdruck-Minutien die relevanten Features eines Fingerabdruck-Minutien-Templates dar.

Da unterschiedliche Systemhersteller verschiedene Konzepte und Algorithmen zur Identifikation von Minutien-Lokationen, -Typen und -Richtungen verwenden, stimmen diese im Allgemeinen nicht vollständig mit der wahren Lokation der Minutien-Landmarke überein sondern sind um diese herum gestreut. Eine Überprüfung, ob eine kompakter Minutien-Datensatz eine verlässliche Repräsentation des ursprünglichen Bildes darstellt, wird nach ISO/IEC IS 29109-1 [ISOq] als semantischer Konformitätstests verstanden und stellt einen notwendigen und wichtigen Schritt im Entwicklungsprozess eines Minutien-Extraktors dar, um hinreichende Interoperabilität und eine akzeptable Gesamtperformance zwischen unterschiedlichen Implementierungen gewährleisten zu können. ISO/IEC IS 29109-1 [ISOq] kategorisiert Konformitätstests in drei Ebenen (Level):

- Level 1 bezieht sich auf grundlegende syntaktische Tests der Datenfelder. Hierbei wird eine feldweise Überprüfung der Daten auf Byte-Ebene durchgeführt, um eine Konformität zum definierten Interchange Record[9] in Bezug auf Anzahl und Art der Felder sowie Wertebereiche der übermittelten Attribute zu gewährleisten.
- Level 2 stellt einen syntaktischen Test dar, der die Inhalte der Datenfelder auf Plausibilität und Konsistenz prüft. Hierbei wird die Konsistenz zwischen Datenfeldern eines nterchange Record überprüft.
- Level 3 beschreibt einen semantischen Test, der die Güte einer erzeugten biometrischen Referenz in Bezug auf die zu Grunde liegenden biometrischen Daten (z.B. das Bild eines Fingerabdrucks) innerhalb spezifizierter Toleranzen bewertet (siehe auch [bus09]).

Level 3 Konformitätstests stellen somit einen bedeutenden Prozessschritt dar, da ohne eine akkurate Repräsentation biometrischer Daten die gewünschte Qualität der Featureextraktionsalgrithmen und

9 Ein Interchange Record ist ein standardisierte Datenaustauschformat, in dem biometrische Daten wie beispielsweise Minutienkoordinaten und zugehörende Metainformationen (z.B. Minutientypen) gespeichert werden.

somit die erforderliche Interoperabilität und biometrische Performanz des Systems nicht erreicht werden kann. Dieser Abschnitt beschäftigt sich daher im Weiteren mit Level 3 Konformitätstests für Fingerabdruck-Minutien-Daten und wendet diese Verfahren auf drei im Projekt ausgewählte Algorithmen an. Die grundlegende Idee des hier angewandten Verfahrens ist in [bus09] beschrieben. Dort wird eine neue Methodik für Level 3 Konformitätstests vorgeschlagen. Zur Durchführung eines solchen Konformitätstest ist ein Clustering von Minutien notwendig. Dazu wurde ein Verfahren in [Lod09]vorgestellt, das in der Untersuchung in diesem Arbeitspaket verwendet wurde.

4.2 Herausforderungen in der Minutien-Detektion

Bei der Anwendung von Feature-Extraktions-Algorithmen (FE-Algorithmen) auf Bilder von Fingerabdrücken können die folgenden Fälle auftreten, die eine Herausforderung an das Vergleichssubsystem stellen:

4.2.1 Nicht vorhandene Platzierung einer Minutie

Der FE-Algorithmus hat im regionalen Bereich der wahren Koordinate einer Minutien-Landmarke keine Minutie detektieren können.

4.2.2 Unpräzise platzierte Minutien

Der FE-Algorithmus hat eine Minutien-Landmarke zwar detektiert, deren Attribute jedoch ungenau bestimmt. Eine unpräzise Detektion von Minutien kann mit folgenden Defiziten verbunden sein:

- ungenaue Positionierung von Minutien
- Zuordnung eines falschen Minutien-Typs
- ungenaue Bestimmung der Richtung
- unzureichende Qualität von Minutien

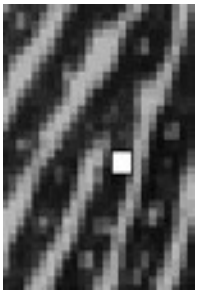

Abbildung 12: Fehlerhaft identifizierter Minuten-Typ: Papillarlinienende anstatt Gabelung.

Ein häufig auftretender Fehler ist die Zuordnung eines falschen Minutien-Typs (vgl. Abbildung 18). Hierbei kann es sich um die Detektion eines tatsächlichen Endes einer Papillarlinie als vermeintliche Gabelung einer Papillarlinie handeln oder umgekehrt. Es gibt jedoch auch Hersteller, die bewusst keine geeignete Typisierung der Minutien vornehmen.

4.2.3 Falsche Minutien-Detektion innerhalb des Fingerabdruckbereichs

Fälschlicherweise automatisch detektierte Minutien können sich in diversen problematischen Bereichen befinden (vgl. Abbildung 13, Rechtecke markieren Papillarlinienenden, Kreuze markieren Gabelungen von Papillarlinien; FE-Algorithmus: NIST mindtct.):

- vernarbtes Gewebe und Hautfurchen
- Papillarpunkte
- sich auf dem Finger befindende Verunreinigungen (Schmutz, Haare)
- Hautkrankheiten (z.B. Ekzeme, Tuberkel)
- Text, Zeichnungen innerhalb des Fingerabdruckbereichs (sofern es sich um eingescannte Bilder von analogen Fingerabdrücken handelt (ink-prints).

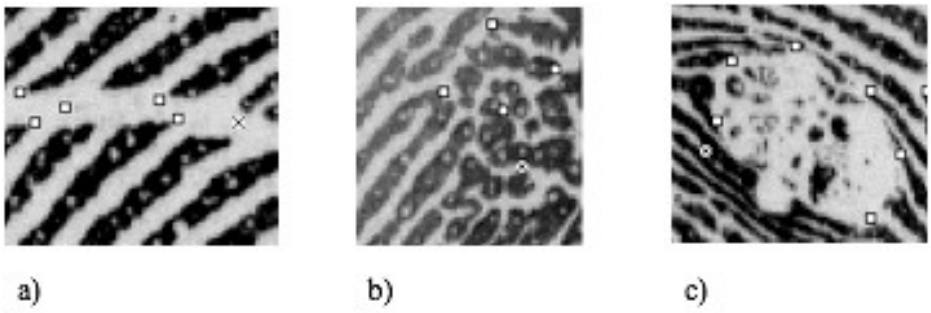

Abbildung 13: Minutien in problematischen Bereichen innerhalb des Fingerabdruckbereichs: a) Hautfurche, b) Papillarpunkte, c) Tuberkel

4.2.4 Falsche Minutien-Detektion außerhalb des Fingerabdruckbereichs

Einige Minutien-Extraktions-Algorithmen erkennen fälschlicher Weise Minutien am Rande oder außerhalb des Fingerabdruckbereichs. Dieser Fehler resultiert in der Regel aus ungeeignetem Maskieren des Vorder- bzw. Hintergrunds und kann durch sich im Hintergrund befindenden Schmutz, Zeichen oder Zeichnungen hervorgerufen werden. Bei Bildern, die aus analogen Fingerabdruckbildern erstellt wurden, tritt dies häufig auf, da sich im Hintergrund gedruckte oder handschriftliche Annotationen zur Fingerabdruckskarte befinden. Abbildung 14 (a) zeigt den Ausschnitt einer aufgedruckten Information im Hintergrundbereich und eine falsch erkannte

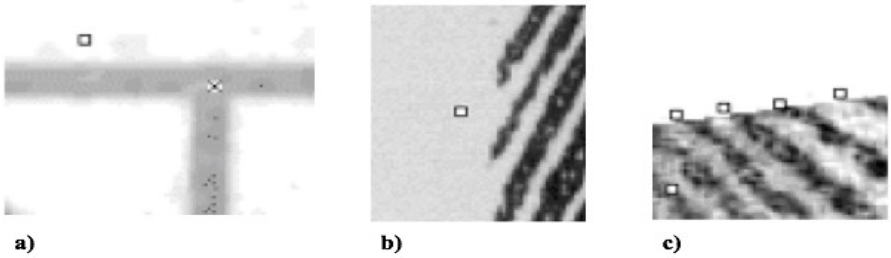

Abbildung 14: Fehlerhaft erkannte Minutien: a) außerhalb des Fingerabdruckbereichs, b) und c) am Rand des Fingerabdruckbereichs.

Minutie vom Typ Papillarlinienende sowie eine weitere Minutie, die am Aufdruck eine Gabelung einer Papillarlinie vermutet.

4.3 Methode zur Durchführung semantischer Konformitätstests

Dieser Abschnitt beschreibt die drei Konformitätsraten, wie sie in [bus09] und [Lod09] entwickelt wurden. Diese Raten ermöglichen es, Aussagen über die Konformität eines FE-Algorithmus auf Basis eines Ground-Truth Datensatzes (GTM) zu treffen. Bei den zum Ground-Truth Datensatz (siehe Abschnitt 4.4) gehörenden Minutien handelt es sich um Cluster-Zentren, die sich nach Anwendung des in diesem Abschnitt beschriebenen Clustering-Algorithmus auf von unterschiedlichen Experten manuell platzierten Minutien ergeben haben.

Die erste Konformitätsrate cr_{gtm} beschreibt, in welchem Rahmen die von einem FE-Algorithmus automatisch platzierte Minutien (AGM) in der Nachbarschaft der Ground-Truth Daten (wahre Koordinate der Landmarke) platziert werden. Die Konformitätsrate cr_{gtm} nimmt den Wert 0 an, wenn keine der automatisch generierten Minutien innerhalb eines gewissen Toleranzbereiches der in der GTM enthaltenden Minutien liegen. Andernfalls berechnet sich die Konformitätsrate cr_{gtm} wie folgt:

$$cr_{gtm} = \frac{\sum_{i=1}^{ngtm} mcs_i}{ngtm}$$

Hierbei bezeichnet *ngtm* die Anzahl der im Ground-Truth Datensatz enthaltenen Minutien und mcs_i die für eine *i*-te Minutie spezifische Minutien-Konformitätsrate. Die Minutien-Konformitätsrate mcs_i berechnet sich nach:

$$mcs_i = \begin{cases} 0 & wenn\ d \geq tol_d \\ 1-p & sonst \end{cases} \quad \text{mit} \quad tol_d = \frac{W}{4}$$

Hierbei bezeichnet *d* den euklidischen Abstand zwischen einer sich in der GTM befindenden Minutie und der nächsten automatisch erzeugten Minutie und *W* beschreibt den Abstand zwischen zwei parallelen Papillarlinien. Des Toleranzbereiches tol_d wird mit W/4 derart gewählt, sodass sich es dabei um den maximal möglichen Radius eines Kreises um eine GTM handelt, wobei für zwei benachbarte GTM keine Überlappung entsteht. Abbildung 26 verdeutlicht dies.

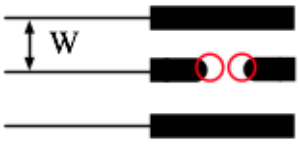

Abbildung 15: Abstand W zwischen Papillarlinien und Minutien-Toleranzbereich.

Die Minutien-Konformitätsrate *mcs* einer Minutie reduziert sich gegebenenfalls um den Bestrafungsfaktor *p,* wenn die Minutienattribute (Richtung und/oder Minutientyp) nicht übereinstimmen. In diesem Fall ergibt sich der Bestrafungsfaktor *p* auf Grund von Richtungsunterschieden $p_{\Delta\Theta}$ oder abweichenden Minutien-Typen $p_{\Delta t}$ und wird wie folgt definiert:

$$p = p_{\Delta\Theta} + p_{\Delta t} \quad \text{mit} \quad p_{\Delta\Theta} = \frac{|\Theta_{gtm} - \Theta_{agm}| * 0.5}{\pi} \quad \text{und} \quad p_{\Delta t} = \begin{cases} 0{,}25 & wenn\, t_{gtm} \neq t_{agm} \\ 0 & sonst \end{cases} .$$

Die Wahl unterschiedlicher Bestrafungsfaktoren für unterschiedliche Attributsfehler ist hierbei beabsichtigt, da Ungenauigkeiten in der Bestimmung der Minutien-Lokation den stärksten Einfluss auf die Interoperabilität und gewünschte Performanz haben, gefolgt von Ungenauigkeiten in der Bestimmung des Minutien-Winkels und in der Identifizierung des Minutien-Typs.

FE-Algorithmen machen häufig Fehler bei der Bestimmung des Minutien-Typs. Dies verursacht Seiteneffekte und führt zu einem Winkel $\Delta\Theta$ der sich nahe in einer Umgebung von π befindet. Um eine Fehlklassifikation nicht doppelt zu bewerten, wird daher der Winkel der AGM Minutie um π erhöht, falls eine Fehlklassifikation des Minutien-Typs festgestellt wurde.

Die zweite Konformitätsrate cr_{agm} beschreibt den Anteil falsch detektierter Minutien außerhalb oder an den Grenzen des Fingerabdruckbereichs:

$$cr_{agm} = \frac{\sum_{i=1}^{nagm} mps_i}{nagm} \quad \text{mit}$$

$$mps = \begin{cases} 0 & wenn\, die\ AGM - Minutie\,(mit\ Index\ agm)\ außerhalb\ des\ Fingerabdruckbereichs\ liegt \\ 0{,}5 & wenn\, die\ AGM - Minutie\ auf\ der\ Grenzlinie\ des\ Fingerabdruckbereichs\ liegt \\ 1 & sonst \end{cases}$$

und *nagm* als die Anzahl der automatisch generierten Minutien.

Die dritte Konformitätsrate cr_{amf} beschreibt, in welchem Ausmaß sich die automatisch extrahierten Minutien tatsächlich wie erwartet auf den Fingerabdruckbereich konzentrieren. Diese Konformitätsrate kann als derjenige Anteil der automatisch generierten Minutien verstanden werden, für die keine entsprechenden Minutien im Ground-Truth Datensatz gefunden werden können:

$$cr_{amf} = 1 - \frac{niagm}{nagm} \quad ,$$

wobei *niagm* die Anzahl der automatisch erzeugten Minutien entspricht, für die keine entsprechenden Minutien im GTM existieren. Es werden mit dieser Konformitätsrate die Anzahl der falschen Minutien im Fingerabdruckfokus bewertet. Für die Paarzuordnung einer GTM Minutie zu einer AGM Minutie gelten die gleichen Nachbarschaftsregeln wie bei der Konformitätsrate cr_{gtm}.

4.4 Ground-Truth Datensatz

Die Durchführung von Level 3 Konformitätstests basierend auf der hier vorgestellten Methodik erfordert eine als Ground-Truth bezeichnete Datenbank (GTM – Ground-Truth Minutien) aus Referenzdatensätzen, die als Prüfdaten verwendet werden können. Die folgenden Unterkapiteln beschreiben den Prozess der Erstellung eines Ground-Truth Datensatzes sowie Methoden zur Evaluation der Güte des Datensatzes.

4.4.1 Sammlung von Ground-Truth Daten

Die im Projekt verwendete GTM Datenbank besteht aus Fingerabdruck-Minutien, die von daktyloskopischen Experten des Bundeskriminalamts erzeugt wurden. Beim Aufbau der Datenbank wurden Minutien von unterschiedlichen Experten des BKA unabhängig voneinander lokalisiert und parametrisiert. Zur Durchführung wurde den Experten ein grafisches Interface zur Verfügung gestellt, über das Angaben über Lokation, Typ, Winkel und Qualität von Minutien innerhalb von Fingerabdruckbildern erfasst wurden (siehe Abbildung 16). Ferner wurden für zukünftige Verwendungszwecke Informationen über Cores, Deltas, Mustertypen und Signalqualität erfasst. Die von den Experten erfassten Daten werden darüber hinaus mit Konfidenz- bzw. Qualitätswerten versehen, um die Zuverlässigkeit einer einzelnen Parametrisierung (z.B. Minutienqualität) bzw. eine Bewertung der Signalqualität des Gesamtbildes zu erhalten. .

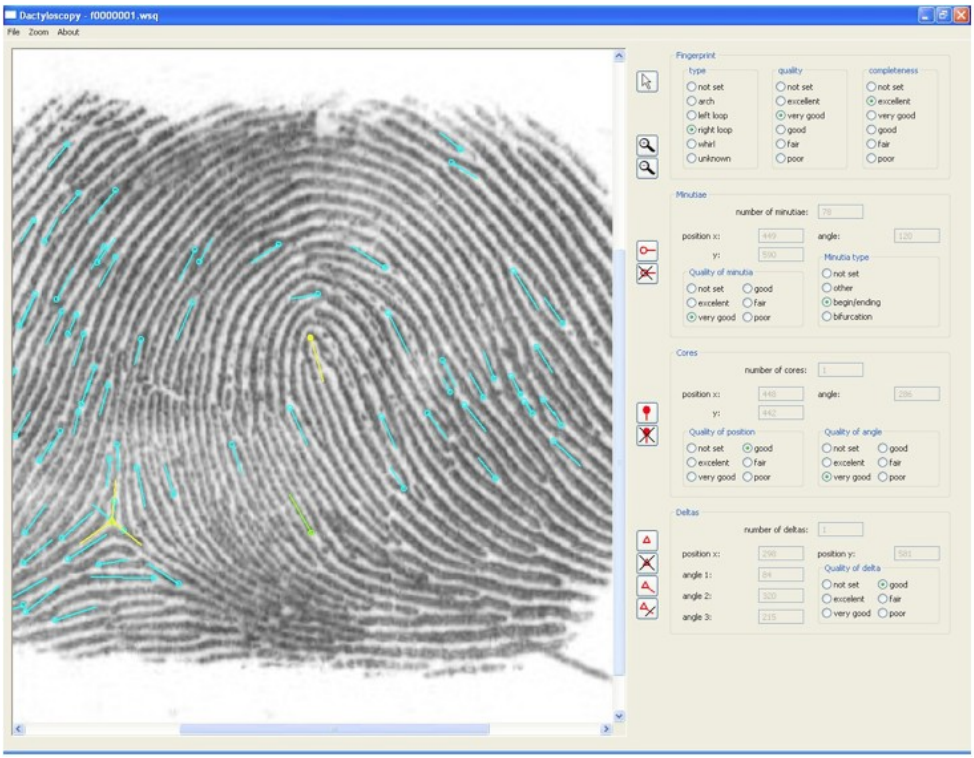

Abbildung 16: Von daktyloskopischen Experten verwendete GUI.

Die hierfür erforderlichen Fingerbilder wurden durch das National Institute of Standards and Technology (NIST) bereitgestellt. Die Bilder wurden aus den NIST Special Databases SD14 (überwiegend gerollte Fingerbilder - überwiegend inked-prints aber auch einige live-scan Bilder) [NIST-SD14] und SD29 (flache Abdrücke aber Durchweg inked-prints) [NIST-SD29] zusammengestellt. Die vom NIST durchgeführte Auswahl sollte systematische Effekte in Bezug auf die Scan-Eigenschaften ausschließen. Fingerbilder sollten wenn möglich in einer ausgewogenen Mischung von männlichen und weiblichen Personen stammen. Weiterhin sollen die ausgewählten Fingerbildpaare die verschiedenen NFIQ-Qualitätslevel repräsentieren.

Durch manuelle Erhebung der Daten unter Verwendung einer eigens hierfür entwickelten grafischen Oberfläche soll jeglicher Einfluss von automatischer oder halbautomatischer Funktionalität eines Automatische Fingerabdruck-Identifikationssystems (AFIS) vermieden werden.

4.4.2 Ground-Truth Minutien Datenformat

Die von den Experten des Bundeskriminalamts erzeugten Informationen werden in einem intern verwendeten *.gtm Dateiformat gespeichert dessen Kodierung dem ISO 19794-2 Standard soweit wie möglich folgt. Hierzu werden die folgenden Datenfelder gespeichert:

1. Grundmuster-Typ

   Die Grundmuster Klassifikation sollte entsprechend der folgenden Klassen-Kodierungen erfolgen: A = Arch, L = Left Loop, R = Right Loop, W = Whirl, U = Unknown.[10]

2. Bewertung der Bildqualität

   Die Bewertung der Bildqualität soll entsprechend der NFIQ-Quantisierungen[11] erfolgen: 1 = exzellent, 2 = sehr gut, 3 = gut, 4 = noch verwendbar, 5 = schlecht.

3. Minutien-Typ (ISO 19794-2 Clause 7.4.2.1)

   Unknown, Papillarlinienende, Gabelung einer Papillarlinie.

4. Minutien-Position (ISO 19794-2 Clause 7.4.2.2)

   Die Bild-Koordinate der Minutie (horizontal X und vertikal Y). Der Ursprung des Koordinatensystems ist wie in ISO/IEC 19794-2 festgelegt die obere linke Ecke. Die Angabe von X und Y Koordinaten erfolgt in Pixeleinheiten.

5. Minutien-Winkel (ISO 19794-2 Clause 7.5.3.1)

   Absoluter Richtungswinkel der Minutie. Der Richtungswinkel wird im Gegenuhrzeigersinn von der rechten Horizontalen gemessen. Der Winkel wird in Grad angegeben.

6. Konfidenzwert für die Minutie (ISO 19794-2 Clause 7.4.2.4)

   Der Konfidenzwert einer Minutie (quality - accuracy) soll ein Messwert für die Position UND den Richtungswinkel sein und rangiert im Bereich von 100 (Maximum) bis 1 (Minimum).

7. Anzahl Cores (ISO 19794-2 Clause 7.5.3.1)

   Die Anzahl der im Fingerbild vorhandenen Cores.

8. Core Position (ISO 19794-2 Clause 7.5.3.3)

   Die Koordinaten der Cores werden bestimmt (horizontal X und vertikal Y).

9. Konfidenzwert für die Core Position (nicht in ISO 19794-2 enthalten)

   Der Konfidenzwert für die Core Position beschreibt nur die Koordinaten-Genauigkeit und rangiert im Bereich von 100 (Maximum) bis 1 (Minimum).

10. Core Richtungwinkel (ISO 19794-2 Clause 7.5.3.4)

    Absoluter Richtungswinkel der Core-Tendenz. Der Richtungswinkel wird im Gegenuhrzeigersinn von der rechten Horizontalen gemessen. Der Winkel wird in Grad angegeben.

10 Ausführungen zu Grundmustern sind zu finden bei Maltoni et al [Mal05]

11 Diese Qualitätsstufen wurden vom Tabassi et al. [Tab04] Definiert. Das Verfahren hat sich als de-facto Standard zur Bewertung der Qualität von Fingerbildern etabliert.

11. Konfidenzwert für Core Richtungswinkel (nicht in ISO 19794-2 enthalten)

    Der Konfidenzwert für den Core Richtungswinkel beschreibt nur die Richtungswinkel-Genauigkeit und rangiert im Bereich von 100 (Maximum) bis 1 (Minimum).

12. Anzahl der Deltas (ISO 19794-2 Clause 7.5.3.5)

    Die Anzahl der im Fingerbild vorhandenen Deltas.

13. Delta Position (ISO 19794-2 Clause 7.5.3.7)

    Die Koordinaten der Delta Position wird angegeben (horizontal X und vertikal Y).

14. Delta Richtungswinkel (ISO 19794-2 Clause 7.5.3.8)

    Die drei Richtungswinkel der Tendenz der Papillarlinien am Delta wird gemessen. Der Richtungswinkel wird im Gegenuhrzeigersinn von der rechten Horizontalen bestimmt. Der Winkel wird in Grad angegeben. Sofern nicht alle drei Richtungswinkel gemessen werden können, da etwa das Bild verrauscht ist oder abgeschnitten wurde, so soll in den Feldern der Richtungswinkel der anderen Tendenzen für das selbe Delta wiederholt werden.

15. Konfidenzwert für das Delta (nicht in ISO 19794-2 enthalten)

    Der Konfidenzwert für ein Delta soll ein Messwert für die Position UND den Richtungswinkel sein und rangiert im Bereich von 100 (Maximum) bis 1 (Minimum).

4.4.3 Daten-Clustering-Methode

Es kann davon ausgegangen werden, dass die von den einzelnen Experten vorgenommene Analyse der Fingerabdruck-Minutien ähnliche Ergebnisse liefern, die jedoch um die wahren Werte der Minutienparameter gestreut sind. Aus diesem Grund ist es notwendig, die von den Experten erzeugten Daten nach zu bearbeiten und zu clustern, bevor sie als Eingabe für den in Abbildung 17 dargestellten Prozess zur Generierung von Konformitätsraten genutzt werden können.

Der erste Verarbeitungsschritt hierbei ist die Analyse von Anhäufungen der von unterschiedlichen Experten erzeugten Minutien innerhalb eines Bildes. Danach wird der Fingerabdruckbereich eines Bildes markiert und der Abstand zwischen Papillarlinien (W) berechnet. Das selbe Bild wird des Weiteren durch den zu testenden Minutien-Extraktions-Algorithmus verarbeitet – in diesem Fall diente der NIST mindtct Algorithmus [nbis] als Beispiel. Die so gewonnenen Informationen dienen als Eingabe zur Bestimmung der Konformitätsraten.

Die Definition eines Cluster-Algorithmus, der die aus der Arbeit zahlreicher Experten vorliegenden Minutien analysiert und die Ground-Truth Minutien generiert, ist nicht-trivial da die Anzahl der vorhandenen Cluster im Voraus nicht bekannt ist. Um dieses Problem zu adressieren, wurde in [Lod09] ein Ansatz entwickelt, der von [wk09] inspiriert wurde. Hierzu werden die von n Experten erzeugten Datensätze zunächst in einem aus Minutien bestehenden Array (in diesem Fall ein `struct` mit Werten zu Position, Winkel, Typ, Qualität, Experten-ID sowie einem booleschen Flag, das Auskunft über den Verarbeitungsstatus gibt) gespeichert. Anschließend wird hieraus ein Array bestehend aus Minutien-Paaren, für jede mögliche paarweise Kombination von Minutien gewonnen, wenn die folgenden Bedingungen erfüllt sind:

- Jede Minutie wurde von einem unterschiedlichen Experten erzeugt.

- Der Abstand zwischen zwei Minutien ist kleiner oder gleich W/2 (d.h. alle Minutien jeweils eines Paares befinden sich innerhalb eines Kreises mit Radius W/4).

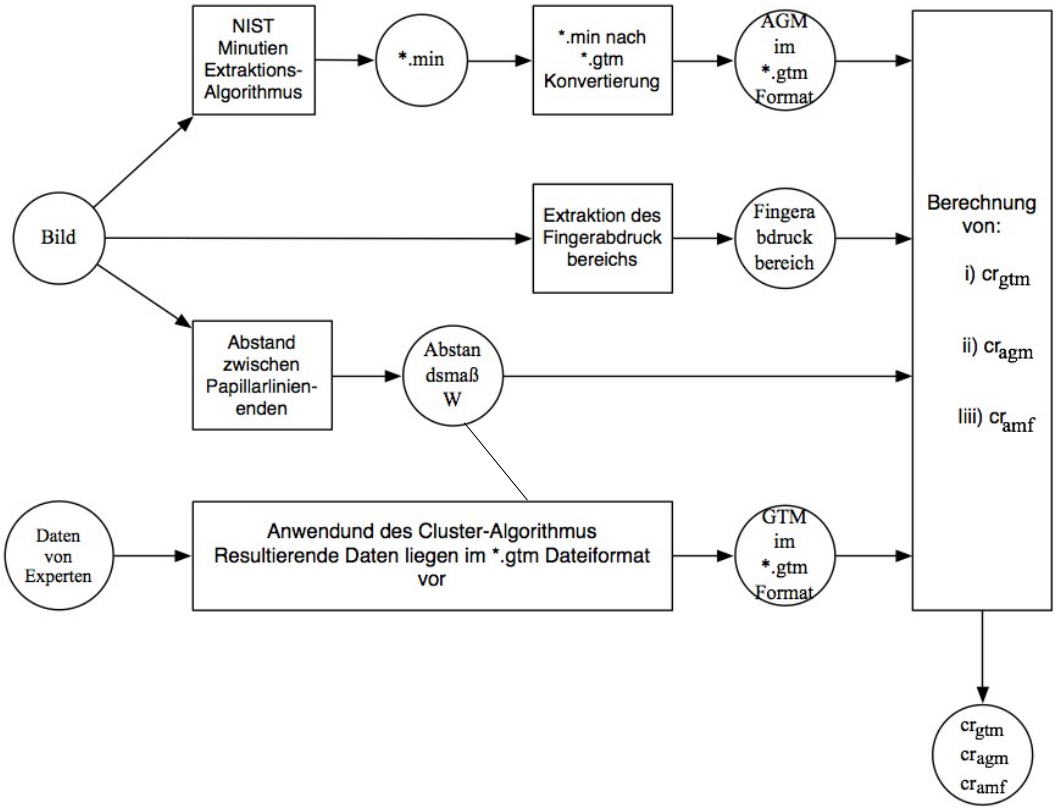

Abbildung 17: Prozessschritte zur Bestimmung von Konformitätsraten. Kreise repräsentieren Dateien bzw. Werte, Rechtecke repräsentieren Softwarekomponenten.

Nachdem auf diese Weise ein Minutien-Paar erzeugt wurde, werden beide Minutien als verarbeitet markiert und das hierdurch gewonnenen Minutien-Paar in einem Minutien-Paar-Array gespeichert.

Analog hierzu wird auf Basis des Minutien-Paar-Arrays ein Array bestehend aus Minutien-Tripel erzeugt. Zwei Minutien-Paare werden zu einem Minutien-Tripel zusammengefasst, wenn die folgenden Bedingungen erfüllt sind:

- Zwei Minutien-Paare verfügen über eine gemeinsame, identische Minutie
- Jede Minutie in einem Minutien-Tripel wurde von einem unterschiedlichen Experten erzeugt
- Der maximale Abstand von Minutien, die dem Tripel-Kandidaten zugeordnet werden ist kleiner oder gleich W/2 (d.h. alle Minutien befinden sich innerhalb eines das Tripel umfassenden Kreises mit Radius W/4)

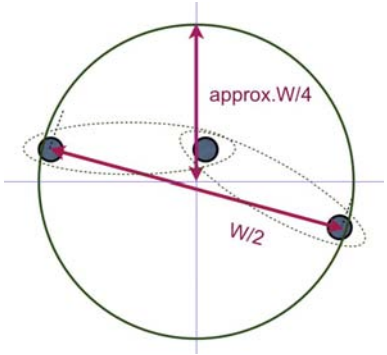

Abbildung 18: Toleranzbereich beim Minutien-Clustering: Zugehörigkeit wird durch maximale Distanz W/2 bedingt.

Tripel werden analog zu Minutien-Paaren erzeugt. Zusätzlich bedarf es beider Minuten-Paare jedoch einer gemeinsamen Minutie, die als Verknüpfung dient (vgl. Abbildung 19). Entsprechend hierzu lassen sich auf Basis der Minuten-Tripel Minutien-Quadrupel erzeugen, sofern die folgenden Bedingungen erfüllt sind:

- Minutien-Tripel verfügen über zwei gemeinsame, identische Minutien (siehe Abbildung 19)
- Jede Minutie in einem Quadrupel-Kandidat wurde von einem unterschiedlichen Experten erzeugt
- Der maximale Abstand aller Minutien innerhalb des Quadripel-Kandidaten ist kleiner oder gleich W/2

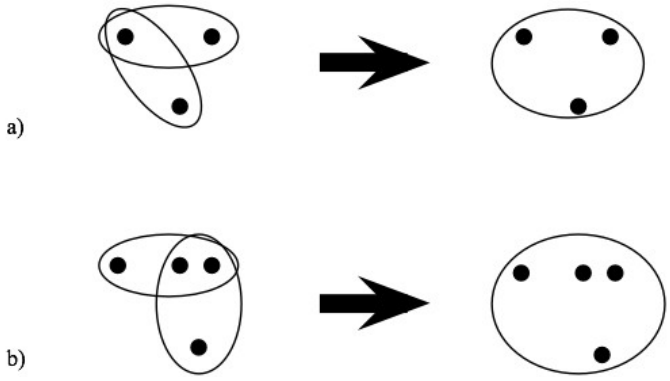

Abbildung 19: Minutien-Clustering: a) Erzeugung eines Minutien-Tripels auf Basis zweier Minutien-Paare, b) Erzeugung eines Minutien-Quadrupels auf Basis zweier Minutien-Tripel.

Dieser Prozess lässt sich bis zur Erzeugung von n-Tupel verallgemeinern, so denn die folgenden Kriterien erfüllt sind:

- (n-1)-Tupel verfügen über (n-2) gemeinsame, identische Minutien
- Jede Minutie in einem n-Tupel Kandidat wurde von einem unterschiedlichen Experten erzeugt
- Der Abstand aller (n-1)-Tupel innerhalb eines n-Tupels ist kleiner oder gleich W/2 (d.h. alle Minutien befinden sich innerhalb eines Kreises mit Radius W/4)

Um im Anschluss das Zentrum eines Clusters identifizieren zu können, wird die durchschnittliche Minutien-Position im Cluster, sowie der durchschnittliche Typ und der durchschnittliche Winkel bestimmt. Die Position des Clusterzentrums erfolgt durch Mittelbildung:

Arithmetischen Mittels: $$X_{GTM} = \frac{\sum_{i=1}^{ngtm} x_i}{ngtm} \quad , \quad Y_{GTM} = \frac{\sum_{i=1}^{ngtm} y_i}{ngtm}$$

Die Auswirkung der Mittelbildung ist in Abbildung 20 dargestellt. Wie leicht zu sehen ist, weist der Ansatz eine Robustheit in Bezug auf Ausreißer auf. Da in diesem Beispiel ausschließlich ein Experte die Minutie als linksseitig klassifizierte und alle anderen Experten die Minutie rechtsseitig klassifizierten, liegt das Zentrum des Clusters näher an der rechten Seite. Der Vorteil dieses Ansatzes liegt darin, dass die Ground-Truth Daten eine Robustheit und Zuverlässigkeit aufweisen, während gleichzeitig die Gefahr des Verwerfens einer automatisch erzeugten Minutie der

Wahrscheinlichkeit entspricht, dass es sich bei der als Ausreißer bewerteten Klassifikation eines einzelnen Experten um die korrekte Klassifikation handelte.

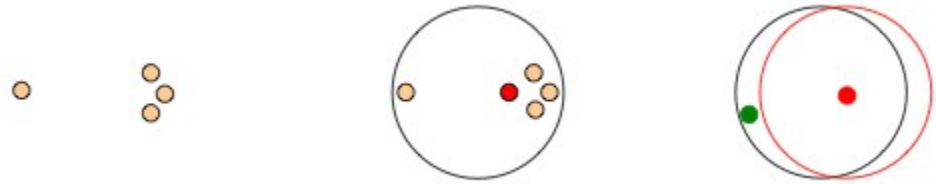

Abbildung 20: Vorgehen bei der Berechnung des Cluster-Zentrums: Ansatz (arithmetisches Mittel) in Abbildungen. Helle Punkte repräsentieren von Experten identifizierte Minuten, rote Punkte repräsentieren die berechneten Zentren der Cluster und der grüner Punk repräsentieren einen Outlier.

Gleichermaßen notwendig ist die Berechnung des durchschnittlichen Minutien-Typs. Die Zuweisung eines expliziten Minutien-Typs in den Ground-Truth Daten erfolgt genau dann, wenn von mehr als 2/3 der Experten der gleiche Minutien-Typ identifiziert wurde. Andernfalls wird der Minutien-Typ auf UKNOWN gesetzt, wodurch das Bestimmen eines Bestrafungsfaktors $p_{\Delta t}$ für einen fehlerhaften Minutien-Typ unmöglich wird. Dieses Vorgehen entspricht ISO Direktiven, in denen eine vorhandene Zweidrittelmehrheit als Konsens betrachtet wird.

Zur Berechnung eines durchschnittlichen Winkels bedarf es der Beachtung unterschiedlicher Fälle. Zum Beispiel kann der Fall eintreten, in dem ein Experte die Richtung einer Minutie als 180° identifiziert, während diese von einem zweiten Experten für die selbe Minutie als 0° bemessen wird. Darüber hinaus könnte es sein, dass drei (oder mehr) Experten zu drei (oder mehr) völlig unterschiedlichen Einschätzungen kommen (z.B. 0°, 120° und 240°). In solch einem Fall ist es angebracht, die Winkel-Angabe in den Ground-Truth Daten auf UNKNOWN zu setzen. Ferner wird der Richtungsparameter auf UNKNOWN gesetzt wenn der Minutien-Typ UNKNOWN ist, da Konsens bei der Bestimmung des Minutien-Typs als Voraussetzung für die Bestimmung des Richtungswerts und zur Generierung zuverlässiger Ground-Truth Daten erachtet wird.

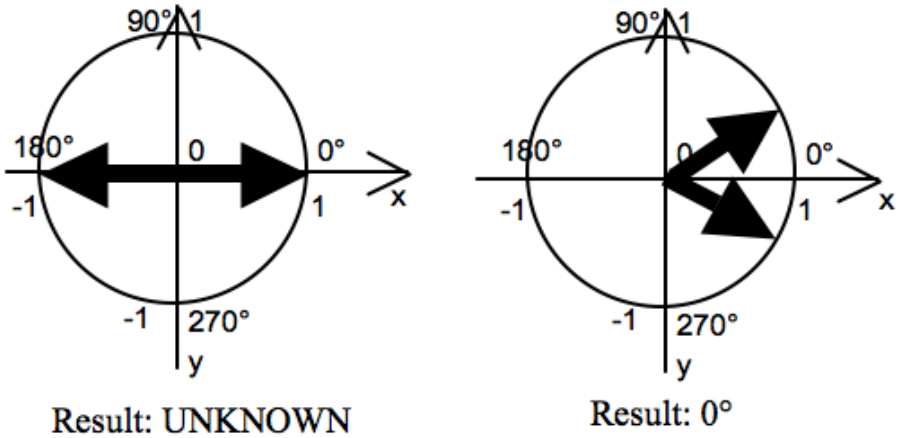

Abbildung 21: Berechnung des durchschnittlichen Winkels.

Ein durchschnittlicher Winkel wird berechnet, indem zunächst alle Winkel in gerichtete Einheitsvektoren konvertiert werden, sodass alle Endpunkte (x_m,y_m) eines Clusters sich auf dem Einheitskreis befinden. Im Anschluss daran werden die mittleren[12] Koordinaten x und y errechnet und als Endpunkte eines daraus resultierenden Vektors angenommen. Die Länge dieses Vektors kann einen Wert kleiner 1 annehmen, wenn die Expertenmeinungen divergieren. Sollte die Länge des Vektors kleiner als 1/3 sein (wie in Abbildung 21 links in einem Beispiel dargestellt) so kann kein Konsens unter den Experten mehr festgestellt werden und die resultierende Richtung wird als UNKNOWN beschrieben. Andernfalls wird der resultierende Vektor in eine entsprechende Richtungsangabe konvertiert und in die Ground-Truth Daten übernommen.

4.4.4 Zuverlässigkeit der Cluster

Zur Berechnung der Konformitätsraten wie in Kapitel 4.3 beschrieben, ist es notwendig, die Zuverlässigkeit jeder im Ground-Truth Datensatz enthaltenen Minutie zu beachten. Diese Zuverlässigkeit hängt wiederum von der Qualität des identifizierten Clusters ab, der die Ground-Truth Minutie erzeugt. Die Qualität eines Clusters letztlich wird von den folgenden zwei Faktoren beeinflusst: einerseits ist die Qualität eines Clusters abhängig von der Anzahl der zur Erzeugung von Minutien-Daten eingesetzten Experten. Wenn ein Bild von zum Beispiel 20 Experten analysiert wurde und lediglich zwei der 20 Experten eine bestimmte Minutie identifiziert haben (und diese möglicherweise darüber hinaus mit einer niedrigen Qualität attributiert haben), kann die daraus resultierende durchschnittliche Minutie, d.h. das Zentrum des Clusters, nicht als zuverlässig betrachtet werden. Andererseits, wenn eine bestimmte Minutie von 18 Experten detektiert (und möglicherweise mit hoher Güte bewertet) wurde, kann von einem zuverlässigen Cluster-Zentrum gesprochen werden. Um dementsprechend unzuverlässige von zuverlässigen Minutien unterscheiden zu können, wird die Qualität eines Clusters wie folgt bestimmt:

$$quality\ of\ cluster\ = \frac{\sum_{i=1}^{ncl} q_i}{nexp},\ \ quality\ of\ cluster \in \{0,..,100\}$$

Hierbei bezeichnet q_i die Qualität der i-ten Minutie des Cluster, *ncl* die Anzahl der Minutien im Cluster und *nexp* die Anzahl der Experten, die das dazugehörige Bild verarbeitet haben. Zur Verdeutlichung: wenn alle Experten eine Minutie mit einer Qualität von 50 erkennen, resultiert dies in einer Cluster-Qualität von ebenfalls 50. Dies entspricht der gleichen Güte, als wenn die Minutie von einer Hälfte der Experten nicht erkannt, von der anderen Hälfte jedoch erkannt und mit einer Minutien-Qualität von 100 bewertet werden würde.

4.4.5 Evaluation des Ansatzes

Zur Evaluation der Methodologie wurden in [Lod09] 11 Bilder aus der NIST SD14, SD29 Datenbank verwendet, die von 11 unterschiedlichen Experten des Bundeskriminalamts ausgewertet wurden. Der durchschnittliche Abstand zwischen parallelen Papillarlinien und dem Fingerabdruckbereich wurde manuell berechnet.

12 Die mittleren Koordinaten ergeben sich durch Mittelwertbildung aller zum Cluster beitragenden Endpunkte der Einheitsvektoren auf dem Einheitskreis.

Abbildung 22 zeigt ein Beispiel in dem von Experten identifizierte Minutien in das zu Grunde liegende Originalbild eingezeichnet wurden. Rechtecke repräsentieren hierbei Enden von Papillarlinien während Dreiecke Minutien des Typs „other“ darstellen. Wie man leicht in Abbildung 22 erkennen kann, sind die Messungen der Experten im Allgemeinen konsistent in Hinblick auf Bestimmung des Minutien-Typs und dessen Lokation. Allerdings gibt es durchaus Fälle in denen eine eindeutige Bestimmung nicht immer unproblematisch ist (siehe die zwei Minutien vom Typ „other“ in der oberen linken Ecke von Abbildung 22

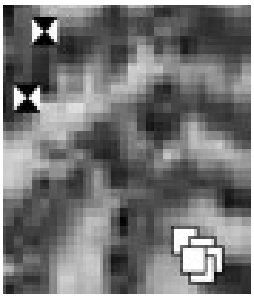

Abbildung 22: Typ und Position von Minutien (8 Experten, weiße Rechtecke repräsentieren Papillarlinienenden, Rechtecke mit schwarzen Dreiecken repräsentieren Minutien des Typs "other")

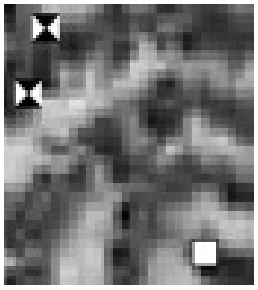

Abbildung 23: Ort und Typ der Minutien-Cluster (weiße Rechtecke repräsentieren Papillarlinienenden, Rechtecke mit schwarzen Dreiecken repräsentieren Minutien vom Typ "other")

Ein möglicher problematischer Fall ist zum Beispiel eine nur sehr kurze Papillarlinie (ein Papillarpunkt). Einige Experten markieren Anfang und Ende dieser kurzen Papillarlinie, während andere den Mittelpunkt des Papillarpunkts markieren und als Minutien-Typ „other“ verwenden. Andere Probleme können durch Unsicherheit bei der Unterscheidung zwischen Enden und Gabelungen von Papillarlinien entstehen.

Abbildung 23 zeigt die Anwendung des Cluster-Algorithmus auf die in Abbildung 22 dargestellten Ergebnisse der Experten. Das identifizierte Cluster (rechte untere Ecke) ist sehr zuverlässig, da die Bewertung der Mehrheit der Experten konsistent war. Wenn dies nicht der Fall ist und die von Experten erzeugten Minutien weiter um die tatsächliche Position gestreut sind, kann der Cluster-Algorithmus anstatt einen, zwei oder mehrere Cluster-Zentren identifizieren.

Um lediglich die zuverlässigsten Cluster in den Ground-Truth Datensatz zu übernehmen, muss die Qualität eines Clusters durch einen Grenzwert gesichert werden. Dies führt einerseits dazu, dass weder Cluster mit aufgenommen werden, die lediglich auf der Beurteilung eines einzigen Experten beruhen, noch darf der Grenzwert zu hoch sein, sodass zu wenige Cluster übernommen werden und

die Konformitätsraten somit auf Basis zu weniger Ground-Truth Minutien bestimmt werden würden.

Um einen geeigneten Grenzwert zu identifizieren, wurden in [Lod09] alle Konformitätsraten für Grenzwerte im Intervall zwischen 0 und 50 berechnet. Im Anschluss wurden Mittelwerte sowie Standardabweichungen dieser Raten bestimmt (vgl. Abbildung 24). Als Grenzwert wurde derjenige Wert ausgewählt, indem sowohl cr_{gtm} als auch cr_{amf} den gleichen Wert aufweisen. Für den weiter

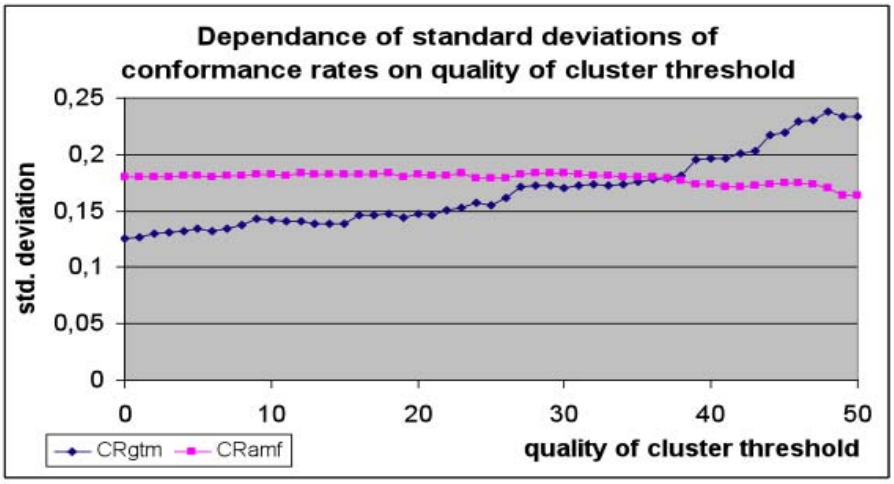

Abbildung 24: Standardabweichung von Konformitätsraten über "quality of custer" Grenzwert

oben beschriebenen einfachen Datensatz ergibt sich somit der Grenzwert 37. Die errechneten Konformitätsraten sind für diesen Evaluationsdatensatz in Tabelle 3 gegeben.

	cr_{gtm}	cr_{agm}	cr_{amf}	*ngtm*	*nagm*
Mittelwert	0,353	0,885	0,662	59	100
Standardabweichung	0,179	0,066	0,178		

Tabelle 3: Ergebnisse für den gewählten „quality of cluster“ Grenzwert (37)

Abbildung 25 zeigt Cluster-Zentren, d.h. Minutien des Ground-Truth Datensatzes, die den bestimmten Qualitäts-Grenzwert von 37 überschreiten. Wie man erkennt, wurden die weiter oben erläuterten Probleme durch die Setzung dieses Grenzwertes behoben und die problematischen Cluster nicht übernommen.

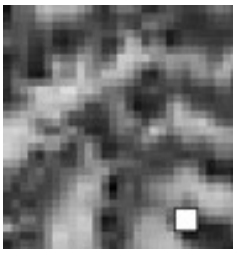

Abbildung 25: Positionierung und Typen der Cluster-Zentren, die dem berechneten Qualitäts-Grenzwert von 37 genügen

4.4.6 Verwendung des Ground-Truth Datensatzes

Die vom Bundeskriminalamt bereitgestellten wahren Minutien-Daten werden als Prüfdaten zur Bewertung der Qualität der FE-Algorithmen eingesetzt, wobei die in [Lod09] definierte Methodology für Conformance Testing Level 3 eingesetzt wird. Es kamen 733 Fingerbilder und der GTM Testdatenbestand zum Einsatz. 486 Bilder entstammen der SD14 und 247 Bilder entstammen der SD29. Auswahlkriterium für die Bilder waren das Vorliegen von mindestens drei Expertenmeinung und geringe Bildstörungen wie zum Beispiel doppelte (übereinanderliegende) Fingerabdrücke.

4.5 Ergebnisse der Qualitätsuntersuchung der FE-Algorithmen

Das vorliegende Testverfahren wurde im Projekt auf einen Datenbestand von 733 Fingerbildern angewandt. Die Tabelle 4 zeigt für die drei getesteten Algorithmen mindtct (NIST) [nbis], Innovatrics [INN] und Neurotechnology [NT] die Ergebnisse für die semantischen Konformitätsraten cr_{gtm}, cr_{agm} und cr_{amf}:, wie sie in Abschnitt 4.3 definiert wurden. Die Parameter *ngtm* und *nagm* geben die Anzahl von Ground-Truth-Minutien bzw. automatisch erzeugten Minutien an. Es wird deutlich, dass die trainierten Experten zwar im Mittel 48 Minutien pro

GTM-DB		cr_{gtm}		cr_{agm}		cr_{amf}		ngtm				nagm			
		mean	std. dev.	mean	std. dev.	mean	std. dev.	mean	std. dev.	min	max	mean	std. dev.	min	max
SD29	**NIST**	0,468	0,122	0,847	0,085	0,522	0,152	48,7	22,6	2	131	88,5	32,3	17	191
SD29	**INN**	0,299	0,100	0,937	0,054	0,670	0,118	48,7	22,6	2	131	64,2	23,9	12	131
SD29	**NT**	0,294	0,117	0,974	0,040	0,692	0,135	48,7	22,6	2	131	53,8	21,9	11	139
SD14	**NIST**	0,464	0,092	0,857	0,063	0,645	0,123	76,4	29,4	7	174	201,9	49,4	87	366
SD14	**INN**	0,296	0,074	0,876	0,078	0,707	0,097	76,4	29,4	7	174	160,4	38,6	74	255
SD14	**NT**	0,299	0,098	0,838	0,102	0,667	0,119	76,4	29,4	7	174	147,1	35,7	60	255

Tabelle 4: Ergebnisse für den SD14/SD29 Testdatensatz mit 733 Fingerbildern und je 3 Expertenmeinungen.

Fingerbild identifizieren konnten, in extremen Situation jedoch nur 2 Minutien, wodurch die Mindestminutienanzahl von 12 Minutien [ISOc] deutlich unterschritten ist. Die erzielten Werte für *nagm* zeigen, dass alle drei getesteten Verfahren in der Regel eine ausreichende Menge an Minutien detektieren konnten. Die Standardabweichung (std.dev) gibt einen Eindruck der Streuung des Ergebnis über die Menge der untersuchten Testbilder in den jeweiligen Testdatenbanken SD14 und SD29.

Für die Konformitätsrate cr_{amf} liegen die kommerziellen Verfahren INN und NT erwartungsgemäß vor der frei verfügbaren Software des NIST, die bekanntermaßen auch im Fingerabdruckfokus

falsche Minutien liefert (Siehe dazu auch die Abbildungen in Kapitel 4.3.) Das Ergebnis ist weitgehend unabhängig von der verwendeten Datenbank.

Mit der Konformitätsrate cr_{agm} wird insbesondere bei der Datenbank SD29 deutlich, dass die kommerziellen Verfahren eine deutlich besseres Ergebnis erzielen, da sie nicht wie das NIST-Verfahren falsche Minutien in großer Anzahl außerhalb des Fingerabdruckbereichs detektieren (siehe dazu Abbildung 14).

Überraschend ist auf den ersten Blick die geringere Konformitätsrate cr_{gtm} beider Produktverfahren. Das Ergebnis ist deutlich schlechter als der Vergleichswert des NIST-Verfahrens. Dies ist insbesondere deshalb verwunderlich, da beide Produktverfahren durch gute Erkennungsleistungen bekannt sind. Eine besseres Ergebnis könnte daher auch für diese semantische Konformitätsrate erwartet werden.

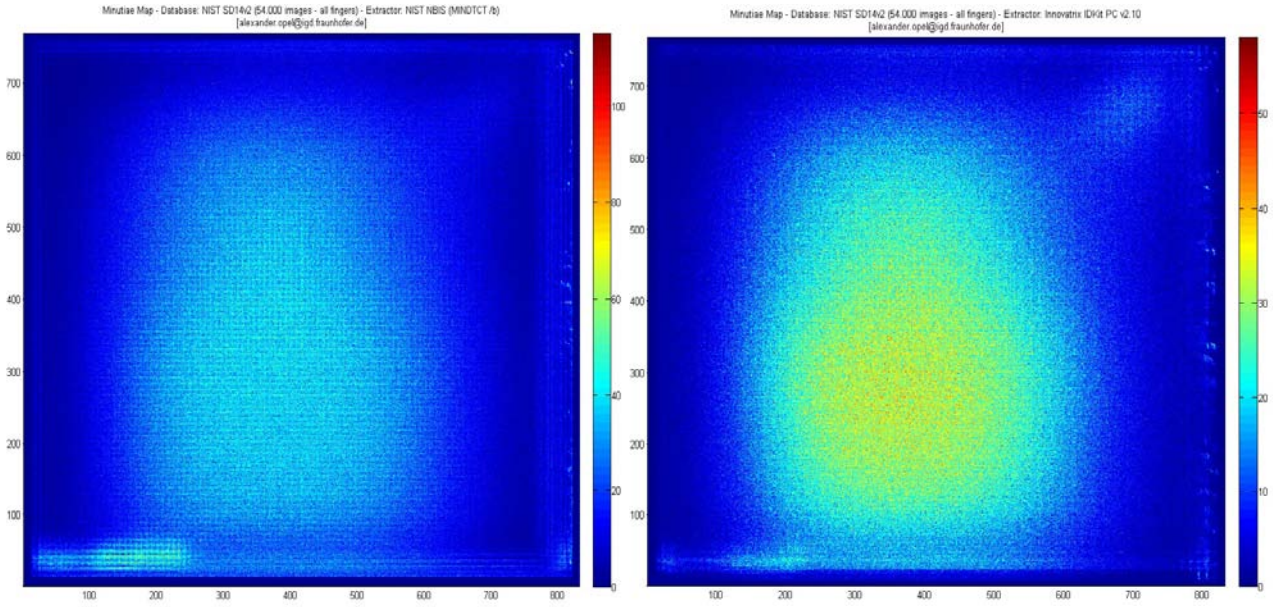

Abbildung 26: Vergleich der Minutien-Verteilung: links - NIST mindct, rechts – Innovatrics mit deutlich erkennbarer Gitterstruktur.

Eine Erklärung ergibt sich aus dem Vergleich der Häufigkeiten einer Minutienposition im Fingerabdruckbild. Die oben stehende Abbildung zeigt links eine um das Zentrum des Fingerbildes gleichmäßige Verteilung der Minutien. Die Produktversion von Innovatrics zeigt jedoch eindeutig ein regelmäßiges Gitter, auf das die AGM Minutien bewusst oder unbewusst projeziert werden. Dieser systematische Fehler kann in der „Intra-Produkt-Testing“-Erkennungsleistung nicht bemerkbar sein, da die Translation sowohl bei Referenz als auch bei der Probe stattfindet. Bei einer „Inter-Produkt-Testing“-Auswertung wird diese räumliche Verschiebung (in Referenz oder Probe) die Ursache für eine zu große Distanz von zugeordneten Minutien haben. Dies wird durch das Distanzmaß der cr_{gtm} Konformitätsrate in gleicher Weise ausgedrückt.

4.6 Bewertung und Ausblick

Dieses Kapitel beschreibt die Anwendung einer Methodik für Level 3 Konformitätstests zur Bewertung von FE-Algorithmen. Der hier beschriebene Ansatz wurde exemplarisch implementiert und die im Rahmen einer vorläufigen Evaluation durchgeführten Tests lieferten vielversprechende Ergebnisse, da sie durch eine niedrige Konformitätsrate cr_{gtm} klar die möglichen Interoperabilitäts-Probleme anzeigen. Zur Prüfung der Qualität der FE-Algorithmen wurde ein Konformitätstest für den NIST mindtct, Innovatrics sowie den Neurotechnology Algorithmus durchgeführt und die gewonnenen Ergebnisse bescheinigen die Plausibilität des Ansatzes. Dennoch ist es notwendig, weitere Tests mit mehreren hundert Bilddateien durchzuführen sowie einigen offenen Problemen und Fragestellungen nachzugehen: (i) Einbeziehung von Konformitätsraten für Cores und Deltas, (ii) qualitätsgesteuerte, semi-automatisierte Definition des Fingerabdruckbereiches, (iii) qualitätsgesteuerte, semi-automatisierte Definition des durchschnittlichen Abstands zwischen Papillarlinien im zu bearbeitenden Bild, (iv) Bestimmung und Validierung von Grenzwerten für alle Konformitätsraten, sodass Minutien-Extraktions-Algorithmen nur dann Konformität bescheinigt wird, wenn alle Grenzwerte überschritten werden und (v) Validierung des „Clustering von Clustern“ bzw. Clustering Ansatzes entsprechend des Minutien-Typs.

5 Integration von Zusatzinformationen

5.1 Motivation

Zahlreiche wissenschaftliche Publikationen sowie die Ergebnisse aus den vorherigen Projekten zeigen, dass Template-Protection-Verfahren die Sicherheit biometrischer Systeme erheblich verbessern und dabei gleichzeitig die Privatsphäre der Benutzer schützen können. Die inhärenten Eigenschaften der Biometrie schränken allerdings die Erkennungsleistung der Verfahren ein. Ähnlichkeiten in den biometrischen Merkmalen unterschiedlicher Benutzer und Variationen in den Merkmalen bei verschiedenen Aufnahmen ein und desselben Benutzers können zu Falschübereinstimmung und Falschnichtübereinstimmung im biometrischen System führen. Um diese Fehlerraten zu minimieren, bedarf es einer zusätzlichen Informationsquelle, beispielsweise einer weiteren Modalität oder eines Passworts.

Die Eingabe einer PIN oder eines Passworts ist eine weit verbreitete Authentisierungsmethode. Längere und damit sichere Passwörter sind jedoch nur schwer im Gedächtnis zu behalten, kürzere PINs hingegen bieten keine ausreichende Sicherheit. Zudem fehlt die direkte Verbindung zwischen dem Benutzer und dem Identitätsnachweis. Pseudonyme Identifikatoren sind meistens binär und lassen sich einfach mit einer PIN oder einem Passwort kombinieren (siehe Abschnitt 2.4.1). Die Integration von Zusatzinformationen in Template-Protection-Verfahren ist durchaus geeignet, sowohl die Sicherheit als auch die Erkennungsleistung des gesamten biometrischen Systems zu erhöhen. In diesem Kapitel konzentrieren wir uns auf minutienbasierte Fingerabdruckerkennungssysteme und analysieren mögliche Integrationsmethoden. Es wird ein Verfahren vorgestellt, das Zusatzinformationen in ein Fuzzy-Vault-Verfahren integriert.

5.2 Die möglichen Integrationsmethoden

Die Integration von Zusatzinformationen zielt auf die Verbesserung der Sicherheit und der Erkennungsleistung; in anderen Worten: die Sicherheit des biometrischen Systems wird erhöht bei verbesserter Benutzbarkeit. In Kapitel 3 wurden unterschiedliche Template-Protection-Verfahren vorgestellt. In den biometrischen Kryptosystemen besteht der pseudonyme Identifikator aus dem Hash-Wert einer geheimen Zeichenkette. PIN und Passwort können auch als Zeichenkette dargestellt werden. Da in den vorherigen BioKey-Projekten biometrische Kryptosysteme verwendet wurden, stehen auch hier Integrationsmethoden für biometrischen Kryptosysteme im Vordergrund. Es wurden drei mögliche Integrationsmethoden identifiziert, die in der Folge vorgestellt und deren Vor- und Nachteile analysiert und dargestellt werden. Die folgenden Parameter werden verwendet, um die Sicherheit und die Performanz zu analysieren:

- *Lp,* die Länge der Zusatzinformation in Bits
- *t,* Schwellwert des Template-Protection-Verfahrens, hier die Anzahl der übereinzustimmenden Minutien
- $FMR_{tp}(t)$, die Falschübereinstimmungsrate des Systems mit Template-Protection tp am Schwellwert *t*.

- $FNMR_{tp}(t)$, die Falschnichtübereinstimmungsrate des Systems mit Template-Protection tp am Schwellwert *t*.
- $FMR(t)$, die Falschübereinstimmungrate eines Template-Protection-Systems mit Zusatzinformation am Schwellwert *t*.
- $FNMR(t)$, die Falschnichtübereinstimmungsrate eines Template-Protection-Systems mit Zusatzinformation am Schwellwert *t*.

5.2.1 Verlängerung der geheimen Zeichenkette

Die einfachste Methode, die Sicherheit zu erhöhen, ist das Verlängern der geheime Zeichenkette mit einer PIN oder einem Passwort. Dabei sind keine Änderungen am Template-Protection-Verfahren selbst notwendig. Der pseudonyme Identifikator (PI) ist der Hash-Wert der geheimen Zeichenkette und der Zusatzinformation. Der PI kann allerdings nur dann richtig generiert werden, wenn die geheime Zeichenkette erfolgreich berechnet und die Zusatzinformation korrekt eingegeben wurde.

Die Verbesserung der Sicherheit kann mit der Länge der Zusatzinformation in Bits quantifiziert werden. Je länger die Zusatzinformation ist, um so sicherer ist das System. Da keine Änderungen am biometrischen Verfahren und am Template-Protection-Verfahren selbst vorgenommen werden, bleibt die Falschnichtübereinstimmungsrate (FNMR) unverändert. Die Falschübereinstimmungsrate (FMR) jedoch wird um den Faktor 2^{-Lp} verkleinert.

$$FNMR(t) = FNMR_{tp}(t)$$

$$FMR(t) = FMR_{tp}(t) \cdot (2^{-Lp})$$

5.2.2 Verlängerung der biometrischen Merkmale

Alternativ zur oben vorgestellten Methode können auch die biometrischen Merkmale um Zusatzinformationen erweitert und als Eingabedaten für Template-Protection herangezogen werden. Solche künstlichen Merkmale werden aus benutzerspezifischen Zusatzinformationen generiert. Sie haben den Vorteil, dass sie bei Genuine-Vergleichen invariant bleiben, aber sich die Unterscheidbarkeit bei Imposter-Vergleichen vergrößert. AP ist die Anzahl der künstlichen Minutien. Unter der Annahme, dass die künstlichen Minutien eindeutig von Zusatzinformationen bestimmt werden, sind die FNMR und FMR des Systems mit Zusatzinformation:

$$FNMR(t) = FNMR_{tp}(t - AP)$$

$$FMR(t) = FMR_{tp}(t - AP) \cdot (2^{-Lp})$$

Die FNMR ist deutlich verkleinert. Die Kollisionswahrscheinlichkeit, dass Zusatzinformationen unterschiedlicher Benutzer identisch sind, ist gleich 2^{-Lp} . Wenn eine Falschübereinstimmung auftritt, muss die Zusatzinformation identisch ist. Obwohl $FMR_{tp}(t - AP) \geq FMR_{tp}(t)$, kann FMR(t) durch 2^{-Lp} verringert werden. Die Sicherheit kann weiter verbessert werden, wenn die Zusatzinformationen erneut verwendet und damit die geheime Zeichenkette verlängert wird. Die Änderung der FMR und FNMR bezüglich der Anzahl der übereinstimmenden Minutien wird in Abbildung 27 gezeigt.

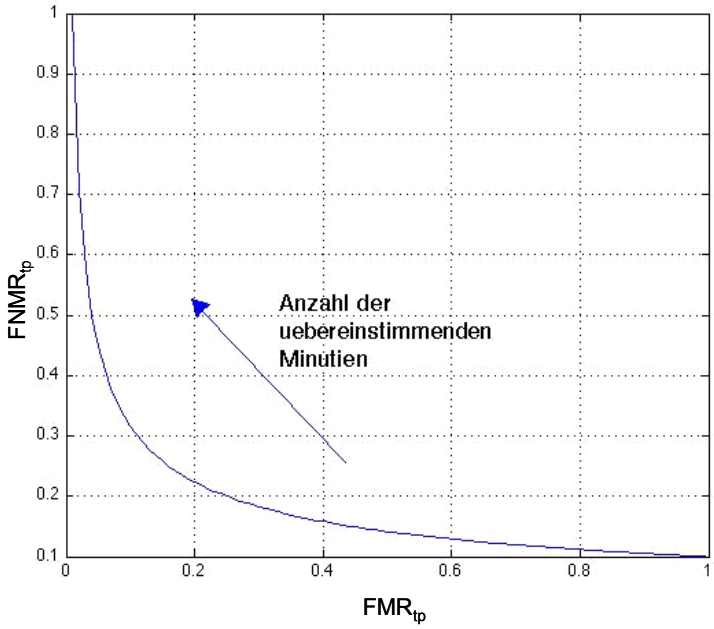

Abbildung 27: DET-Kurve: Änderung der FMR_{tp} und $FNMR_{tp}$ bezüglich der Anzahl der übereinstimmenden Minutien

Mit dieser Methode kann sowohl die Sicherheit als auch die Erkennungsleistung des Systems verbessert werden. Darüber hinaus sind umfangreichere Einstellmöglichkeiten für *FMR* und *FNMR* gegeben.

5.2.3 Härtung Fuzzy Vault

In Abschnitt 3.3.1.2 wird der Vernüpfungs-Angriff auf Template-Protection-Verfahren erläutert. Wegen der Offenlegung der unterstützenden Punkte im Vault Set ist das Fuzzy-Vault-Verfahren besonders anfällig für diesen Angriff. In [NNJ07] wird der Ansatz „Härtung des Fingerabdruck-Fuzzy-Vault mittels Passwort“ aufgezeigt. Mit Hilfe eines Passworts werden Minutien eines Fingerabdrucks rotiert und verschoben, um die Resistenz gegen einen Verknüpfungs-Angriff zu erhöhen. Dabei werden die Informationen individueller Minutien sowie deren Position und Winkel als 16-Bit-Zeichenfolge dargestellt. Der gesamte Fingerabdruckbereich wird in vier Quadranten unterteilt. Ein 64-Bit Passwort besteht aus acht Zeichen. Aus dem Hash-Wert des Passworts wird eine Permutation für die vier Quadranten berechnet. Dazu wird das 64-Bit-Passwort in vier Gruppen geteilt, die jeweils die Translation und Rotation der Minutien in jedem Quadranten bestimmen. Die geänderten Minutien dienen sodann als Eingabedaten für das Fuzzy-Vault-Verfahren. Das Vault Set wird jedoch noch zusätzlich durch Verschlüsselung gesichert, wobei der Schlüssel aus dem bereits definierten Passwort erzeugt wird. Die Verschlüsselung am Ende des Prozesses ist ein wichtiger Schritt für die Sicherheit, da allein die Änderungen an den Minutien nicht ausreichend sind, um die Verknüpfungsproblematik zu vermeiden. Die Änderungen der Minutien erfolgt in jedem Quadranten, es gibt kaum Änderungen der relativen Positionen zwischen den Minutien innerhalb eines Quadranten. Tritt eine ausreichende Überlappung zwischen den Quadranten zweier Enrolment-Bilder auf, kann man von einer Korrelation der beiden Vault Sets ausgehen. Darüber hinaus wird in [NNJ07] berichtet, dass eine leichte Steigerung der FNMR in den durchgeführten Experimenten beobachtet werden konnte. Dies kann von den vorkommenden Änderungen der

relativen Minutien-Positionen verursacht werden. Die Sicherheit des Verfahrens liegt in der Komplexität, *Lp*-Bit-lange Passwörter richtig zu schätzen und im Aufwand, das richtige Polynom gleichzeitig zu konstruieren. Verknüpfungen zwischen den geschützten Referenzen unterschiedlicher Anwendungen werden damit vermieden.

5.2.4 Fazit

Die ersten beiden Methoden wurden im Rahmen dieses Projekts entwickelt, während die letzte Methode der Literatur entnommen ist. Alle drei vorgestellten Verfahren sind dazu geeignet, die Sicherheit der Systeme zu verbessern. Die ersten beiden Verfahren können Verknüpfungen der pseudonymen Identifikatoren verhindern, da eine Verifikation ohne die Eingabe der korrekten Zusatzinformation nicht möglich ist. Aber Verknüpfungen durch die Minutieninformationen in den Vault Sets können nicht vermieden werden. Das dritte Verfahren ist resistenter gegen diese Art der Verknüpfung.

Die FMR kann durch die Hinzunahme der Zusatzinformation reduziert werden. Das größere Problem in biometrischen Systemen stellt jedoch eine höhere FNMR dar. Das erste geschilderte Verfahren kann die FNMR nicht ändern, im dritten Verfahren wird die FNMR sogar noch gesteigert. Lediglich im Verfahren durch Verlängerung der biometrischen Merkmale gibt es die Möglichkeit, die FNMR zu reduzieren. Dies ist auch der Grund, weshalb in diesem Projekt das zweite Verfahren ausgewählt wurde.

Im nächsten Abschnitt wird ein Verfahren beschrieben, das die biometrischen Merkmale mit Zusatzinformationen verlängert. Das Verfahren wird für ein Fingerabdruckerkennungssystem implementiert, das auf Fuzzy Vault basiert und die experimentellen Ergebnisse mit einer Fingerabdruckdatenbank werden demonstriert.

5.3 Ein effizientes Template-Protection-Verfahren mit Zusatzinformation

Im letzten Abschnitt wurden drei unterschiedliche Implementierungsansätze vorgestellt. Von diesen ermöglicht das zweite Verfahren der „Verlängerung der biometrischen Merkmale“ die besten Einstellmöglichkeiten zwischen FNMR, FMR und der Länge der geheimen Zeichenkette. Damit wird es einfach, ein ausgewogenes Verhältnis zwischen Erkennungsleistung und Sicherheit herzustellen. In diesem Abschnitt wird ein effizientes Template-Protection-Verfahren mit Zusatzinformation für Fuzzy-Vault dargestellt. Das Verfahren wird in ein minutienbasiertes Fingerabdruckerkennungssystem implementiert und mit Hilfe der NIST SD14 Datenbank getestet. Die experimentellen Ergebnisse sowie die Evaluierung der Leistungsfähigkeit und der Sicherheit werden aufgezeigt.

In diesem Fuzzy-Vault-Verfahren wird jede Minutie als Kette von 16-Bit Länge dargestellt. Um eine genügend große Anzahl künstlicher Minutien zu generieren, bedarf es ausreichender Zusatzinformationen. Angenommen, es wird eine 4-stellige Zahl als Zusatzinformation benutzt. Das entspräche $\log_2 10^4 = 13.29$ Bits und es ist offensichtlich, dass dies nicht ausreicht, um genügend künstliche Minutien zu generieren. Eine Permutation der Zusatzinformation ist nicht zulässig, da die generierten Minutien miteinander korrelieren können. Stattdessen kann die Zusatzinformation als Seed (Initialisierungswert) für die Erzeugung einer Pseudo-Zufallszahl dienen, um so mehrere

unabhängige Minutien zu erstellen. Somit wirkt sich eine Verlängerung des Passworts nur auf den Initialisierungswert des Pseudo-Zufallszahlen-Generators für die Erzeugung unabhängiger künstlicher Minutien aus. Die Anzahl der künstlichen Minutien ist ein Systemparameter, der unabhängig von der Länge des Passworts ist. Die Anzahl der künstlichen Minutien darf dabei jedoch nicht die Anzahl der Polynom-Koeffizienten übersteigen, da sonst keine biometrische Information bei der Verifikation herangezogen wird.

5.3.1 Enrolment

Im Enrolment-Prozess werden neben den extrahierten echten Fingerbildminutien auf Basis der Zusatzinformation K mit Hilfe des Pseudo-Zufallszahlen-Generators künstliche Minutien generiert. Dabei muss jedoch ein gewisser Abstand zwischen den unabhängigen künstlichen Minutien und den Fingerbildminutien gewahrt werden. Dies ist notwendig, da Polynomprojektion und -rekonstruktion zu Effizienzzwecken in einem endlichen Körper (16-Bit) durchgeführt werden. Dies geschieht in Anlehnung an das original Fuzzy-Vault-Verfahren wie es in [FbFV] gezeigt wurde. Da alle Minutien somit als 16-Bit-Werte im endlichen Körper verarbeitet werden, findet eine Diskretisierung statt. Aus diesem Grund müssen die einzelnen echten und künstlichen Minutien voneinander einen Abstand aufweisen, damit die Minutien im endlichen Körper voneinander eindeutig trennbar sind. Der Mindestabstand wird dabei in x- und y-Richtung eingehalten. Nachdem die künstlichen Minutien (Artificial Points) erzeugt wurden, werden sie mit den echten Minutien des Fingerabdruckbilds (Enrolment Points) fusioniert (Kasten Template Fusion in Abbildung 28) und als Eingabedaten für das normale Fuzzy-Vault-Verfahren genutzt.

Im ersten Schritt wird ein binarisiertes Geheimnis S zufällig generiert. Dann wird das Geheimnis zu d+1 Zahlen im endlichen Körper konvertiert, wobei d der Grad des Fuzzy-Vault-Polynoms ist. Daraus wird zusammen mit der Koeffizientenanzahl das Fuzzy-Vault-Polynom im endlichen Körper generiert. Die echten und künstlichen Minutien werden nun in den endlichen Körper übertragen (nun als α-Werte bezeichnet), wo sie anschließend auf das Polynom projiziert werden (β-Werte). Dabei wird der α-Wert im endlichen Körper wie folgt als 16-Bit-Wert generiert: Die x- und die dazugehörigen y-Minutienkoordinaten werden jeweils auf 8-Bit diskretisiert und anschließend zu einem 16-Bit-Wert aneinandergereiht. Dies entspricht somit einer Zahl im endlichen Körper.

Anschließend werden zur Erhöhung der Sicherheit weitere zufällige Streupunkte (Chaff Points) generiert, die ebenfalls einen Abstand zu den echten und künstlichen Minutien wahren müssen. Zu diesen Streupunkten werden künstliche Polynom-Projektionen im endlichen Körper generiert, die jedoch nicht auf dem Fuzzy-Vault-Polynom liegen. Die nun vorhandenen echten, künstlichen und Streupunkte mit deren zugehörigen Projektionen werden nun zufällig verwürfelt („Liste Scrambling“ in Abbildung 28) und bilden zusammen ein Vault Set. Weiterhin wird ein aus dem binarisierten Geheimnis (S) und Passwort (K) gebildeter Hash-Wert ($h([S,K])$) zusammen mit einer benutzerspezifischen ID (User ID) und dem Vault Set gespeichert. In Abbildung 28 kann der gesamte Enrolment-Prozess schematisch betrachtet werden.

In diesem Verfahren ist es jedoch möglich, dass unterschiedliche Passwörter die gleichen künstlichen Minutien generieren, weil die Reihenfolge der Minutien keine Rolle spielt. Beispielsweise können AP künstliche Minutien $AP! = \prod_{i=1}^{AP} i$ verschiedene Zeichenketten generieren, die aus unterschiedlichen Passwörtern erzeugt werden können. Jedoch beeinflusst dieser

Effekt die Falschübereinstimmungsrate nicht, da ein Vergleich nicht erfolgreich sein kann, wenn die jeweiligen Passwörter nicht identisch sind.

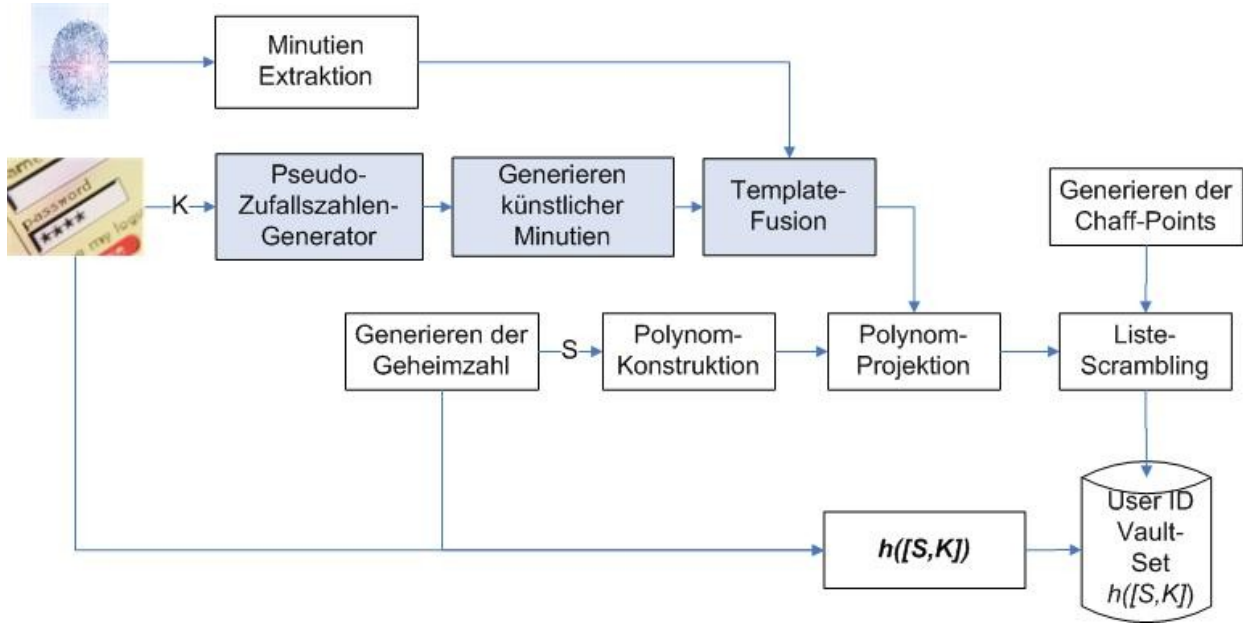

Abbildung 28: Enrolment-Prozess

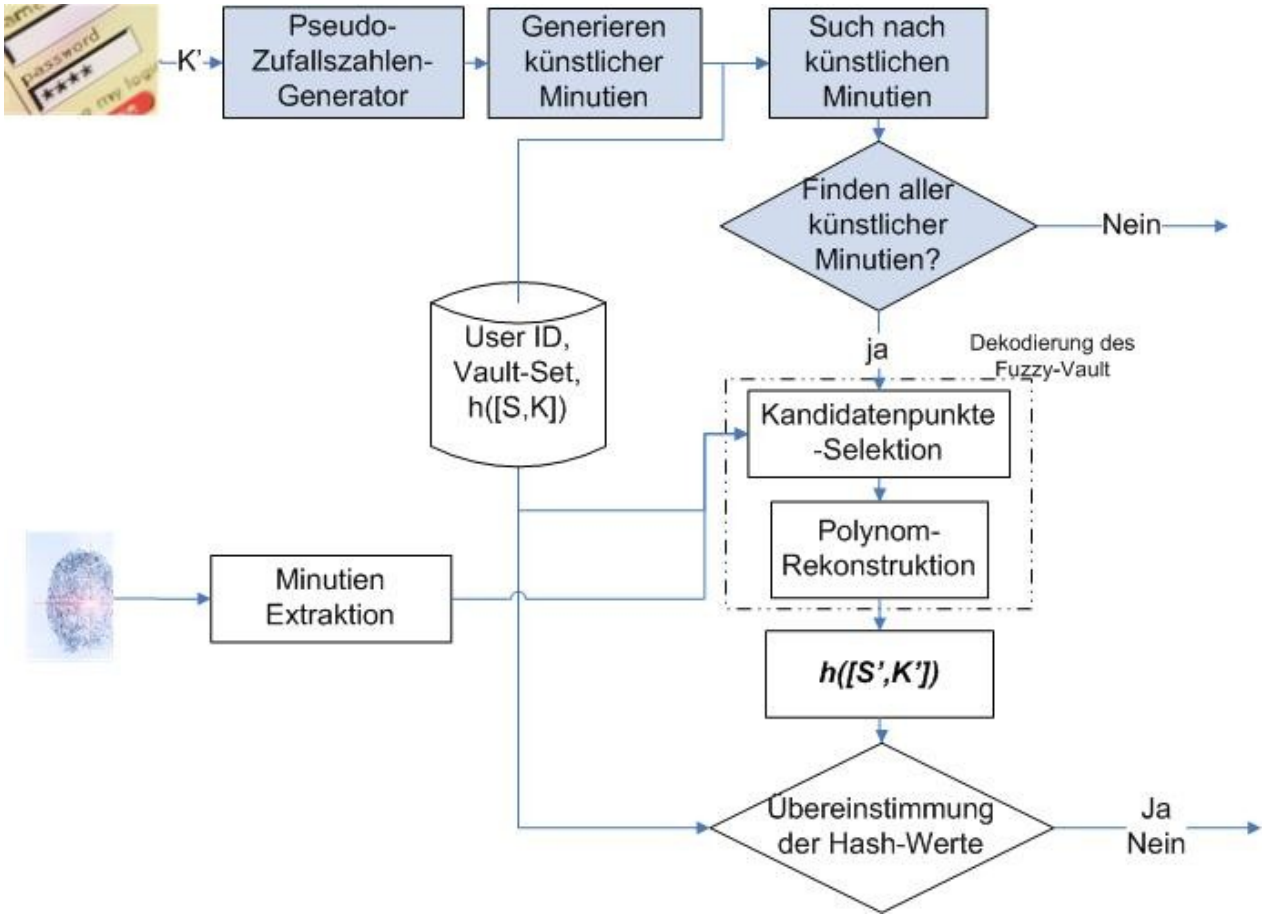

Abbildung 29: Verifikations-Prozess

5.3.2 Verifikation

Bei der Verifikation (siehe Abbildung 29) werden anhand der Zusatzinformation K' künstliche Minutien generiert. Diese stimmen bei der Benutzung derselben Zusatzinformation mit den künstlichen Minutien des Enrolments überein. Anhand der User ID wird nun das passende Vault Set geladen und die nun in den endlichen Körper übertragenen neu generierten künstlichen Minutien mit allen Punkten im Vault Set verglichen. Wenn nicht alle künstlichen Minutien exakt gefunden werden können, wurde die falsche Zusatzinformation für die Erzeugung genutzt und die Verifikation wird vorzeitig als nicht erfolgreich abgebrochen. Wenn der Vergleich positiv verläuft, wird versucht, das Polynom zu rekonstruieren (siehe gestrichelten Kasten in Abbildung 29). Dazu werden die Minutien eines Verifikations-Fingerbilds mit den Minutieninformationen im Vault Set (ohne die künstlichen Minutien) verglichen. Die dabei übereinstimmenden Minutieninformationen werden als Index-Liste vom Minutienkomparator zurückgegeben. Die den Indizes entsprechenden Minutieninformationen des Vault Sets dienen als Kandidaten für echte Minutien entsprechend denen im Enrolment-Prozess. Zusammen mit den künstlich erzeugten Minutien wird nun versucht, das Polynom zu rekonstruieren. Dabei muss die Anzahl der übereinstimmenden echten zusammen mit den künstlichen Minutien mindestens der Zahl der Koeffizienten des Fuzzy-Vault-Polynoms entsprechen, um dass Polynom erfolgreich rekonstruieren zu können. Dabei findet die Rekonstruktion im endlichen Körper statt (siehe Abschnitt 5.3.1). Dazu wird die Lagrangesche Interpolation verwendet. Wenn der Minutienkomparator jedoch mehr Kandidatenpunkte als notwendig extrahiert, müssen alle möglichen Kombinationen der Projektionen in den endlichen Körpern getestet werden, um das Polynom erfolgreich zu rekonstruieren. Ist dies erfolgreich absolviert und das Polynom konnte erfolgreich rekonstruiert werden, wird mit dessen Hilfe das binarisierte Geheimnis S' berechnet. Zusammen mit der binarisierten Zusatzinformation K' wird erneut ein Hash-Wert gebildet und mit dem gespeicherten Hash-Wert verglichen. Stimmen beide überein, wurde das richtige Polynom rekonstruiert und die Verifikation war erfolgreich. In Abbildung 29 ist der gesamte Verifikation-Prozess schematisch dargestellt.

5.3.3 Sicherheit des Verfahrens

Im Folgenden wird die Sicherheit des Verfahrens evaluiert. Die folgenden Parameter wurden in der Evaluierung verwendet:

Systemparamter:

- Grad des Polynoms: d
- Anzahl der Koeffizienten des Polynoms: k = d+1
- Anzahl der Elemente im endlichen Körper: p
- Anzahl der verwendeten echten Minutien im Enrolment: EP
- Anzahl der Streupunkte (Chaff Points): CP
- Anzahl der künstlichen Minutien (Artificial Points): AP
- Anzahl der mindestens zu übereinstimmenden Minutien: $M_{ü}$

Benutzer-Parameter:

- Länge des Passworts in Bits (Lp, Lp=0 für das Verfahren ohne Zusatzinformation)

Die *Länge der geheime Zeichenkette*, *Ls,* wird durch Einsetzen des Passworts vergrößert:

$$L_s = L_p + (d+1)\cdot p$$

Die Zufälligkeit des pseudonymen Identifikators (PI) ist von der Länge der geheime Zeichenkette abhängig, da sowohl Passwörter als auch das binäre Geheimnis zufällig gewählt werden und statistisch unabhängig voneinander sind. Dadurch erhöht sich auch die Sicherheit des PI und dessen Erneuerbarkeit.

Die Sicherheit des Verfahrens liegt im *gesamten Aufwand, die originale Zeichenkette des Hash-Werts richtig zu schätzen*. Die Zeichenkette besteht aus dem Passwort und den Koeffizienten des Polynoms. Die durchschnittliche Anzahl der Versuche, um ein Passwort zu schätzen, ist gleich 2^{L_p-1}. Sobald das Passwort erfolgreich geschätzt ist, sind *AP* künstliche Minutien bekannt. Um das Polynom zu konstruieren, sollen *d+1-AP* Stützpunkte in den restlichen Vault-Set-Punkten (*CP+EP)* gefunden werden. $\binom{CP+EP}{d+1-AP}$ ist die Anzahl der Kombinationen mit $d+1-AP$ Punkten im gesamten Punkte-Set (*CP+EP)*. $\binom{EP}{d+1-AP}$ ist die Anzahl der Kombinationen mit den $d+1-AP$ echten Stützpunkten. Die erwartete Anzahl der Versuche, um das Polynom zu rekonstruieren ist (ähnlich der Sicherheitsevaluierung in [FbFV]):

$$\frac{\binom{CP+EP}{d+1-AP}}{\binom{EP}{d+1-AP}}$$

Die Sicherheit wurde verbessert, weil der Einsatz des Passworts den Schätzungsaufwand des Hash-Werts erhöht. Im Vergleich zum Template-Protection-Verfahren ohne Zusatzinformation, wird die Widerstandsfähigkeit des Verfahrens gegen Verknüpfungen verbessert. Selbst wenn ein paar echte Minutien-Punkte in zwei Vault Sets gleicher Benutzer durch deren Ähnlichkeit gefunden werden könnten, sind diese nicht ausreichend, um die originale Zeichenkette des PI zu rekonstruieren. Es fehlen noch die Passwort-Informationen.

Die Falschübereineinstimmungsrate ändert sich ebenfalls. Eine Falschübereineinstimmung würde passieren, wenn die Passwörter unterschiedlicher Benutzer gleich und deren Minutien auch ähnlich gelagert sind. Wenn $FMR(AP, d+1)$ die FMR des Template-Protection-Verfahrens mit Passwort bei Benutzung der *AP* künstlichen Minutien und des Polynom von Grad *d* und $FMR_{tp}(d+1-AP)$ die FMR des Template-Protection-Verfahrens ohne Passwort mit dem Polynom vom Grad *d+1-AP* sind, ergibt sich:

$$FMR(AP, d+1) = 2^{-L_p}\cdot FMR_{tp}(d+1-AP) \quad .$$

Obwohl $FMR_{tp}(d+1) \leqslant FMR_{tp}(d+1-AP)$, wird meistens $FMR(AP, d+1) < FMR_{tp}(d+1)$ aufgrund des Verkleinerungsfaktors 2^{-L_p} sein.

5.4 Evaluierungsvorbereitung

5.4.1 Minutienkomparator

Für die Verifikation des in Abschnitt 5.3 beschriebenen Template-Protection-Verfahrens wird ein Minutienkomparator benötigt, der eine Index-Liste der übereinstimmenden Minutien zurückliefert. Somit konnte der ursprünglich geplante NIST Bozorth3-Algorithmus nicht verwendet werden. Die Wahl fiel aus diesem Grund auf das VeriFingerSDK von Neurotechnology, um die NIST SD14v2 Fingerbilder zu verarbeiten. Das SDK bietet die Möglichkeit, beim Vergleich zweier Fingerbilder die übereinstimmenden Minutien auszugeben. Dies ist notwendig für die weitere Verarbeitung unter Verwendung von Template Protection und Template Protection mit Zusatzinformationen.

Minutien-Informationen bestehen u.a. aus Position, Winkel und Typ. In diesem Fuzzy-Vault-Verfahren werden nur die x- und y- Koordinaten der Minutien verwendet. Aus diesem Grund wird der Neurotechnology Komparator nur auf Basis der Minutien-Koordinaten genutzt. Hierzu wurden aus den SD14-Fingerbildern nur die Minutien-Koordinaten extrahiert und als CSV-Dateien abgelegt. Diese Daten bilden die Datenbasis für die weitere Verarbeitung. Aus diesen Daten werden nun neue Templates mittels des Neurotechnology VeriFingerSDK angelegt, auf denen nun die Vergleiche und die Ausgabe der übereinstimmenden Minutien stattfinden.

Als Ergebnis eines solchen – auf rein Minutien-Koordinaten basierenden – Vergleichs liefert der Neurotechnology-Algorithmus die übereinstimmenden Minutien und einen Score. Die übereinstimmenden Minutien werden als Indexliste, d.h. welche Minutie des Datenbasistemplates mit welcher Minutie des Vergleichstemplates übereinstimmt, zurückgegeben. Der Score liefert Werte größer gleich 0 und beschreibt die Zuverlässigkeit des Vergleichs. Je höher der Score, desto besser ist die Übereinstimmung.

5.4.2 Vorbereitung der Datenbasis

Abbildung 30: Minutienfilterung mit Hilfe der NIST MINDTCT Quality Map

Da für das Enrolment jeweils nur ein Fingerbild zur Verfügung steht, müssen stabile Minutien gefunden werden, die auch bei einem Genuine-Vergleich im Verifikationsbild auffindbar sind. Ebenfalls müssen Ausreißer in den Minutiensätzen entfernt werden. Aus diesem Grund werden die extrahierten Minutien (rote Markierungen in Abbildung 30) eines jeden Minutiensatzes mit der jeweiligen Quality Map des NIST MINDTCT verglichen. Dabei werden alle Minutien entfernt, die nicht der höchsten Qualitätsstufe angehören. Die Ergebnisse zeigen, dass fälschlich erkannt Minutien (wie bspw. solche, die an den Kartenaufdrucken der SD14-Datenbank entstehen) somit entfernt werden können. In Abbildung 30 ist ein Beispiel zu sehen, bei dem nur Minutien beibehalten werden (grüne Dreiecke), die in der NIST Quality Map der höchsten Stufe entsprechen. Fälschlich erkannte Minutien bzw. Minutien schlecht Qualität werden somit entfernt.

Da die Verarbeitung nur auf Minutien-Koordinaten basieren, ist eine Betrachtung der Datenbasis notwendig. Dazu wurden zuerst Genuine-Vergleiche auf den bereinigten Minutiendatensätzen durchgeführt, die wie oben beschrieben jeweils einen Score und die Anzahl der übereinstimmenden Minutien liefern. Als Ergebnis ist erkennbar, dass es zu jedem Genuine-Vergleich (es wurden hierfür alle 2700 Fingerbilder des linken Zeigefingers betrachtet) übereinstimmende Minutien gibt. Jedoch zeigt sich, dass nicht alle einen Score-Wert besitzen, der ungleich 0 ist. Der Grund hierfür ist das Fehlen weiterer Informationen, die für den Matching-Vorgang von Vorteil wären. Dazu gehören unter anderem die Winkelinformationen oder der Typ der jeweiligen Minutie. Tests, bei denen der Vergleich mit allen Informationen durchgeführt wurde, zeigen, dass immer Score-Werte größer 0 zurückgegeben werden (vorausgesetzt, es konnten überhaupt Minutien extrahiert werden).

In Abbildung 31 sind die übereinstimmenden Minutien eines Genuine-Vergleichs dargestellt. Der Score-Wert des Vergleichs ist 431 und somit größer 0. Es ist eindeutig zu erkennen, dass die übereinstimmenden Minutien sehr ähnlich gelagert sind.

In Abbildung 32 ist ein Vergleich mit einem Score-Wert von 0 dargestellt. Es werden jedoch auch 25 übereinstimmende Minutien gefunden. Es ist jedoch auch visuell erkennbar, dass die Übereinstimmungen nicht in der Art entsprechen, wie dies in Abbildung 31 der Fall ist.

Da die Vergleiche die einen Score-Wert von 0 haben einen großen Einfluss auf die Weiterverarbeitung für die Integration von Template Protection bzw. Template Protection mit

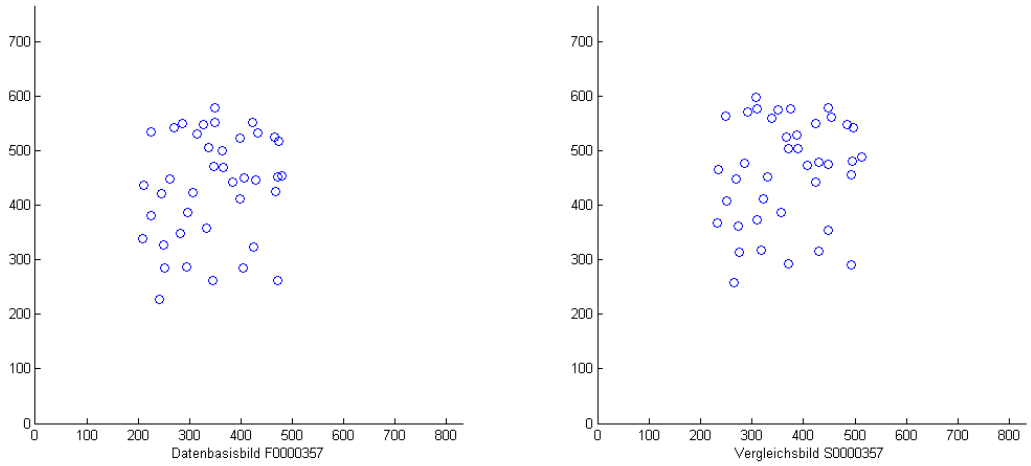

Abbildung 31: Pixelpositionen der übereinstimmenden Minutien; Mü = 40, Score-Wert = 431

Zusatzinformationen haben, wird die Datenbasis in der Hinsicht eingeschränkt, dass nur Fingerbilder genutzt werden, die beim Genuine-Vergleich einen Score-Wert größer 0 haben. Würden alle Fingerbilder genutzt werden, die einen Score-Wert von 0 haben, würde dies eine große FNMR nach sich ziehen, da die übereinstimmenden Minutien unzuverlässig sind.

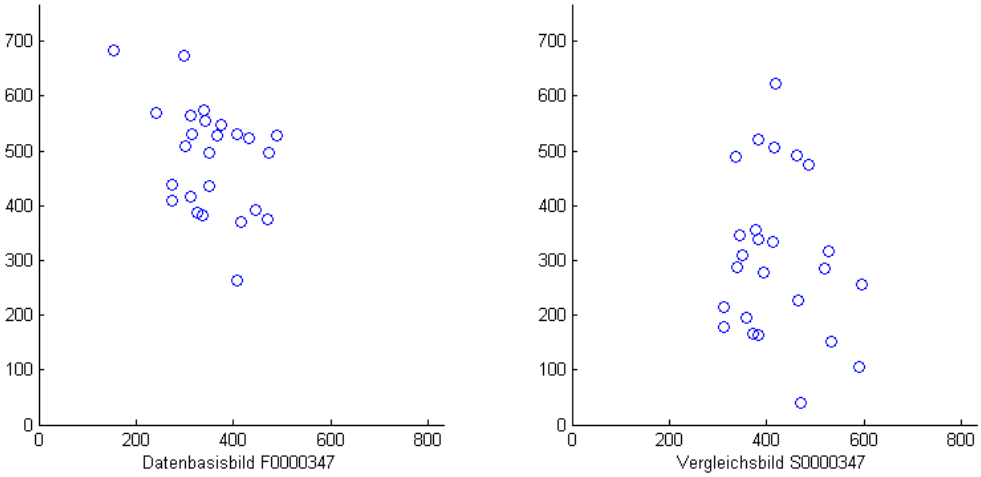

Abbildung 32: Pixelpositionen der übereinstimmenden Minutien; Mü = 25, Score-Wert = 0

5.4.3 Evaluierungsergebnisse

5.4.3.1 Testplan

Für die Evaluierung wurden alle linken Zeigefingerbilder der 2700 vorhandenen Personen der NIST SD14-Datenbank herangezogen – jeweils 2700 Enrolment-Bilder und 2700 Verifikations-Bilder. Da die Qualität der Fingerbilder nicht immer ausreichte, um Minutien zu extrahieren, standen 2676 Bilderpaare zur Verfügung, bevor diese mit Hilfe der NIST MINDTCT Quality Map gefiltert wurden (siehe Abschnitt 5.4.2). Aufgrund der verfahrensbedingten Verarbeitung auf Koordinatenbasis, konnten nur 954 Bilderpaare nach einem Genuine-Vergleich mit Hilfe des Neurotechnology Komparators (mit einem Score-Wert über 0) für die Evaluation herangezogen werden. Ohne die Filterung der Datenbasis mit Hilfe der MINDTCT Quality Map wurden ca. 200 Bilderpaare weniger gefunden. Dies zeigt, dass Ausreißer und schlechte Minutien eine erhebliche Auswirkung auf den Minutien-Komparator haben, wenn die Vergleiche nur auf Basis der (x,y)-Koordinaten durchgeführt werden.

Zur Evaluierung des Verfahrens wurden die in Tabelle 5 aufgelisteten Parameter herangezogen. Dabei wurde immer die Zahl der übereinstimmenden Minutien in einem Genuine- und Imposter-Vergleich gemessen, um anschließen *FMR* und *FNMR* zu berechnen. Ein Imposter-Vergleich wurde dabei so durchgeführt, dass ein Enrolment-Bild gegen zehn unabhängige Verifikationsbilder verglichen wurde, d.h. gegen zehn Verifikationsbilder verschiedener Nutzer. So wurden für die Berechnung der *FMR* und *FNMR* 954 Genuine- und 9540 Imposter-Vergleiche durchgeführt.

Ohne Template Protection und ohne Zusatzinformationen	
Parameter:	–
Template Protection ohne Zusatzinformationen	
Parameter:	CP = 150 EP = 20-50, 30, 50
Template Protection mit Zusatzinformationen	
Parameter:	CP = 150 EP = 20-50, 30, 50 (jeweils mit AP = 5, 10, 15, 20 bei jeweils 4-stelligem und 8-stelligem Passwort)

Tabelle 5: Parameter der einzelnen Evaluierungen

5.4.3.2 Testergebnisse – ohne Template Protection, ohne Zusatzinformationen

Für die Evaluierung wurden alle 954 Fingerbildpaare herangezogen. Die Performanz basiert auf der Anzahl der übereinstimmenden Minutien ($M_{ü}$) mit Hilfe des Neurotechnology Komparators. Ein Vergleich wird nur anhand von (x,y)-Koordinaten durchgeführt.

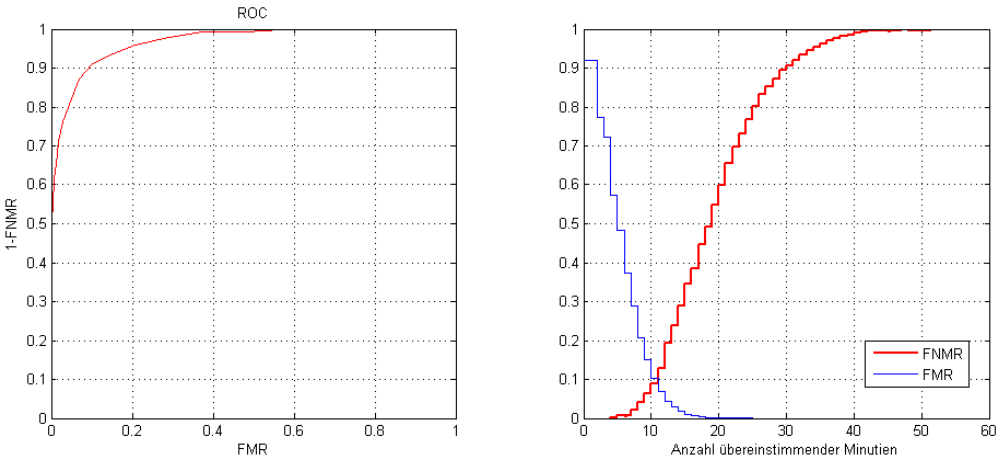

Abbildung 33: ROC-Kurve, False Match und False Non-Match Raten

Aufgrund der für die Verarbeitung nicht vorhandenen Winkel- und Typinformationen der einzelnen Minutien und der daraus resultierenden schlechteren Ergebnisse des Minutien-Komparators, kommt es zu einer schlechten Performanz. Die Equal Error Rate liegt bei elf übereinstimmenden Minutien bei einer Falschnichtübereinstimmungrate und Falschübereinstimmungsrate von etwa neun Prozent (siehe Abbildung 33). Die stufenförmige Darstellung der FMR- und FNMR-Kurven entsteht durch die diskrete Berechnung zu einer bestimmten Anzahl übereinstimmender Minutien.

Zum Vergleich ist in Abbildung 34 die Performanz mit allen Minutieninformationen dargestellt. Man kann eindeutig sehen, dass durch Hinzunahme von u.a. Winkel- und Typinformationen die Leistung gesteigert werden. So liegt die Equal Error Rate bei etwa fünf Prozent bei 44 übereinstimmenden Minutien.

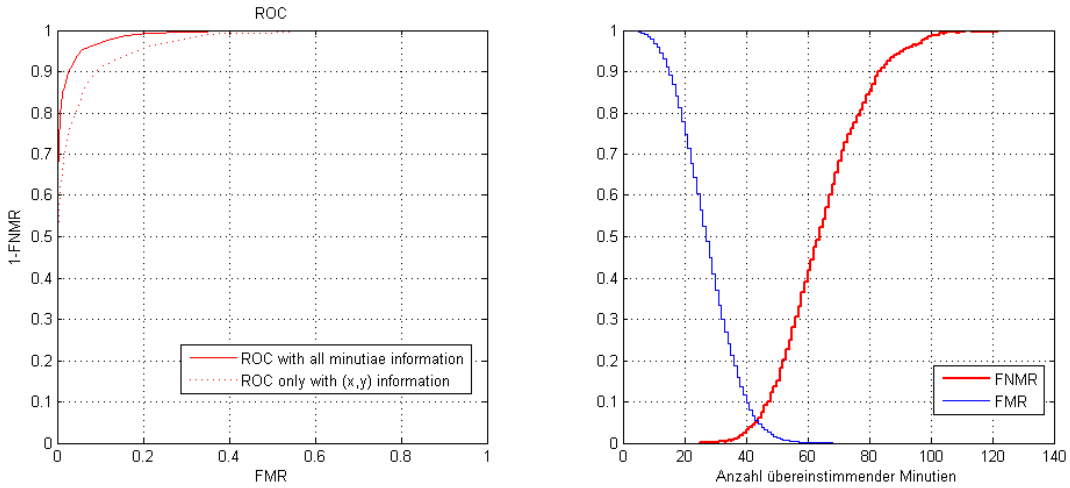

Abbildung 34: ROC-Kurve, FMR und FNMR unter Einbezug aller Informationen (Winkel, Typ,...)

5.4.3.3 Testergebnisse – Template Protection mit/ohne Zusatzinformationen

Für die Evaluierung wurden alle 954 Fingerbildpaare herangezogen. Die Performanz des Verfahrens ohne Zusatzinformationen basiert auf der Anzahl der Polynom-Koeffizienten ($k=d+1$), wobei diese in diesem Fall der Anzahl der übereinstimmenden Minutien ($M_{ü}$) entspricht. Ist die Anzahl der übereinstimmenden Minutien mindestens so groß wie die Anzahl der Koeffizienten ($M_{ü} \geq k$), kann das Polynom erfolgreich rekonstruiert werden.

Die Performanz des Verfahrens mit Zusatzinformationen basiert ebenfalls auf der Anzahl der Polynom-Koeffizienten k. In diesem Fall kann ein Polynom erfolgreich rekonstruiert werden, wenn mindestens k minus die Anzahl der künstlichen Minutien (AP) übereinstimmende Minutien gefunden werden ($M_{ü} \geq k\text{-}AP$).

Für eine bestimmte Anzahl von Polynom-Koeffizienten k gilt:

$$FMR = \frac{Anzahl\ erfolgreicher\ Rekonstruktionen\ der\ Imposter\ Vergleiche}{Anzahl\ aller\ Imposter\ Vergleiche}$$

$$FNMR = \frac{Anzahl\ nicht\ erfolgreicher\ Rekonstruktionen\ der\ Genuine\ Vergleiche}{Anzahl\ aller\ Genuine\ Vergleiche}$$

Alle Untersuchungen wurden mit einer Streupunkteanzahl von 150 durchgeführt ($CP = 150$).

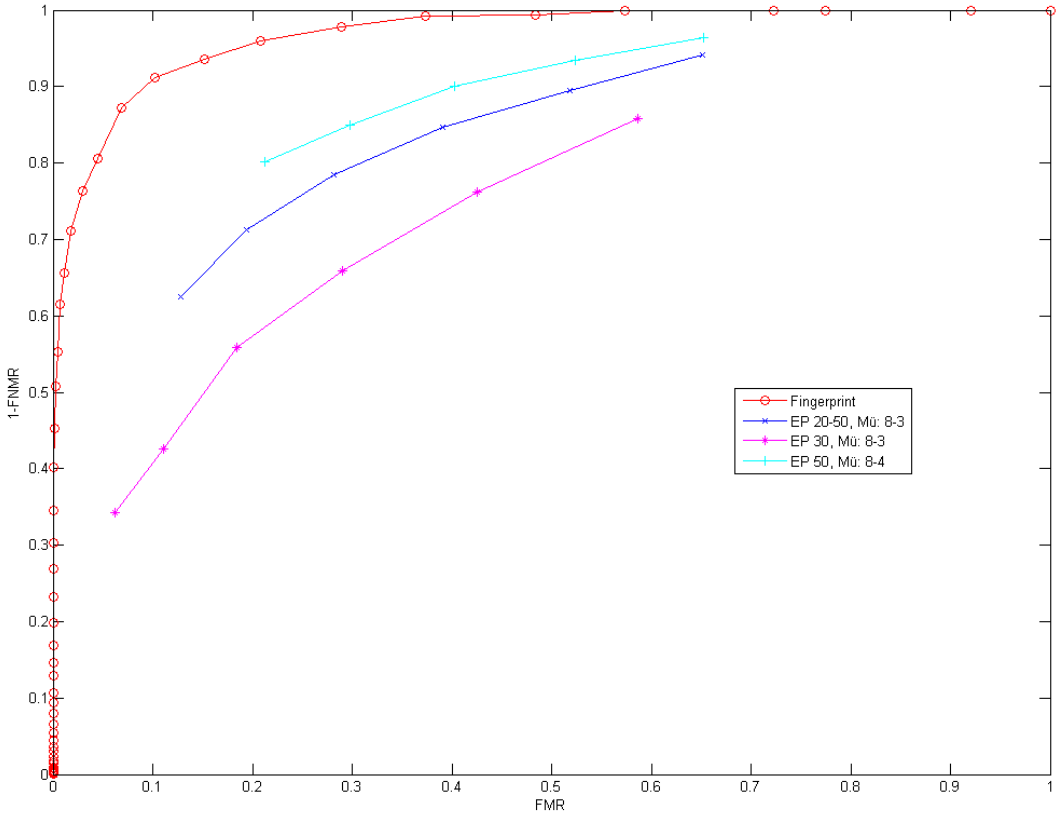

Abbildung 35: ROC-Kurven für das Fuzzy-Vault-Verfahren ohne Zusatzinformation

In Abbildung 35 können Ergebnisse des Fuzzy-Vault-Verfahrens ohne Zusatzinformation betrachtet werden (siehe dazu Tabellen 7, 9 und 11). Es ist (in den Kurven zu EP 20-50, EP 30 und EP 50) zu sehen, dass mit abnehmender Anzahl übereinstimmenden Minutien die FMR jeweils zunimmt. Bei gleichbleibender Anzahl übereinstimmenden Minutien und zunehmender Anzahl von Enrolment Punkten nimmt die FNMR ab. Dabei reiht sich die Evaluierung mit 20 bis 50 Enrolment Punkten zwischen den Untersuchungen mit 30 und 50 EP ein, da meist über 30 Enrolment Punkte in dieser Einstellung gefunden werden.

Die FMR beim Template-Protection-Verfahren mit Zusatzinformation bleibt immer Null. In Abschnitt 5.2.2 wurde gezeigt, dass $FMR(t) = FMR_{tp}(t) \cdot (2^{-Lp})$. Die erwartete *FMR* ist sehr klein. Für eine kleine *FMR* müssen viel mehr Imposter-Vergleich durchgeführt werden, um eine Falschakzeptanz beobachten zu können. Deswegen erscheint die *FMR* gleich null mit der aktuellen Anzahl der Imposter-Vergleich.

Die jeweilige *FNMR* bleibt bei gleichbleibender Anzahl von echten Minutien in etwa konstant (siehe Tabelle 8, 10 oder 12 bei bspw. fünf übereinstimmenden Minutien). Abweichungen ergeben sich durch die unterschiedlich generierten zufälligen Minutien (Chaff Points) und dem darauf aufbauenden Neurotechnology-Komparator. Da das Passwort nur einen Einfluss auf den Pseudo-Zufallszahlen-Generator hat und die *FNMR* von der Anzahl der künstlichen Minutien abhängt, sind die *FNMR*-Werte ebenfalls ähnlich zwischen 4-stelligem und 8-stelligem Passwort.

Die *FNMR* entspricht bei gleichbleibender Anzahl übereinstimmender Minutien und gleichbleibender Anzahl von Enrolment Punkten mit 4-stelligem und 8-stelligem Passwort annähernd dem Wert des Verfahrens ohne Zusatzinformation. Da jedoch das Passwort überprüft wird, ist nur bei dem Verfahren ohne Zusatzinformation eine FMR von größer null messbar.

Bei gleichbleibender Anzahl übereinstimmender Minutien und gleichbleibender Enrolment-Einstellung (EP), bleibt die FNMR ungefähr gleich. Jedoch ist das Verfahren mit Zusatzinformationen sicherer, da die Koeffizientenanzahl des Polynoms und der Aufwand der Polynomrekonstruktion ansteigt. So steigt beispielsweise die Koeffizientenanzahl bei 5 übereinstimmenden Minutien und 50 Enrolment-Punkten von 9 auf 14 und 19 auf 24, wenn jeweils 5 zusätzliche künstliche Minutien einbezogen werden. Die FNMR bleibt in diesen vier Fällen annähernd konstant.

Die Simulationsergebnisse zeigen, dass die FNMR mit der zunehmenden Anzahl der während des Enrolments verwendeten echten Minutien sinkt. Beispielsweise sinkt zwischen dem Enrolment mit 30 und 50 echten Minutien und bei gleichem Polynomgrad d=9 und gleicher Anzahl künstlicher Minutien (AP=5), die FNMR von ca. 30 Prozent auf 7 Prozent (4-stelliges Passwort). Dazwischen befindet sich das Enrolment-Setting mit mindestens 20 und maximal 50 echten Minutien, welches eine FNMR von 14 Prozent zur Folge hat. Die Performanz ist schlechter als mit 50 EP, da auch Fingerbilder herangezogen worden sind, die weniger als 50 EP besitzen. Jedoch steigt auch mit ansteigender EP (siehe Tabelle 6) die Failure-to-Enrol-Rate.

Anzahl Enrolment Points	***Failure-to-Enrol-Rate***
mindestens 20 bis maximal 50	1,05%
genau 30	7,34%
genau 50	45,18%

Tabelle 6: FTE-Raten der Datenbasis in Abhängigkeit der Enrolment Points

Wegen der schlechten Qualität der Fingerabdruckbilder ist die Übereinstimmungsrate der gefundenen Minutien sehr klein. Da nur ein Enrolment-Bild pro Fingerabdruck vorhanden ist, ist es schwer, zuverlässige Minutien zu detektieren. Aus diesem Grund müssen viel mehr Minutien beim Enrolment benutzt werden, um eine ausreichende Erkennungsleistung zu erreichen. Bedauerlicherweise wird hierdurch die Sicherheit stark reduziert. Bei $CP=150, d=9$ reduziert sich die erwartete Anzahl der Versuche, um das Polynom zu rekonstruieren, von 4902 bei $EP=30$ auf 330 bei $EP=50$.

Anzahl echter Minutien: EP = mindestens 20 bis maximal 50

$M_ü$	d	FMR	FNMR
4	3	0,518750	0,105930
5	4	0,390780	0,153600
6	5	0,281140	0,216100

Tabelle 7: Verfahren ohne Zusatzinformationen (EP20-50)

			Passwortlänge 4		Passwortlänge 8	
AP	$M_ü$	d	FMR	FNMR	FMR	FNMR
5	4	8	0	0,094280	0	0,101690
5	5	9	0	0,146190	0	0,153600
5	6	10	0	0,205510	0	0,219280
10	4	13	0	0,099576	0	0,091102
10	5	14	0	0,157840	0	0,156780
10	6	15	0	0,218220	0	0,211860
15	4	18	0	0,097458	0	0,105930
15	5	19	0	0,151480	0	0,154660
15	6	20	0	0,212920	0	0,213980
20	4	23	0	0,095339	0	0,098517
20	5	24	0	0,150420	0	0,151480
20	6	25	0	0,204450	0	0,210810

Tabelle 8: Verfahren mit Zusatzinformationen (EP20-50)

Anzahl echter Minutien: EP = genau 30

$M_ü$	d	FMR	FNMR
4	3	0,425790	0,238690
5	4	0,290050	0,340500

Tabelle 9: Verfahren ohne Zusatzinformationen (EP30)

			Passwortlänge 4		Passwortlänge 8	
AP	$M_ü$	d	FMR	FNMR	FMR	FNMR
5	4	8	0	0,205880	0	0,203620
5	5	9	0	0,299770	0	0,315610
10	4	13	0	0,190050	0	0,213800
10	5	14	0	0,297510	0	0,319000
15	4	18	0	0,228510	0	0,205880
15	5	19	0	0,313350	0	0,313350
20	4	23	0	0,197960	0	0,201360
20	5	24	0	0,302040	0	0.30882

Tabelle 10: Verfahren mit Zusatzinformationen (EP30)

Anzahl echter Minutien: EP = genau 50

$M_ü$	d	FMR	FNMR
5	4	0,523330	0,065010
6	5	0,402870	0,099426
7	6	0,297320	0,151050

Tabelle 11: Verfahren ohne Zusatzinformationen (EP50)

			Passwortlänge 4		Passwortlänge 8	
AP	$M_ü$	d	FMR	FNMR	FMR	FNMR
5	5	9	0	0,070746	0	0,051625
5	6	10	0	0,112810	0	0,080306
5	7	11	0	0,160610	0	0,151050
10	5	14	0	0,072658	0	0,059273
10	6	15	0	0,110900	0	0,107070
10	7	16	0	0,151050	0	0,164440
15	5	19	0	0,051625	0	0,059273
15	6	20	0	0,105160	0	0,108990
15	7	21	0	0,152960	0	0,141490
20	5	24	0	0,063098	0	0,070746
20	6	25	0	0,084130	0	0,103250
20	7	26	0	0,160610	0	0,149140

Tabelle 12: Verfahren mit Zusatzinformationen (EP50)

5.4.3.4 Laufzeitangaben zur Integration von Zusatzinformation

Für die Umsetzung des Verfahrens zur Integration von Zusatzinformationen wurde ein Intel Core 2 Duo P9700 mit 2,80 GHz genutzt, welcher mit 4 GB Arbeitsspeicher bestückt war. Die verwendete

MATLAB-Version war 7.8.0 (2009a). Als Betriebssystem kam Windows XP mit Service Pack 3 zum Einsatz.

Die Integration und die Verifikation hängen von den verwendeten Parametern ab. So haben sowohl die Anzahl der echten Minutien, der künstlichen Minutien und der Chaff Points Einfluss auf die Laufzeit, als auch der Grad des Polynoms. Weiterhin beeinflussen die zufälligen Positionen der Chaff Points den Minutien-Komparator, da dieser aufgrund der wenigen Informationen (nur x- und y-Koordinaten) zu falschen Übereinstimmungen führen kann. So kann es trotz ausreichender Anzahl übereinstimmender Minutien-Paare zwischen Enrolment- und Verifikationsbild zu einer nicht erfolgreichen Polynomrekonstruktion bzw. Verifikation kommen. Auch die Anzahl der übereinstimmenden Minutien-Paare wirkt sich stark auf die Polynomrekonstruktion aus.

Exemplarisch werden folgend ein paar Laufzeitmessungen angegeben. Hierzu wurde folgendes Setup genutzt:

- Anzahl echter Minutien: mind. 20 max. 50
- Anzahl künstlicher Minutien: 5
- Anzahl Chaff Points: 150
- Abstand der Minutien: 12 Pixel
- Grad des Polynoms: 9

ID	*Passwortlänge*	*Passwort*	*Enrolmentzeit*	*Verifikationszeit*	*Anzahl übereinstimmender Minutien*
1	4	1234	0,606647 s	1,337140 s	5
2	4	5042	0,587675 s	14,190649 s	8
3	8	03845759	0,592456 s	2,179227 s	10
4	8	73966590	0,593533 s	1,159286 s	9

5.5 Zusammenfassung

In diesem Kapitel wurden verschiedene Möglichkeiten vorgestellt, um Template-Protection-Verfahren mit Zusatzinformationen zu versehen. So wurden Verfahren beschrieben, die Zusatzinformationen durch Verlängerung der geheimen Zeichenkette oder durch Verlängerung des biometrischen Merkmals einbetten. Weiterhin wurde eine Methode aus der Literatur vorgestellt, die eine Härtung des Fuzzy Vaults aufzeigt. Nach einer Analyse zeigte sich, dass das zweite Verfahren das beste Potenzial im Rahmen dieses Projekts bietet, da sowohl die Sicherheit als auch die Erkennungsleistung des Systems verbessert werden kann. Darüber hinaus sind umfangreichere Einstellmöglichkeiten für Falschübereinstimmungsrate und Falschnichtübereinstimmungsrate gegeben und die Abwägung zwischen Sicherheit und Leistungsfähigkeit wurde vereinfacht. Die ausgewählte Methode nutzt dabei die Zusatzinformationen, um zusätzliche künstliche biometrische

Merkmale zu generieren und wurde im Rahmen dieses Kapitels näher beschrieben und evaluiert. Die Ergebnisse zeigen, dass durch die Nutzung von Zusatzinformationen (hier 4- oder 8-stellige Passwörter) die Falschübereinstimmungsraten stark reduziert werden können. Außerdem kann der Grad des Polynoms im Fuzzy-Vault-Verfahren vergrößert werden, ohne die Anzahl der übereinstimmenden Minutien zu vergrößern. Die Auswertungen der Evaluierungen zeigen auch, dass die Falschnichtübereinstimmungsraten von der Anzahl der übereinstimmenden Minutien bestimmt werden. In den vorgenommenen Experimenten zeigte sich, dass bei einem Enrolment mit 50 echten Minutien die besten Erkennungsraten bei acht übereinstimmenden Minutien erzielt werden konnten.

Neben der allgemeinen Evaluierung wurde auch die Sicherheit des neuen Verfahrens überprüft. Es zeigte sich, dass neben dem Aufwand der Polynomrekonstruktion auch der Aufwand das Passwort richtig zu schätzen die Sicherheit erhöht. Ebenfalls wurde gezeigt, dass das Verfahren eine verbesserte Resistenz gegen Verknüpfungsangriffe bietet.

Die für die Evaluierung verwendete Datenbank NIST SD14 enthält für jeden Finger nur zwei Fingerbilder, wobei ein Großteil dieser eine schlechte Qualität aufweisen. Aus diesem Grund musste für das Enrolment eine hohe Anzahl von Minutien herangezogen werden, die wesentlich größer als die Anzahl der Koeffizienten des Fuzzy-Vault-Polynoms waren. Dies spiegelt sich in einer schlechten Leistungsfähigkeit und Sicherheit wieder. So bietet das Verfahren noch Potenzial zur Verbesserung, wenn eine bessere Datenbank zur Evaluierung bzw. Praxiseinsatz herangezogen wird. Auch die Umsetzung einzelner Teile des Verfahrens können noch optimiert werden. Da der Umgang mit Verknüpfungsangriffe noch nicht vollständig gelöst ist, bietet das Verfahren noch Potenzial die Sicherheit des Systems weiter zu verbessern.

Zusammenfassend kann gesagt werden, dass das Verfahren zur Verlängerung der biometrischen Merkmale durch Zusatzinformationen die Sicherheit und Leistungsfähigkeit steigern kann. Jedoch ist es nur für biometrische Verifikationen und nicht für Identifikationen geeignet.

6 Template Protection in Identifikationssystemen

Dieses Kapitel untersucht die Anwendung von Identifikationsverfahren im Kontext zentraler Datenbanken zur Speicherung geschützter biometrischer Referenzdaten. Der Schutz der hoch sensiblen Biometriedaten wird dabei durch das in Abschnitt 3.2.2.2 beschriebene Template Protection Verfahren Fuzzy Vault [JS02] gewährleistet. Da Template Protection Verfahren den Vergleich zweier biometrischer Referenzdaten im Allgemeinen sehr erschweren, werden verschiedene Ansätze zur effizienten Unterstützung des Identifikationsprozesses in großen Datenbanken analysiert. Bei den betrachteten biometrischen Referenzdaten handelt es sich dabei um Fingerabdrücke und daraus extrahierte Minutienmengen.

Im Rahmen des Projekts BioKeyS-Multi [KMN09] wurde bereits eine praktikable Methode für Minutien-basierte Fingerabdruckerkennung implementiert. Die Authentifikation einer Person erfolgt dabei ebenfalls auf Basis von entsprechend verschleierten Referenzdaten der Anwender. Eine naive Übertragung dieses Ansatzes auf die Aufgabe der Identifikation bewirkt im schlechtesten Fall ein Durchsuchen der gesamten Datenbank, bis eine Entscheidung getroffen werden kann. Die im Folgenden vorgeschlagenen Ansätze sollen nun dazu dienen diesen Prozess dahingehend zu beschleunigen, dass die zu untersuchenden Datenbankreferenzen entsprechend einer gewissen Präferenz betrachtet werden.

6.1 Fragestellung der Identifikationslösung

Wie im Abschnitt 3.2.2.2 über Fuzzy Vault bereits ausführlich beschrieben wurde, bestehen für die Speicherung von Biometriedaten durch ihre Persistenz strenge Sicherheitsanforderungen. Aus diesem Grunde werden biometrische Daten wie Fingerabdrücke sowie daraus extrahierten Merkmale wie beispielsweise Minutien vor der Speicherung kodiert. Das in Abschnitt 3.2.2.2 beschriebene Template Protection Verfahren Fuzzy Vault erzeugt zu diesem Zweck zufälliges Rauschen (sogenannte Chaff-Points), welches den ursprünglichen Daten hinzugefügt wird. Abbildung 36 zeigt die Prozesskette der Datenaufbereitung eines Fingerabdrucks bis zur Speicherung in der Datenbank. Die so aufbereiteten Daten werden im Rahmen dieses Kapitels als Templates bezeichnet. Im Allgemeinen wirkt eine derartige Kodierung der Datenbanktemplates allerdings auch dem Prozess des direkten Vergleichs mit einem Anfragetemplate entgegen. Bisherige Verfahren zum Vergleich zweier Fingerabdrücke basierend auf Minutien, wie bspw. [JPH+99] sind auf Grund des starken Rauschens bei derartig kodierten Daten nicht anwendbar.

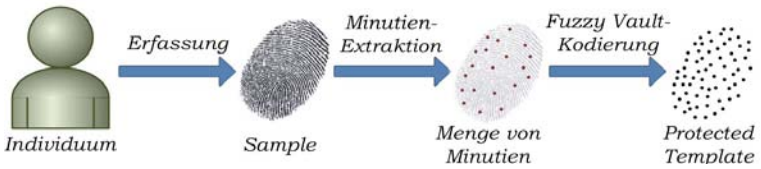

Abbildung 36: Generierung eines geschützten Templates für den Identifikationsprozess

In der Publikation von Korte et al. [KMN09] wurde bereits ein Verfahren zum Abgleich eines ungeschützten Anfragetemplates mit einem geschützten Datenbanktemplate vorgestellt. Im Rahmen des Projektes BioKeyS_Multi wurde dieses Verfahren für die Problemstellung der Authentifikation eingesetzt. Die Identifikation eines Datenbanktemplates erfolgt hier über Zusatzinformationen wie bspw. eine PIN (Abbildung 37).

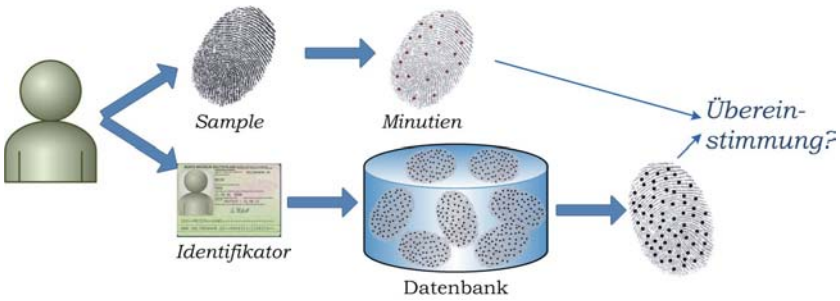

Abbildung 37: Prozess der Authentifikation

Eine naive Übertragung eines Authentifikationsansatzes auf die Identifikation entspricht einer Authentifikationsanfrage für jedes Datenbanktemplate, welche der Reihe nach bis zur Feststellung einer Übereinstimmung abzuarbeiten sind. Im schlimmsten Fall entspricht die Anzahl der gestellten Authentifikationsanfragen dabei der Datenbankgröße (Abbildung 38 oben).

Um Identifikationsanfragen auf großen Datenbanken in akzeptabler Zeit zu beantworten, ist daher neben einem effizienten Authentifikationsprozess zudem die Anzahl der zu überprüfenden Datenbanktemplates geeignet einzuschränken. Indexstrukturen sowie Filterarchitekturen ermöglichen eine entsprechende Vorauswahl von Datenobjekten durch verschiedene Approximationstechniken (Abbildung 38 unten).

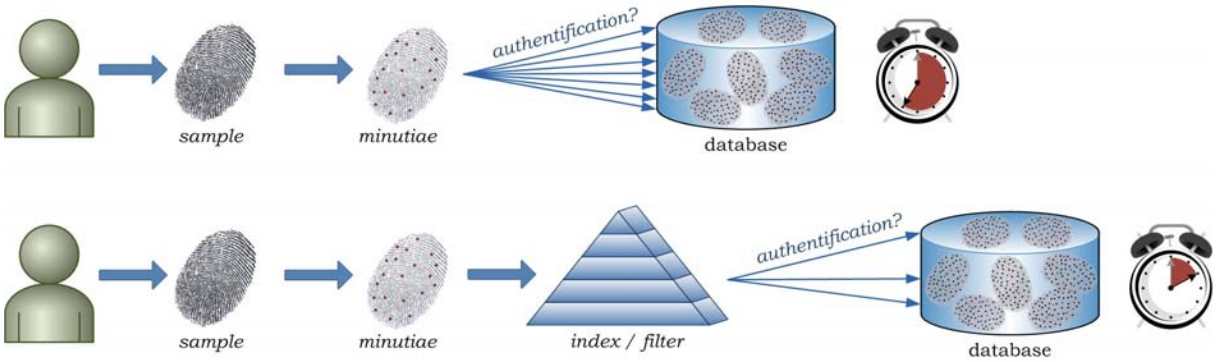

Abbildung 38: Prozess der Identifikation: naiv (oben) und beschleunigt (unten)

Da auf Grund der durch die Kodierung hinzugefügten Unsicherheit über die Daten nahezu keine Objekte mit Sicherheit für eine Übereinstimmung ausgeschlossen werden können, verfolgen die im Folgenden vorgestellten Ansätze das Ziel, eine Rangfolge der Datenbankreferenzen zu erstellen, sodass tendenziell ähnlichere Referenzen für den Authentifikationsvorgang priorisiert werden. Eine zu einer Identifikationsanfrage passende Datenbankreferenz wird somit schneller behandelt, wodurch sich im Allgemeinen eine deutliche Reduktion der Anzahl durchzuführender Authentisierungsanfragen erreichen lässt.

6.2 Identifikationslösung

Die nun folgenden Identifikationslösungen dienen allesamt dazu, dem Nutzer eine bestimmte Reihenfolge der gespeicherten Datenbankreferenzen anzubieten, in der anschließend der Authentifikationsabgleich durchgeführt werden soll. Dabei kommt es darauf an, dass trotz vielfältiger Approximationstechniken die tatsächliche Datenbankreferenz in dieser Rangliste eine möglichst frühe Position einnimmt.

Im Allgemeinen liegen für die extrahierten Minutien eines Fingerabdrucks eine Vielzahl von Informationen, wie beispielsweise die Richtung, der Typ oder gegebenenfalls der Ridge-Count, vor. Im Folgenden wird für die Minutieninformationen eine Beschränkung auf die Ortskoordinate vorgenommen. Ausgangssituation für die Verfahren ist somit eine Menge von Datenbankreferenzen $(R \in DB)$ (im Folgenden Datenbanktemplates genannt) sowie eine Anfrageprobe (Q) (im Folgenden Anfragetemplate genannt), welche jeweils eine Menge von 2-dimensionalen Objekten $(m=(m_x, m_y))$ beinhalten. Im Falle des Anfragetemplates (Q) handelt es sich bei diesen 2-dimensionalen Objekten ausschließlich um Minutien. Im Gegensatz dazu enthält ein Datenbanktemplate $(R \in DB)$ neben den Minutien zusätzlich eine große Menge an zufällig eingestreuten Chaff-Points, sodass dieses im Allgemeinen um einen gewissen Faktor *k* größer ist, als ein Anfragetemplate. Für jedes erfasste Subjekt *S* existieren Templates zu allen 10 Fingern $(S_i = \{R_{i,1}, \ldots, R_{i,10}\})$. Da für jedes Template der zugehörige Fingertyp $(T = \{1,\ldots,10\})$ bekannt ist, kann angenommen werden, dass alle Fingertypen $(t \in T)$ separat verwaltet werden $(\forall\, i, \forall\, t \in T, \forall\, t' \in T \setminus \{t\}: R_{i,t} \in DB_t \wedge R_{i,t} \notin DB_{t'})$.

6.2.1 Ansatz GeoMatch

Bei dem Verfahren GeoMatch handelt es sich um eine Filterstruktur, welche lediglich approximierte Ähnlichkeiten zwischen einem Anfragetemplate und einem Datenbanktemplate berechnet. Um gegen globale Rotationen sowie Verschiebungen des Anfragesamples robust zu sein, werden bei diesem Ansatz keine globalen Lagepositionen der Minutien für den Vergleich verwendet sondern lediglich relative Positionen der Minutien zueinander. Für diese relativen Lagepositionen eignen sich geometrische Figuren wie *n*-Ecke, durch welche die relative Anordnung von n Punkten zueinander eindeutig beschrieben wird.

Dieser Ansatz sieht es vor aus den jeweils zu vergleichenden Templates *n*-Ecke zu extrahieren und anschließend einen Vergleich lediglich basierend auf diesen *n*-Ecken durchzuführen (vergleiche Abbildung 39). Werden genügend viele n-Ecke eines Anfragetemplates im Vergleichsobjekt der Datenbank wiedergefunden, so kommt das betreffende Datenbanktemplate für den genauen Abgleich in Frage. Je größer die n-Eckübereinstimmug ist, desto wahrscheinlicher ist eine tatsächliche Templateübereinstimmung und entsprechend höher ist daher eine vozunehmende Priorisierung für den genauen Abgleich.

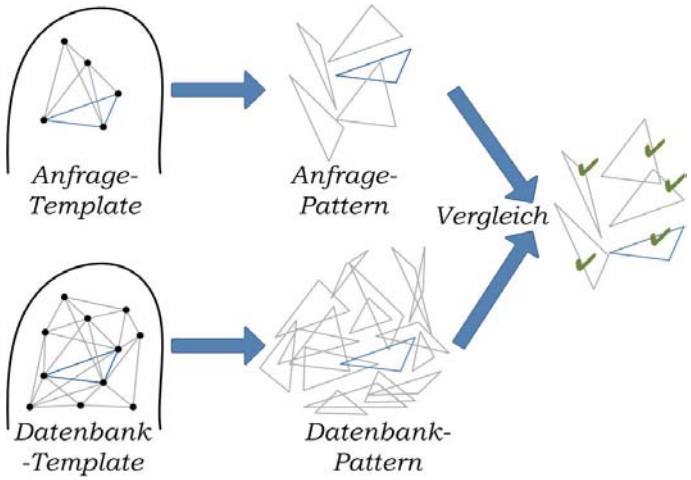

Abbildung 39: Generelle Idee des GeoMatch Ansatzes anhand von Dreiecken

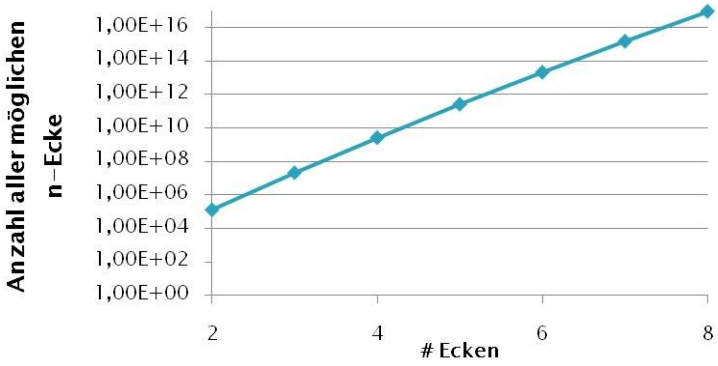

Abbildung 40: Anzahl aller n-Ecke für festes m=400

Die Anzahl aller möglichen *n*-Ecke in einem Template mit *m* 2-dimensionalen Objekten beträgt $\binom{m}{n} = \frac{m!}{n!*(m-n)!}$. Wie der Graph in Abbildung 40 zeigt, wächst diese Anzahl bei konstantem *m* exponentiell in *n*. Um den Platzbedarf für ein Datenbankpattern, sowie die benötigten Vergleichsoperationen gering zu halten, sollte *n* daher eher klein gewählt werden. Für steigendes *n* besitzt ein *n*-Eck indessen eine wesentlich Beschreibungsgenauigkeit, was tendenziell ein größeres n befürwortet. Für diesen Ansatz wurden Dreiecke (n=3) gewählt, welche sich mit 3 Werten, z.B. zwei Seitenlängen und einem Winkel, exakt beschreiben lassen. Ein Anfrage- sowie ein Datenbankpattern ist somit eine Menge von Dreiecken, wobei jedes Dreieck wiederum ein dreielementiges Tupel von Double-Werten ist.

Für einen exakten Vergleich zweier Templates ist eine diskriminative Beschreibung entsprechend der eines vollständigen Graphen für jedes Template ausreichend. Um einen vollständigen Graphen zu rekonstruieren sind jedoch nicht alle $\binom{m}{n}$ vorkommenden n-Ecke notwendig. Da durch diesen Ansatz ferner kein exakter Vergleich, sondern lediglich ein approximativer Vergleich angestrebt wird, kann die Menge der Dreiecke im Anfrage-, sowie im Datenbank-Pattern im Allgemeinen stark eingeschränkt werden. Dies ist z.B. durch Festlegen einer oberen, sowie einer unteren Schranke für die Seitenlängen eines Dreiecks möglich.

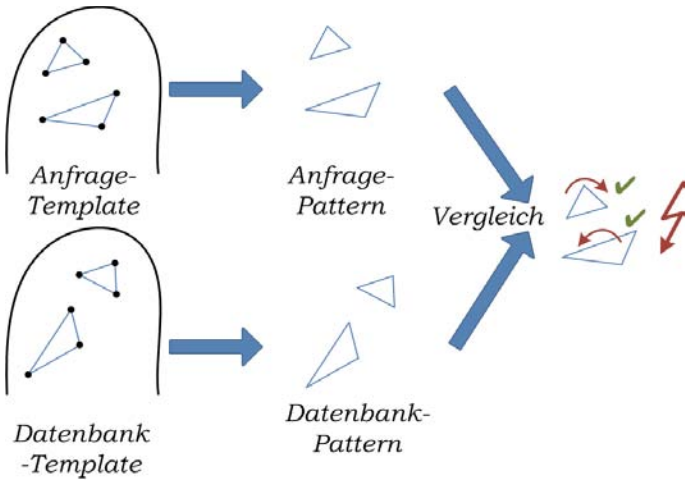

Abbildung 41: Unbeabsichtigte Robustheit gegenüber lokalen Rotationen

Da für die Dreiecke lediglich die geometrisch-relevanten Merkmale gespeichert werden und keine globalen Lageinformationen, sind die so erstellten Anfrage- und Datenbankpatterns offensichtlich robust gegen globale Verschiebung sowie Rotation. Für den Vergleich zweier Patterns können jedoch die Dreiecke beliebig rotiert sein (siehe Abbildung 41). Um diese Robustheit gegen 'lokale' Rotationen aufzuheben, wird zusätzlich die globale Orientierung der n-Ecke mitberücksichtigt. Dazu wird, wie in dem kleinen Beispiel aus Abbildung 42 gezeigt, zu jedem der Dreiecke der Winkel zwischen kürzester Seite und der Horizontalen gespeichert. Anstatt also lediglich zu vergleichen, wie viele Dreiecke des Anfrage-Patterns in einem Datenbank-Pattern wiedergefunden werden, kann so verglichen werden, wie viele Dreiecke mit gleicher Orientierung wiedergefunden werden.

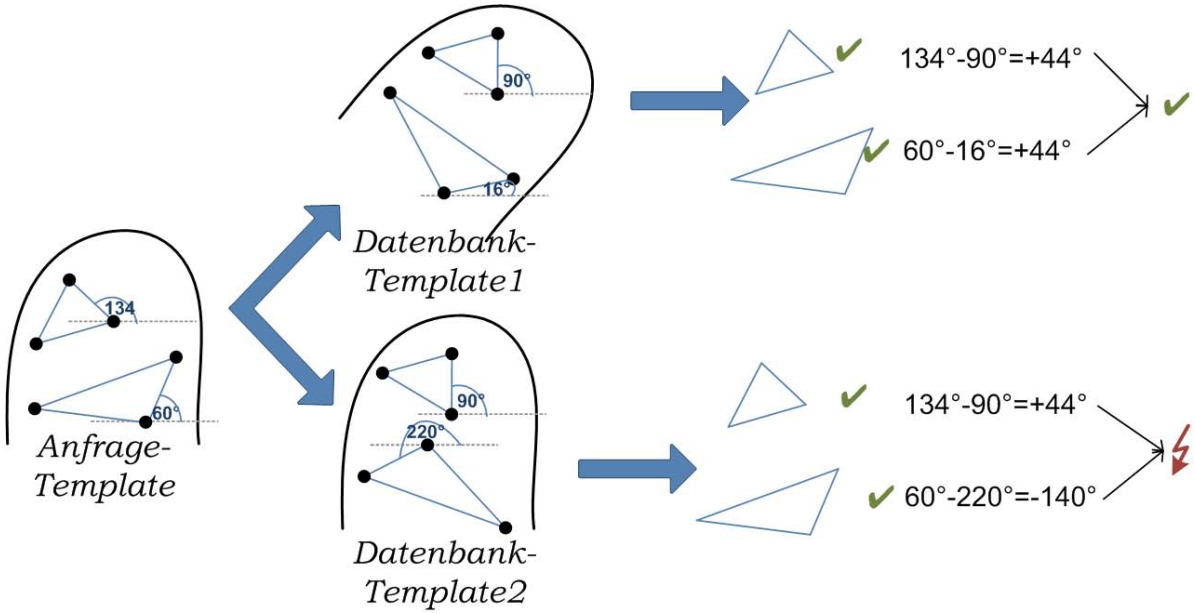

Abbildung 42: Einbezug der globalen Orientierung der n-Ecke mittels des Winkels zwischen kürzester Seite und der Horizontalen

Neben der höheren Aussagekraft eines derartigen Patternvergleichs bietet dieses Verfahren den Vorteil, dass zusätzlich die unterstellte globale Rotation des Anfragetemplats bekannt ist, welches den nachgelagerten Authentifikationsvergleich vereinfacht.

Ausgangssituation für das Verfahren GeoMatch ist die Repräsentation aller Datenbanktemplates $(R \in DB)$, sowie Anfragetemplates (Q) als eine Menge $T = \{t_1, \ldots t_l\}$ von Dreiecken (für welche ggf. gewissen Beschränkungen der Seitenlängen bestehen).

Für jedes Dreieck t_j werden die drei Seitenlängen sowie ein Winkel gespeichert. Die Seitenlängen liegen dabei sortiert vor. Für ein Dreieck $t_j = (s_{j1}, s_{j2}, s_{j3}, \alpha_j)$ wird zunächst die Kürzeste der drei Seiten gespeichert, die verbleibenden zwei Seiten werden anschließend im Uhrzeigersinn aufgezählt. Für jedes Dreieck wird zudem der Winkel α_j zwischen kürzester Seite s_{j1} und der Horizontalen gespeichert (Abbildung 43).

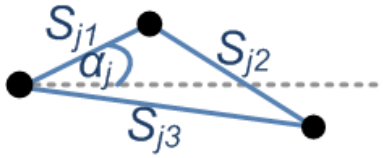

Abbildung 43: Gespeicherte Maße eines Dreiecks t_j

```
Eingabe: Datenbankpatterns T(DB), Anfragetemplate Q, Toleranzparameter δ

 Build Anfragepattern T_Q
 Build Rankingvector rank[1,...,|DB|]
 angle_person = new int[360]
 angle_triangle = new int[360]
 ranking = new int[|DB|]
 FOR EACH T_R in T(DB)
   reset(angle_person)
   FOR EACH t_Q in T_Q
     reset(angle_triangle)
     FOR EACH t_R in T_R
       IF (t_Q ~_δ t_R) THEN angle_triangle[ Δα_{Q,R} ] = 1
     angle_person.add(angle_triangle)
   rank[R.id] = maximum(angle_person)
```

Codelisting 1: Algorithmischer Ablauf des Rankingverfahrens basierend auf GeoMatch

Mit Hilfe des GeoMatch-Ansatzes ist es nun möglich approximierte Ähnlichkeiten zwischen allen Datenbankpatterns und einem Anfragepattern zu berechnen und basierend auf diesen ein Ranking aller Datenbanktemplates zu erstellen. Das generelle Vorgehen dieses Verfahrens ist in Codelisting 1 dargestellt.

Für den Vergleich zweier Patterns Q, R werden für alle Dreiecke $t_Q \in T_Q \wedge t_R \in T_R$ die drei gespeicherten Seitenlängen miteinander verglichen. Da im Allgemeinen kleinere lokale Abweichungen der Minutienpositionen toleriert werden sollten, wird zwischen beiden Dreiecken keine exakte Übereinstimmung getestet, sondern lediglich ihre Ähnlichkeit bewertet. Zwei Dreiecke gelten dabei als ähnlich $(t_Q \sim_\delta t_R)$, wenn die einander entsprechenden Seiten keine Differenz größer als δ aufweisen: $t_Q \sim t_R \Leftrightarrow dist_\infty(t_Q, t_R) \leq \delta$, wobei die Maximumsnorm $dist_\infty(t_Q, t_R)$ definiert ist als: $dist_\infty(t_Q, t_R) = max\{|s_{Qk} - s_{Rk}| \; | k \in \{1,2,3\}\}$. Sind zwei Dreiecke ähnlich zueinander, so wird zudem die Differenz $\Delta\alpha_{Q,R}$ ihrer Winkel berechnet: $\Delta\alpha_{Q,R} = dist(\alpha_Q, \alpha_R)$. Für jedes Dreieck $t_q \in Q$ des Anfragepatterns wird auf diese Weise eine Menge $\Delta t_Q = \{\Delta\alpha_{Q,R} | t_Q \sim t_R\}$ der vorkommenden Rotationsdifferenzen der ähnlichen Dreiecke erstellt. Im Anschluss lässt sich für jeden Winkel $0 \leq w_i \leq 360°, w_i \in \mathbb{N}$ bestimmen, in wie vielen der Mengen Δt_Q dieser Winkel enthalten ist. Die Anzahl für den am häufigsten auftretenden Winkel stellt die Gewichtung für die Priorisierung für den anschließenden Authentifikationsvergleich dar. Je häufiger eine Rotationsdifferenz wiedererkannt wurde, desto wahrscheinlicher liegt eine globale Rotation des gleichen Templates vor.

6.2.1.1 Physische Datenbankstruktur

Für das Verfahren GeoMatch werden aus allen Datenbanktemplates die entsprechenden Dreiecke extrahiert und als Datenbankpatterns gespeichert. Dieser Prozess wird einmal für den gesamten Datenbestand durchgeführt und in einer Datenbank gespeichert, sodass alle Datenbankpatterns zur Anfragezeit zur Verfügung stehen. Für das Anfragetemplate wird dieser Prozess erst zur Anfragezeit durchgeführt. Das zum Vergleich eingesetzte Pattern hat die in Tabelle 14 dargestellte Struktur. Für ein Pattern wird die PatternID sowie eine Liste von Dreiecken gespeichert. Für jedes Dreieck werden die drei Seitenlängen (sortiert) sowie ein Winkel gespeichert. Für ein Dreieck t_j wird zunächst die Kürzeste der drei Seiten gespeichert, die verbleibenden zwei Seiten werden

anschließend im Uhrzeigersinn aufgezählt. Für jedes Dreieck wird zudem der Winkel α_j zwischen kürzester Seite und Horizontalen gespeichert (Abbildung 43). Insgesamt ist die Liste aller Dreiecke aufsteigend nach der Länge der kürzesten Seite sortiert. Auf diese Weise kann für die Suche nach ähnlichen Dreiecken die binäre Suche genutzt werden und so der Anfrageprozess beschleunigt werden.

PatternID				
triangle	*shortest side*	*side two*	*side three*	*angle*
t_1	s_{11}	s_{12}	s_{13}	α_1
t_2	s_{21}	s_{22}	s_{23}	α_2
⋮	⋮	⋮	⋮	⋮
t_i	s_{i1}	s_{i2}	s_{i3}	α_i

Tabelle 13: Datenstruktur eines Anfrage- sowie Datenbank-Patterns

6.2.2 Matrix-Comparator

Bei dem Verfahren Matrix-Comparator handelt es sich um ein Distanzmaß zur Bestimmung der Ähnlichkeit eines Anfragetemplates *Q* mit einem Datenbanktemplate *R*, wobei Letzteres neben den wahren Minutien zusätzlich eine Menge von Streupunkten, die Chaff-Points, enthält.

Die Grundidee des Verfahrens besteht darin, den Vergleich beider Templates basierend auf den zugehörigen Distanzmatrizen durchzuführen. Zu diesem Zweck werden sowohl für das Anfragetemplate *Q* als auch für das Datenbanktemplate *R* je eine Distanzmatrix mit den paarweisen Distanzen der in einem Template enthaltenen Minutien berechnet. Anschließend werden die paarweisen Distanzen des Anfragepatterns für *Q* in der Distanzmatrix des Datenbankpatterns *R* gesucht. Eine Distanzmatrix enthält die paarweisen Abstände aller 2-dimensionalen Objekte eines Templates (Minutien, bzw. Chaff-Points) und hat die folgende Form:

	m_1	m_2	...	m_n
m_1	$D_{1,1}$	$D_{1,2}$		$D_{1,n}$
m_2	$D_{2,1}$	$D_{2,2}$		$D_{2,n}$
...				
m_n	$D_{n,1}$	$D_{n,2}$		$D_{n,n}$

Tabelle 14: Datenstruktur eines Anfrage- und Datenbank - Patterns

Die Werte $D_{i,j}$ stehen dabei für den euklidischen Abstand zweier 2-dimensionaler Objekte m_i und m_j. Der euklidische Abstand für zwei 2-dimensionale Punkte $x=(x_1, x_2)$ und $y=(y_1,y_2)$ ist definiert als:

$$eucl(x, y)=\sqrt{(x_1-y_1)^2+(x_2-y_2)^2}$$

Für das Anfragetemplate Q sowie das Datenbanktemplate R ergibt sich somit jeweils eine Distanzmatrix, wobei die Distanzmatrix von R aufgrund der zusätzlichen Chaff-Points größer als die von Q ist.

6.2.2.1 Ähnlichkeit von Matrizen: Grundidee

Im Folgenden bezeichne D^L die Distanzmatrix für ein Template L, D_i^L bezeichnet die i-te Zeile der Distanzmatrix für L, $|D^L|$ steht für die Anzahl der Zeilen/Spalten in D^L und $D_{i,j}^L$ bezeichnet die Einträge in der Matrix D^T an der Position i,j.

Die Ähnlichkeit *similarity(Q, R)* zwischen Q und R wird nun anhand der Anzahl der Zeilen aus D^Q berechnet, welche zu einem vorgegebenen Anteil β auf die Zeilen aus D^R passen. Eine Zeile D_i^Q passt dabei genau dann auf eine Zeile D_k^R, falls $\beta \cdot |D^Q|$ Werte aus D_i^Q auf Werte aus D_k^R passen. Ein Eintrag D_{ij}^Q aus D_i^Q passt genau dann auf einen Eintrag D_{kl}^R aus D_k^R, wenn die Abweichung $|D_{i,j}^Q - D_{k,l}^R|$ einen benutzerdefinierten Wert δ nicht übersteigt, was gleichbedeutend damit ist, dass der Abstand zwischen Minutien m_i und m_j maximal um δ vom Abstand zwischen m_k und m_l abweicht. Das Ähnlichkeitsmaß ist definiert als die Anzahl der passenden Zeilen aus D^Q in D^R.

Formal lautet die Definition für die Ähnlichkeit zweier Templates Q und R:

$$similarity(Q, R) = |\{D_i^Q \mid \exists i \in \{1..|D^Q|\} \exists l \in \{1..|D^R|\} (match(D_i^Q, D_k^R)) \geq \beta |D^Q|\}|$$

$$\text{wobei } match(D_i^Q, D_k^R) = |\{D_{i,j}^Q \mid \exists j \in \{1..|D^R|\} \exists l \in \{1..|D^R|\} | D_{i,j}^Q - D_{k,l}^R| \leq \delta\}|$$

Die Idee des Verfahrens besteht also darin, für jede Minutie m_i des Anfragetemplates zu überprüfen, ob ein entsprechendes 2-dimensionales Objekt m_k innerhalb des durch Chaff-Points verrauschten Datenbanktemplates existiert, welches Nachbarn in gleichen Abständen wie m_i besitzt. In diesem Fall kann von einer Übereinstimmung von m_i mit m_k ausgegangen werden. Die Anzahl der übereinstimmenden Objekte beider Templates entspricht der Gesamtähnlichkeit zwischen Q und R.

Diese Idee ist noch einmal in Abbildung 44 dargestellt. Für die ausgewählte Minutie m_i des Anfragetemplates wird ein entsprechendes Objekt m_k im Datenbanktemplate gesucht. Dieses wird als gefunden markiert, falls für $\beta = 100\,\%$ aller Nachbarn von m_i ebenfalls Nachbarn mit gleichen Distanzen für m_k gefunden werden.

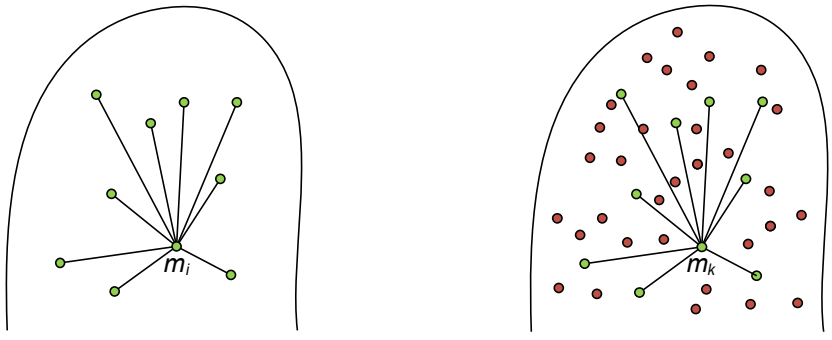

Abbildung 44: Die Minutie m_i des Anfragetemplates wird einer Minutie m_k des Datenbanktemplates zugewiesen.

6.2.2.2 Effiziente Realisierung

Die Suche aller Distanzen einer Zeile D_i^Q aus Q in allen Zeilen der Matrix D^R erfordert effiziente Vergleichsmethoden, da die naive Suche mit quadratischer Komplexität $O(|D^Q|^2 \cdot |D^R|^2)$ pro Templatepaar *(P,R)* einhergeht. Aus diesem Grund werden die im Folgenden beschriebenen Optimierungen der Suche vorgenommen.

Einen ersten Effizienzgewinn erhält man, indem die übereinstimmenden Minutien aus Q und R in Beziehung zueinander gestellt werden. Andernfalls brächte die Reihenfolge der Abstände in den Distanzmatrizen prinzipiell keine Zusatzinformation und könnte somit vernachlässigt werden. Sinnvoll erweist sich hier eine aufsteigende Sortierung der Distanzen der Zeilen beider Matrizen D^Q und D^R vor der Suche, sodass anschließend die effizientere Binärsuche erfolgen kann. Für jedes Element aus einer Zeile D_i^Q beträgt die Suchkomplexität nun nur noch $O(\log|D^R|)$, was eine Gesamtkomplexität für einen Zeilenvergleich von $O(|D^Q| \cdot \log|D^R|)$ zur Folge hat.

Die Beschleunigung, die durch die binäre Suche erreicht wird, lässt sich weiter erhöhen, indem die Sortierung der Zeilen von D^Q mitberücksichtigt wird. Wurde für D_{ij}^Q ein entsprechendes D_{kl}^R gefunden, so kann das nachfolgende Element $D_{i(j+1)}^Q$ nicht auf ein Element D_{km}^R mit $m < l$ passen, da sowohl $D_{i,j+1}^Q \geq D_{i,j}^Q$ als auch $D_{k,l}^R \geq D_{k,m}^R$ gilt. Somit kann bei jeder Suche der Suchbereich für das Element $D_{i(j+1)}^Q$ eingeschränkt werden, indem die Suche bei der zuletzt gefundenen Position beginnt. Geht man nun vereinfachend davon aus, dass nach jeder Suche der Suchbereich um $|D^R|/|D^Q|$ verkleinert wird, verringert sich, wie aus Formel 1 folgt, die Anzahl der Suchschritte für eine Zeile D_i^Q , wodurch insgesamt ein weiterer Geschwindigkeitsgewinn erzielt wird.

$$\sum_{j=0}^{q-1} \log\left(r - \frac{j \cdot r}{q}\right) = \sum_{j=0}^{q-1} \log\left(rq - jr\right) \overset{!}{\leq} \sum_{j=0}^{q-1} \log\left(\frac{r \cdot q}{q}\right) = q \cdot \log(r)$$

Formel 1: Abschätzung der Anzahl der Suchschritte einer Zeile. Diese ist kleiner als die Suchkomplexität bei Standard-Binärsuche. Zur Vereinfachung steht q für $|D^Q|$ und r für $|D^R|$.

6.2.2.3 Unscharfe binäre Suche

Im vorherigen Abschnitt wurde der Suchalgorithmus für die Suche einer Zeile D_i^Q in D_k^R vorgestellt. Dieser ging davon aus, dass ein Wert D_{ij}^Q in einer Zeile D_k^R existiert. Da im Falle des Minutienvergleichs jedoch kleinere lokale Verschiebungen auftreten können, sodass die Abstände nicht exakt miteinander übereinstimmen, muss die Suche angepasst werden. Sei δ eine vom Benutzer definierte Abweichungstoleranz zwischen zwei Distanzen D_{ij}^Q und D_{kl}^R , die auftreten darf. So ändert sich die Suche im Vergleich zum Standard-Binärsuchalgorithmus wie in Java-Codelisting 2 zu sehen ist.

Die angepasste binäre Suche behält die Komplexität des Standard-Binärsuchalgorithmus von *O(log N)* und kann mit der in Abschnitt 6.2.2.2 beschriebenen Technik beschleunigt werden.

Für das Verfahren GeoMatch werden aus allen Datenbanktemplates die entsprechenden Dreiecke extrahiert und als Datenbankpatterns gespeichert. Dieser Prozess wird einmal für den gesamten Datenbestand durchgeführt und in einer Datenbank gespeichert, sodass alle Datenbankpatterns zur

Anfragezeit zur Verfügung stehen. Für das Anfragetemplate wird dieser Prozess erst zur Anfragezeit durchgeführt. Das zum Vergleich eingesetzte Pattern hat die in Tabelle 14 dargestellte Struktur. Für ein Pattern wird die PatternID sowie eine Liste von Dreiecken gespeichert.

```
1  public static int suche(double v, double[] array, double epsilon) {
2      int ergebnis = -1;
3      int erstes = 0;
4      int letztes = array.length - 1;
5
6      while (erstes <= letztes && ergebnis < 0) {
7          final int mitte = (letztes + erstes) / 2;
8          if ( Math.abs(array[mitte] - v) < delta ) {
9              ergebnis = mitte;   // Wert gefunden
10         } else if (array[mitte] < v) {
11             erstes = mitte + 1; // rechts weitersuchen
12         } else if (array[mitte] > v) {
13             letztes = mitte - 1; // links weitersuchen
14         }
15     }
16     return ergebnis;
17 }
```

Codelisting 2: Algorithmus für die unscharfe binäre Suche. Die gelb markierte Zeile (Zeile 8) deutet die Änderung des Standard-Binärsuchalgorithmus an.

In Codelisting 3 findet sich der Pseudocode für das Erstellen eines Rankings der gegebenen Datenbankpatterns D^R ($R \in DB$) für gegebenes Anfragetemplate Q, Fehlertoleranzparameter δ, sowie den Schwellwert β. Für jedes Datenbankpattern wird die Anzahl der übereinstimmenden Zeilen (Minutien) unter Berücksichtigung von des Schwellwertes β gemessen. Um bei gleicher Anzahl übereinstimmender Zeilen stärker zwischen zwei Datenbankpatterns differenzieren zu können, bietet es sich z.B. an neben der Zeilenzahl ebenfalls die Gesamtzahl der Übereinstimmenden Zellen für den Rankingscore zu verwenden. Dieser könnte beispielsweise gebildet werden, indem an die Anzahl der übereinstimmenden Zeilen $round\,(\log_{10}(|D^Q|^2))$ viele Ziffern für die Gesamtzahl der übereinstimmenden Zellen gehängt werden.

6.2.2.4 Platzkomplexität

Für alle bisher beschriebene Verfahren des Matrix-Comparators beträgt die Platzkomplexität für einen Vergleich $O(|D^Q|^2+|D^R|^2)$ pro Fingerabdruck. Für den Speicherplatzbedarf zur Erstellung eines Rankings basierend auf dem Matrix-Comparator-Ansatz besteht also eine quadratische Abhängigkeit von der Länge des Anfragetemplates, sowie des Datenbanktemplates.

```
Eingabe: Datenbankpatterns D^R für alle R∈DB , Anfragetemplate Q,
Toleranzparameter δ , Schwellwert β

 Build Anfragepattern D^Q
 Build Rankingvector rank[1,...,|DB|]
 new ranking = [(r1,0),(r2,0),...,(r|DB|,0)]
 FOR EACH D^R in DB
   int matchedcells = 0
   int matchedminutiae = 0
   FOR EACH  D_i^Q  in  D^Q
     int max = 0
     FOR EACH  D_k^R  in  D^R
       int found := 0
       FOR EACH  D_ij^Q  in  D_i^Q
         IF( suche(D_ij^Q, D_k^R, δ)!=-1 )THEN
             found++
       max = maximum(max,found);
     IF ( max ≥ β*|D^R| ) THEN matchedminutiae++;
     matchedcells += max;
   rank.set(R.id, concatenate(matchedminutiae,matchedcells))
```

Codelisting 3: Algorithmischer Ablauf des Rankingverfahrens basierend auf Matrix-Comparator

6.2.2.5 Physische Datenbankstruktur

Für das Verfahren Matrix-Comparator werden für alle Datenbanktemplates sowie das Anfragetemplate die in Tabelle 14 dargestellte Datenbankstruktur extrahiert. Für jede Minutie, bzw. jeden Chaffpoint, werden die Abstände zu allen anderen Minutien sowie gegebenenfalls Chaffpoints bestimmt. Diese Abstände werden anschließend aufsteigend nach Größe sortiert.

PatternID				
minutia	$dist_1$	$dist_2$	...	$dist_n$
m_1	d_{11}	d_{12}	...	d_{1n}
m_2	d_{21}	d_{22}	...	d_{2n}
⋮	⋮	⋮	⋮	⋮
m_n	d_{n1}	d_{n2}	...	d_{nn}

Tabelle 15: Datenstruktur eines Anfrage- sowie Datenbank-Patterns

6.2.3 BioSimJoin

Ausgangssituation für das Verfahren BioSimJoin ist ein Datenraum, in dem die Minutien aller Personen verschleiert, also mit zusätzlichen Chaff-Points gespeichert sind. Dabei werden die Minutien und Chaff-Points verschiedener Finger, also beispielsweise Daumen und Zeigefinger, in separaten Datenräumen gespeichert. Die Chaff-Points repräsentieren zufällige Punkte aus dem Raum der potenziellen Minutien, die zusammen mit den echten Minutien abgespeichert werden und diese damit verschleiern. Diese Verschleierung gewährleistet eine sichere Speicherung der sicherheitskritischen biometrischen Merkmale. In einem ersten Schritt wird nun zu jeder Minutie der angefragten Person eine Bereichsanfrage mit Radius r durchgeführt. In diesem Fall soll also z.B. eine Anfrage der Form "Finde alle Minutien bzw. Chaff-Points, deren Position im Bereich [225...250] liegt." beantwortet werden. Dieses Vorgehen ist schematisch an Abbildung 45 dargestellt.

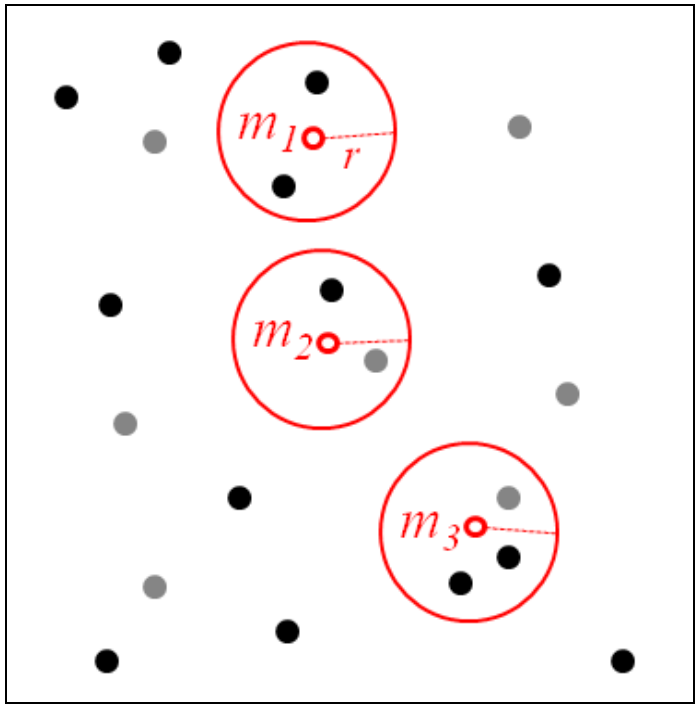

Abbildung 45: Ausgangssituation für das Verfahren BioSimJoin

In diesem Beispiel enthält der angefragte Fingerabdruck drei unterschiedliche Minutien m_1, m_2 und m_3. Für jede dieser Minutien m_i werden diejenigen Minutien, aber auch Chaff-Points bestimmt, die sich innerhalb des Bereiches mit Radius r um m_i befinden. Formal bedeutet dies, dass sich die räumlichen Koordinaten (hier X und Y) zweier Minutien m_a und m_b höchstens um einen euklidischen Abstand r unterscheiden dürfen. Dieses Distanzmaß definiert sich wie folgt:

$$dist(m_a, m_b) = \sqrt{(m_{a_x} - m_{b_x})^2 + (m_{a_y} - m_{b_y})^2}$$

Dieser Bereich ist durch die roten Kreise um m_i dargestellt. „Echte" Minutien sind durch schwarze Punkte dargestellt, graue Punkte repräsentieren Chaff-Points. Um die Sicherheit der biometrischen Merkmale nicht zu gefährden, liegt diese Information dem Verfahren BioSimJoin allerdings nicht vor. Vielmehr werden sowohl Minutien, als auch Chaff-Points in gleicher Weise behandelt. Für jede Minutie m_i kann anschließend eine Liste von Minutien bzw. Chaff-Points erstellt werden, die sich innerhalb des Bereichs um m_i befinden. Für jede solche Minutie bzw. Chaff-Point ist dabei bekannt, von welcher Person er stammt. Wir möchten an dieser Stelle darauf hinweisen, dass diese Informationen sowohl für Chaff-Points als auch für echte Minutien vorliegen, da Chaff-Points lediglich dazu dienen die biometrischen Merkmale zu verschleiern. Sie dürfen sich daher nicht in ihrer Darstellung von echten Minutien unterscheiden. Ansonsten wäre für einen potentiellen Angreifer ohne weiteres erkennbar, um welche Art von Merkmal es sich handelt. Diese Repräsentation wird mittels Abbildung 46 genauer erklärt.

Für die Minutie m_3 der angefragten Person konnten beispielsweise zwei echte Minutien und ein Chaff-Point innerhalb eines Radius r als Treffer identifiziert werden. Der graue Chaff-Point stammt von Person 1. Die linke Minutie bezieht sich auf den Fingerabdruck der Person mit der Identifikationsnummer 3 und die rechte Minutie gehört, wie auch der Chaff-Point zu Person 1. Insgesamt ergeben sich über alle drei Minutien der angefragten Person hinweg folgende Kandidaten: Vier Treffer für Person 1, zwei Treffer für Person 2 und ein Treffer für die Person 3. Diese Reihenfolge entspricht schließlich dem Ergebnis des Verfahrens BioSimJoin. Es dient dazu dem Nutzer eine effiziente Identifikation zu ermöglichen, da es sich in diesem Beispiel sehr wahrscheinlich bei Person 1 um die angefragte Person handelt. Unsere Testergebnisse zeigen, dass der Suchraum durch BioSimJoin stark eingeschränkt werden kann.

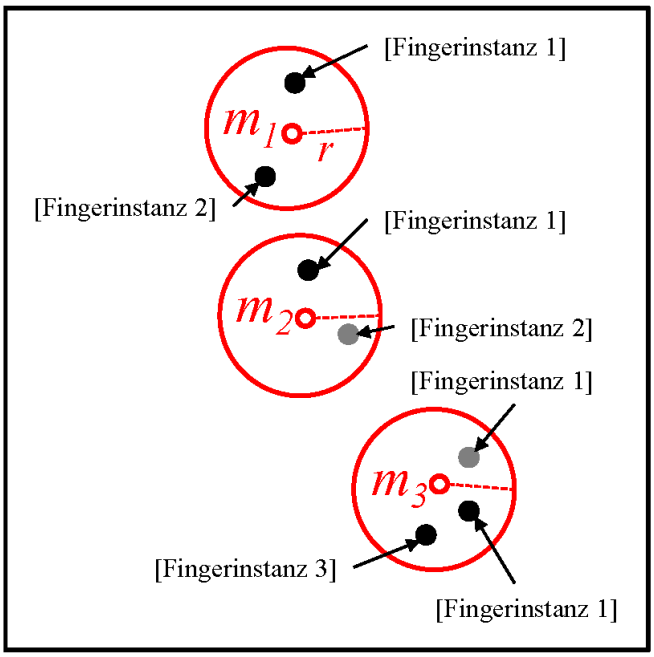

Abbildung 46: Information der Treffer für BioSimJoin

```
Algorithmus filter_BioSimJoin(minutiae_query, r)

Eingabe: Minutien der Anfrageperson
         Radius der Bereichsanfrage

candidates        = [(p_1,0), (p_2,0), …, (p_n,0)]
minutiae_DB       = [(x_1,y_1,p_1), (x_2,y_2,p_2), …, (x_m,y_m,p_m)]

FOR EACH Minutia m_q IN minutiae_query DO{
    FOR EACH Minutia m_DB IN minutiae_DB DO{
        IF m_q.fingertype = m_DB.fingertype
        AND dist(m_q,m_DB) ≤ r DO{
                candidates.increment(m_DB.p)
        }

    }
}
Ausgabe: Kandidatenliste sortiert nach Treffer-Werten
```

Abbildung 47: Algorithmischer Ablauf des Verfahrens BioSimJoin

Abbildung 47 fasst den algorithmischen Ablauf nochmals schematisch zusammen. Die Methode `filter` von BioSimJoin erhält zwei Parameter. Zum einen alle Minutien der angefragten Person und zum anderen den Radius *r*, der die Bereichsanfrage spezifiziert. Zu Beginn des Algorithmus wird eine Datenstruktur `candidates` erstellt, die für alle Personen der Referenzdaten die Anzahl der Treffer mit den Minutien der angefragten Person speichert. Diese Datenstruktur wird mit jeweils 0 Treffern je Person initialisiert. Des Weiteren liegt dem Algorithmus eine Datenstruktur $\texttt{minutiae}_{\texttt{DB}}$ vor, in der sowohl Minutien als auch Chaff-Points aller Personen der Referenzdaten gespeichert sind. Für jede solche Minutie sind die Angaben über X- und Y-Koordinate, sowie die Identifikationsnummer der entsprechenden Person, zu der die Minutie gehört, verfügbar. Für jede Minutie der angefragten Person wird anschließend der euklidische Abstand zu allen Minutien bzw. Chaff-Points der Referenzdaten des passenden Fingertyps bestimmt. Falls dieser den Wert von *r* nicht überschreitet, wird ermittelt, von welcher Person diese Minutie bzw. der Chaff-Point stammt. Die Anzahl der Treffer dieser Person kann somit um 1 erhöht werden. Nachdem die äußere Schleife vollständig durchlaufen wurde, speichert `candidates` die Anzahl der tatsächlich ermittelten Treffer für jede Person der Referenzdaten. Abschließend wird diese Liste nach Treffern absteigend sortiert und von dem Verfahren BioSimJoin als Ergebnis zurückgeliefert.

6.2.4 BioSimJoin*

BioSimJoin* erweitert das zuvor vorgestellte Verfahren BioSimJoin dahingehend, dass die Bereichsanfrage durch eine spezielle Indexstruktur unterstützt und somit beschleunigt wird.

Genauer gesagt, werden bei BioSimJoin* die X- und Y-Koordinaten aller Minutien in der Datenbank in einem R-Baum [Gutt84] organisiert. Dabei wird pro Fingertyp jeweils ein separater Baum aufgebaut. Der R-Baum ist eine Indexstruktur, die ursprünglich zur Speicherung von hochdimensionalen Daten entwickelt wurde, und eignet sich daher für die Verwaltung von biometrischen Punktdaten. Der algorithmische Ablauf entspricht dem von BioSimJoin der in Abbildung 47 beschrieben ist.

6.2.4.1 Physische Datenbankstruktur

Der R-Baum ist eine hierarchische Indexstruktur, die sich aus zwei Arten von Seiten zusammensetzt. Die *Datenseiten* enthalten die zu indizierenden räumlichen Daten, in unserem Fall die zweidimensionalen Minutien. In den inneren Seiten, den sog. *Indexseiten*, werden minimale rechteckige Datenregionen gespeichert, die alle im Teilbaum darunter liegenden Datenregionen vollständig umschließen. Abbildung 48 zeigt ein solches minimal umgebendes Rechteck zum einen bezüglich einer Menge von Minutien (links) und zum anderen bezüglich einer Menge von minimal umgebenden Rechtecken (rechts). Letzteres entspricht einer Indexseite des R-Baums auf höherer Ebene.

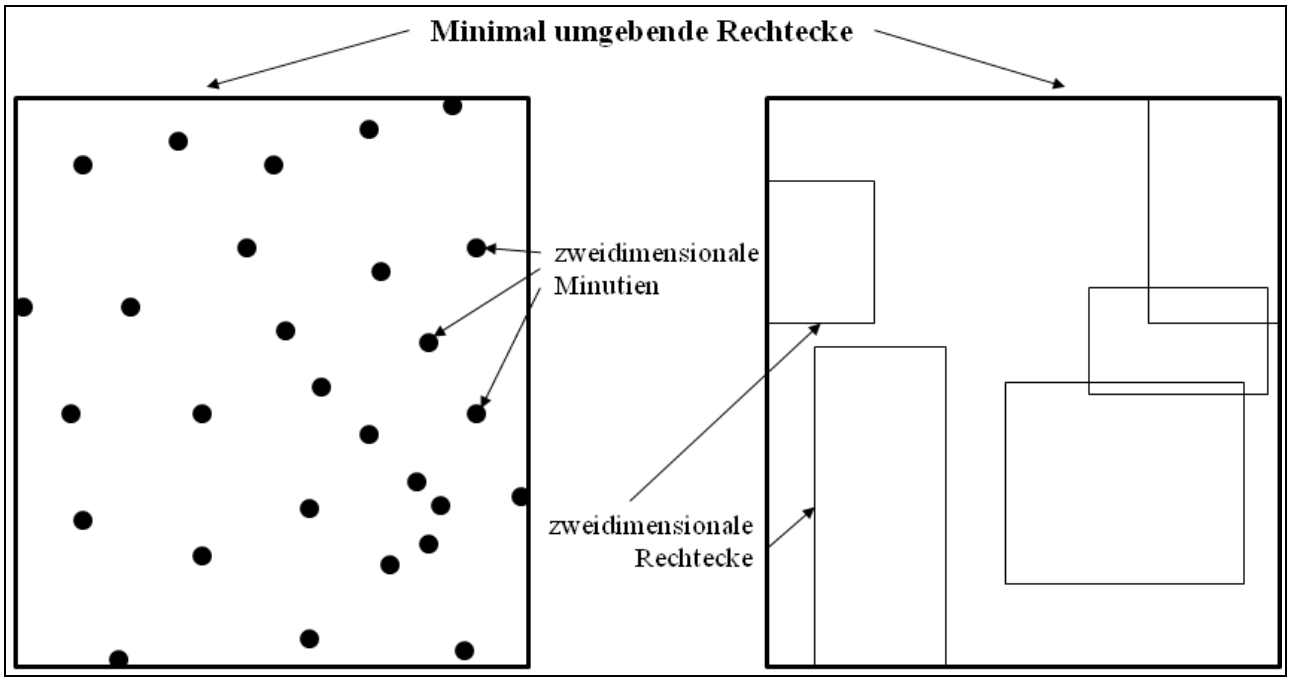

Abbildung 48: Minimal umgebendes Rechteck für eine Menge von Punkt- bzw. Rechteckdaten

Jedes Rechteck in einer Indexseite des R-Baums umfasst alle minimal umgebenden Rechtecke bzw. Punktdaten in allen Index- oder Datenseiten, die im zugehörigen Teilbaum liegen. Abbildung 49 illustriert für ein kleines Beispiel die durch einen R-Baum vorgegebene Partitionierung des

Datenraums. Die roten Rechtecke R_1 bis R_8 enthalten insgesamt 23 Minutien. Jedes dieser Rechtecke approximiert minimal die ihm zugeordneten Punktdaten. Auf der nächsten Ebene werden diese Rechtecke durch die grünen grob-gestrichelten Rechtecke zusammengefasst. Jeweils zwei dieser Rechtecke werden wiederum durch die blauen fein-gestrichelten Rechtecke vereint. Die Wurzel des R-Baums repräsentiert den gesamten Datenraum, indem sie mit dem minimal umgebenden Rechteck $R_{1,2,3,4,5,6,7,8}$ die Rechtecke bzw. Punktdaten zusammenfasst. Über einen Parameter *c* ist geregelt, wie viele Einträge maximal in einer Seite des R-Baums abgelegt werden dürfen. Da die gewählte Seitenkapazität stark vom verwendeten Datensatz abhängt, wird in Abschnitt Evaluierung der Identifikationslösungen eine Evaluierung für die optimale Seitenkapazität bzgl. der von uns verwendeten Datenbanken präsentiert.

Der R-Baum wird beim Anlegen der Datenbasis erstellt und kann durch effiziente Einfüge- und Löschoperationen dynamisch aktualisiert werden. Ein Aufbau bei jeder Anfrage an die Datenbank ist daher nicht erforderlich.

6.2.4.2 Beschleunigung der Bereichsanfrage

Für jedes in einer Seite des R-Baums gespeicherte Rechteck steht dimensionsweise ein Intervall zu Verfügung, das die Ausdehnung des Rechtecks in X- und Y-Richtung angibt. Mithilfe dieser Intervalle kann die Bearbeitung der zentralen Fragestellung von BioSimJoin und BioSimJoin* „Welche Minutien bzw. Chaff-Points befinden sich innerhalb eines Radius *r* um die Minutie der angefragten Person“ signifikant beschleunigt werden.

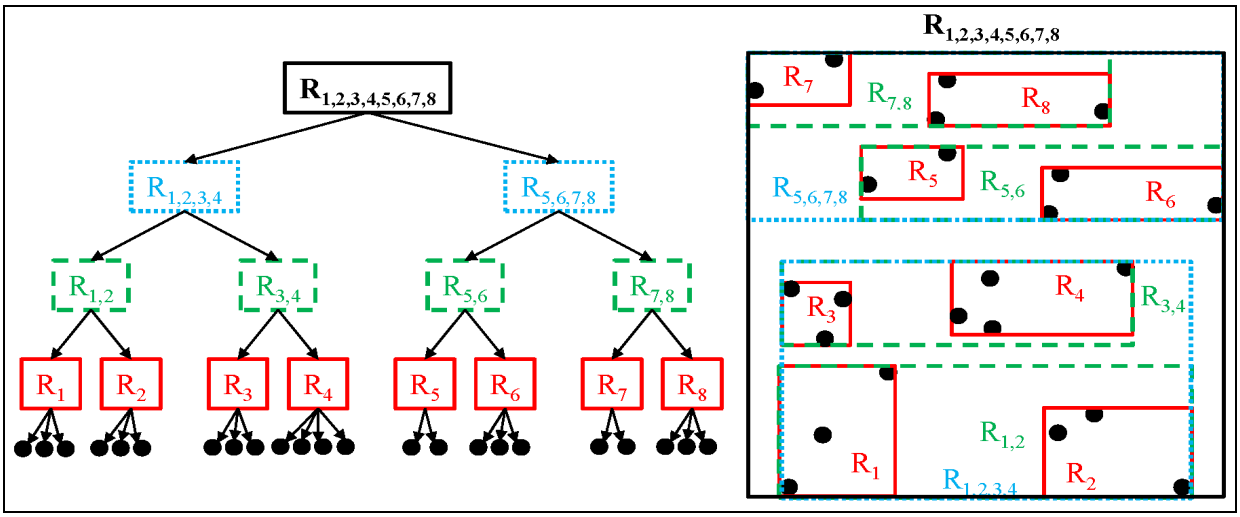

Abbildung 49: Partitionierung des Datenraums durch einen R-Baum

Die Beantwortung einer solchen Bereichsanfrage ist ein rekursiver Prozess. Das bedeutet, dass der R-Baum von der Wurzel hin zu den Blättern abwärts durchsucht wird. In jedem Schritt wird überprüft, welche minimal umgebenden Rechtecke der aktuellen Seite innerhalb des Intervalls bezüglich dem Radius *r* und den Koordinaten der angefragten Minutie liegen. Falls es sich bei den minimal umgebenden Rechtecken um Indexseiten handelt, wird dieser Test so lange durchgeführt bis man auf tiefster Ebene bei den tatsächlichen Punktdaten in den Datenseiten endet. Im besten Fall muss nur ein Pfad von der Wurzel bis zu einer Datenseite durchlaufen werden. Da die minimal umgebenden Rechtecke eine konservative Approximation der Daten darstellen, gehen bei der Suche keine Treffer verloren.

6.2.5 BioNN

Bei der Anfragebearbeitung sucht BioNN einen zu einem Anfragefinger ähnlichen Datenbankfinger anhand geometrischer Übereinstimmungen. BioNN beschränkt sich dabei auf Dreiecke. Ein solches Dreieck besteht aus einer Minutie, sowie einer weiteren Minutie, die dieser am nächsten liegt, dem sogenannten ersten nächsten Nachbarn der Minutie (1NN), und einer weiteren Minutie, die dieser am zweit nächsten liegt, dem sogenannten zweit nächsten Nachbarn der Minutie (2NN). Die Entfernungen zwischen den Minutien werden dabei anhand euklidischer Distanz bestimmt. Dieser Ansatz berücksichtigt explizit Probleme, die durch Rotation und/oder Verschiebung zwischen Anfragefinger und dessen Referenzfinger in der Datenbank zustande kommen können, da nicht die Position des Dreiecks auf dem Finger, sondern lediglich die Ähnlichkeit der Dreiecke maßgeblich ist. So können auch Dreiecke gefunden werden, deren Minutien auf unterschiedlichen Positionen auf dem jeweiligen Finger liegen. Allerdings gilt dies nicht für die Entfernungen zwischen den einzelnen Minutien innerhalb eines Dreiecks, da sich die Entfernungen zwischen den einzelnen Minutien im Anfragefinger und Datenbankfinger aufgrund von Stauchen oder Dehnen der Finger unterscheiden können. Aus diesem Grund wird nicht nach einem gleichen Dreieck im Anfrage- und Datenbankfinger, sondern nach zwei einander ähnlichen Dreiecken gesucht. Zwei Dreiecke werden dann als ähnlich eingestuft, wenn sich ihre Seiten maximal um einen Parameter δ voneinander unterscheiden. Dies ist in Abbildung 50 illustriert.

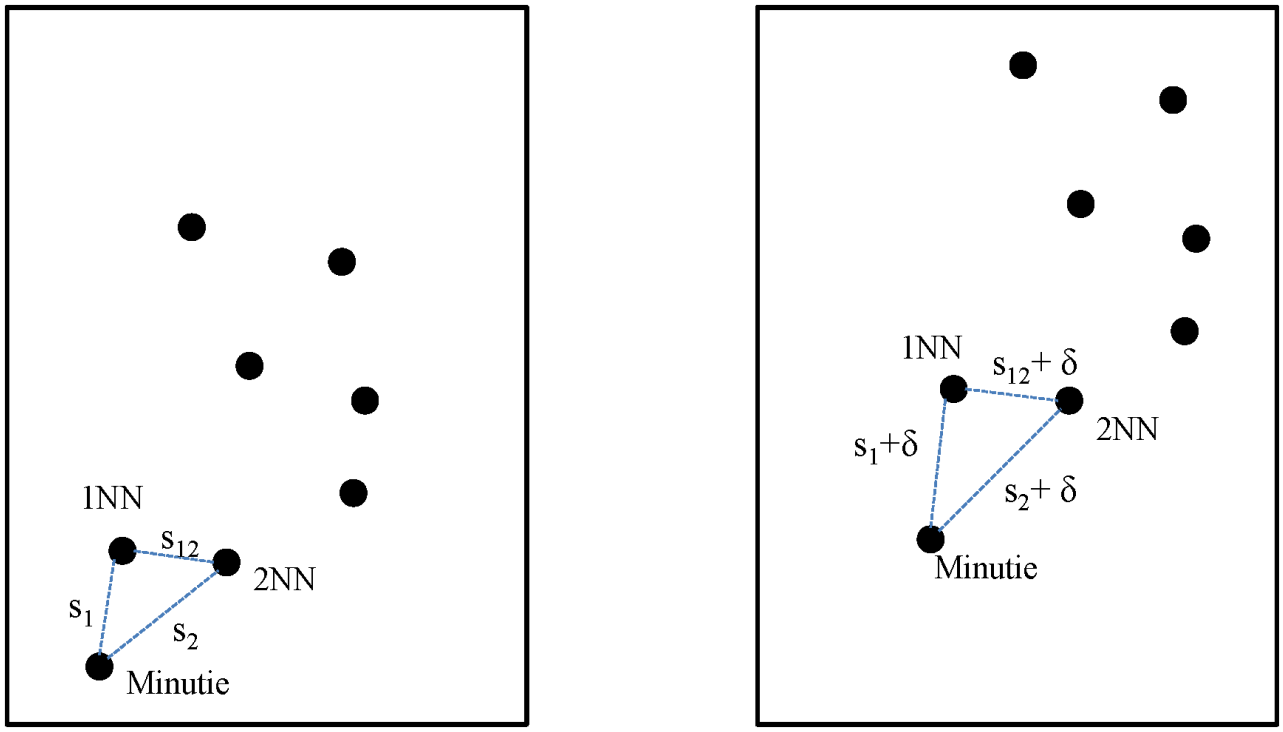

Abbildung 50: Dreiecke in Anfrage- und Referenzfinger, deren Seiten sich um einen Wert δ unterscheiden

Aufgrund des Fuzzy-Vaults enthalten die Fingerabdrücke in der Datenbank neben den echten Minutien auch Chaff-Points. Aus diesem Grund können die Dreiecke der Datenbankfinger im Gegensatz zu den Anfragefingern sowohl echte Minutien als auch Chaff-Points beinhalten. Daher muss sich die Suche nach einem ähnlichen Dreieck in den Datenbankobjekten nicht nur auf die zwei nächsten Nachbarn einer Minutie, sondern auf die k nächsten Nachbarn erstrecken (vgl. Abbildung 51).

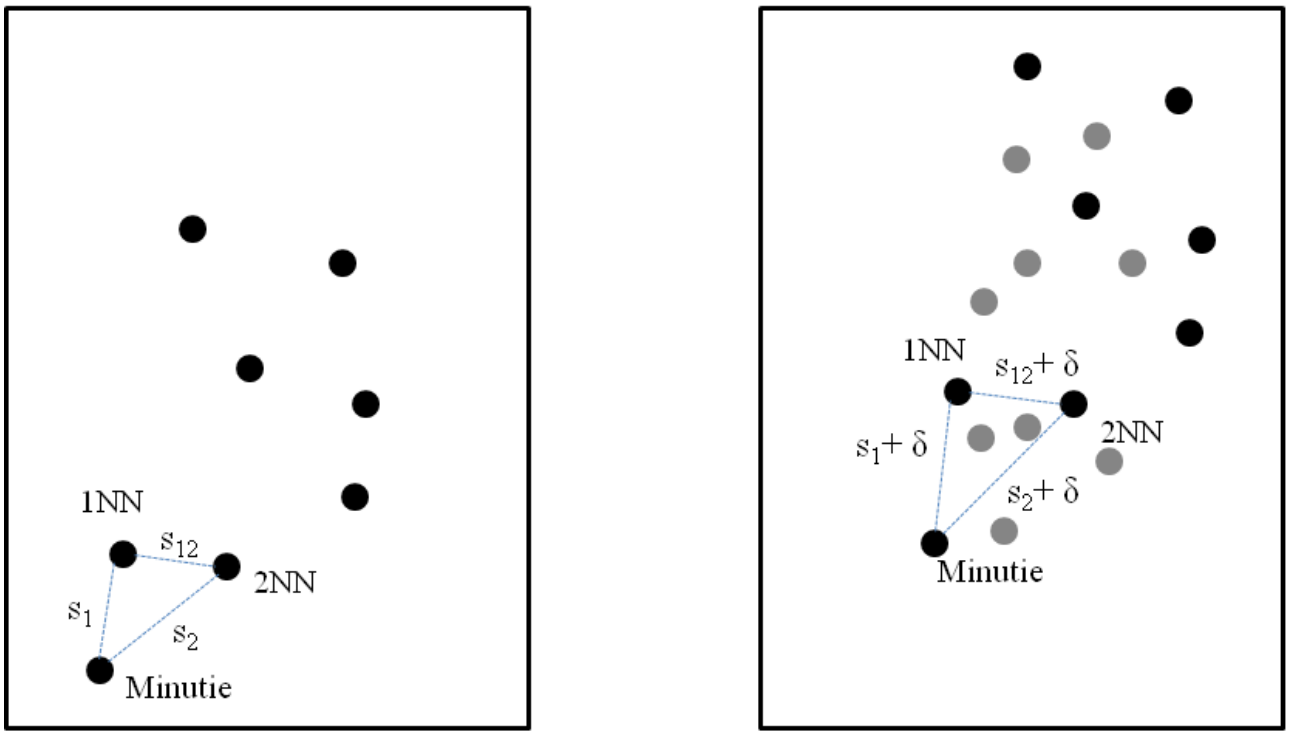

Abbildung 51: Ähnliche Dreiecke in Anfrage- und Referenzfinger, wobei der Referenzfinger zusätzlich Chaff-Points enthält

6.2.5.1 Physische Datenbankstruktur

Bei BioNN sind die Daten aus der Datenbank in der sogenannten *DatabaseStructure* gespeichert. Diese Struktur wird initial bei Anlegen der Datenbank aufgebaut, und muss nicht bei jeder Anfrage erneut generiert werden. Für den Anfragefinger hingegen, muss bei der Anfrage die sogenannte *QueryStructure* generiert werden.
In der bei der Anfrage generierten *QueryStructure* sind für jede Minutie des Anfragefingers die Koordinaten der jeweils ersten und zweiten nächsten Nachbarn sowie jeweils die Distanz zwischen den beiden Minutien gespeichert (vgl. Abbildung 52). Auf diese Weise kann anhand der Einträge in der *QueryStructure* für jede Minutie die Position der zwei nächsten Nachbarn sowie jeweils die Distanz zu diesen Nachbarn abgefragt werden. Anhand der Positionen der 1NN und 2NN kann deren Distanz zueinander bestimmt und somit das Dreieck ermittelt werden das diese drei Minutien umschreibt.

In der Datenbank ist für jeden Finger eine ähnliche Struktur, die sogenannte *DatabaseStructure* gespeichert. Da die Finger in der Datenbank zusätzlich zu den echten Minutien noch Chaff-Points enthalten, werden in dieser Struktur für jede Minutie und jeden Chaff-Point jeweils die Distanz zu den *k* nächsten Nachbarn sowie deren Koordinaten gespeichert (vgl. Abbildung 53). Da dem Algorithmus keine Information darüber vorliegt, ob es sich bei einer Minutie um eine echte Minutie oder einen Chaff-Point handelt, können unter den *k* NN einer Minutie daher sowohl echte Minutien als auch Chaff-Points vorkommen.

Durch das Ablegen der Koordinaten der zwei bzw. *k* nächsten Nachbarn einer Minutie in der *QueryStructure* respektive in der *DatabaseStructure,* ist es möglich zur Laufzeit die Distanz zwischen zwei beliebigen nächsten Nachbarn einer Minutie zu berechnen.

QueryStructure		
	1NN	2NN
m_1	dist(m_1, 1NN), $XPos_{1NN}$, $YPos_{1NN}$	dist(m_1, 2NN), $XPos_{2NN}$, $YPos_{2NN}$
m_2	dist(m_2, 1NN), $XPos_{1NN}$, $YPos_{1NN}$	dist(m_2, 2NN), $XPos_{2NN}$, $YPos_{2NN}$
m_3	dist(m_3, 1NN), $XPos_{1NN}$, $YPos_{1NN}$	dist(m_3, 2NN), $XPos_{2NN}$, $YPos_{2NN}$
…	…	…
m_i	dist(m_i, 1NN), $XPos_{1NN}$, $YPos_{1NN}$	dist(m_i, 2NN), $XPos_{2NN}$, $YPos_{2NN}$

Abbildung 52: Datenstruktur für die Speicherung des Anfragefingers bei BioNN

DatabaseStructure					
		1NN	2NN		kNN
$Finger_1$					
	m_1	dist(m_1, 1NN), $XPos_{1NN}$, $YPos_{1NN}$	dist(m_1, 2NN), $XPos_{2NN}$, $YPos_{2NN}$	…	dist(m_1, kNN), $XPos_{kNN}$, $YPos_{kNN}$
	m_2	dist(m_2, 1NN), $XPos_{1NN}$, $YPos_{1NN}$	dist(m_2, 2NN), $XPos_{2NN}$, $YPos_{2NN}$	…	dist(m_2, kNN), $XPos_{kNN}$, $YPos_{kNN}$
	…	…	…	...	
	m_j	dist(m_j, 1NN), $XPos_{1NN}$, $YPos_{1NN}$	dist(m_j, 2NN), $XPos_{2NN}$, $YPos_{2NN}$	…	dist(m_j, kNN), $XPos_{kNN}$, $YPos_{kNN}$
…					
$Finger_k$					
	m_1	dist(m_1, 1NN), $XPos_{1NN}$, $YPos_{1NN}$	dist(m_1, 2NN), $XPos_{2NN}$, $YPos_{2NN}$	…	dist(m_1, kNN), $XPos_{kNN}$, $YPos_{kNN}$
	m_2	dist(m_2, 1NN), $XPos_{1NN}$, $YPos_{1NN}$	dist(m_2, 2NN), $XPos_{2NN}$, $YPos_{2NN}$	…	dist(m_2, kNN), $XPos_{kNN}$, $YPos_{kNN}$
	…	…	…	…	
	m_j	dist(m_j, 1NN), $XPos_{1NN}$, $YPos_{1NN}$	dist(m_j, 2NN), $XPos_{2NN}$, $YPos_{2NN}$	…	dist(m_j, kNN), $XPos_{kNN}$, $YPos_{kNN}$

Abbildung 53: Datenstruktur für die Speicherung aller Datenbankeinträge bei BioNN

Die Funktionsweise von BioNN lässt sich an einem Beispiel für den Abgleich eines Anfragefingers durch seinen Referenzfinger aus der Datenbank nachvollziehen. Einfachheitshalber wird in dem Beispiel angenommen, dass der Anfragefinger nur fünf Minutien m_1, m_2, m_3, m_4 und m_5 enthält und für den Referenzfinger in der Datenbank außer diesen fünf Minutien neun Chaff-Points c_1-c_9 abgelegt sind (vgl. Abbildung 54). Der Parameter *k* für die Anzahl nächster Nachbarn wird entsprechend dem Mischverhältnis echte Minutien/ Chaff-Points auf den Wert 6 gesetzt. Damit werden für jede Minutie zusätzlich deren sechs nächste Nachbarn in der *DatabaseStructure* abgelegt.

Nachdem bei der Anfrage die *QueryStructure* für den Anfragefinger generiert wurde, startet BioNN mit einer beliebigen Minutie des Anfragefingers(vgl. Abbildung 55), in diesem Beispiel mit Minutie m_1(Die aktuell betrachtete Minutie ist rot markiert). Anhand der Einträge zu dieser Minutie in der

QueryStructure werden die Minutien m_2 und m_3 jeweils als die erste bzw. zweite benachbarte Minutie von m_1 erkannt (Die nächsten Nachbarn der aktuell betrachteten Minutie rot umrandet). In der *QueryStructure* sind die Entfernungen zwischen m_1 und m_2 und zwischen m_1 und m_3 gespeichert. Anhand der gespeicherten Positionen von m_2 und m_3 kann zusätzlich deren Distanz berechnet und somit das Dreieck bestehend aus den Minutien m_1, m_2 und m_3 ermittelt werden. Nun wird in den Einträgen der *DatabaseStructure* unter allen Minutien des Referenzfingers nach einer Minutie gesucht, die mit zwei beliebigen Minutien ihrer sechs nächsten Nachbarn, unter Berücksichtigung des Toleranzwertes , ein ähnliches Dreieck umschreibt. Hierzu werden die Minutien eines Referenzfingers sequentiell in der *DatabaseStructure* durchlaufen. Unter den sechs nächsten Nachbarn der Minutie m_1 (c_1, c_2, c_3, c_4, m_2, m_3) des Referenzfingers findet BioNN ein ähnliches Dreieck ($m_1/m_2/m_3$) wie das Dreieck $m_1/m_2/m_3$ im Anfragefinger, und erhöht den Zähler für die Anzahl erkannter Dreiecke in dem Referenzfinger um eins.

Diese Prozedur wird für jede Minutie des Anfragefingers durchlaufen. Es können in diesem Beispiel demnach maximal fünf Dreiecke aus dem Anfragefinger im Referenzfinger erkannt werden. Die Anzahl der gefundenen Dreiecke dient als Maß für die Übereinstimmung zwischen Anfragefinger und Referenzfinger.

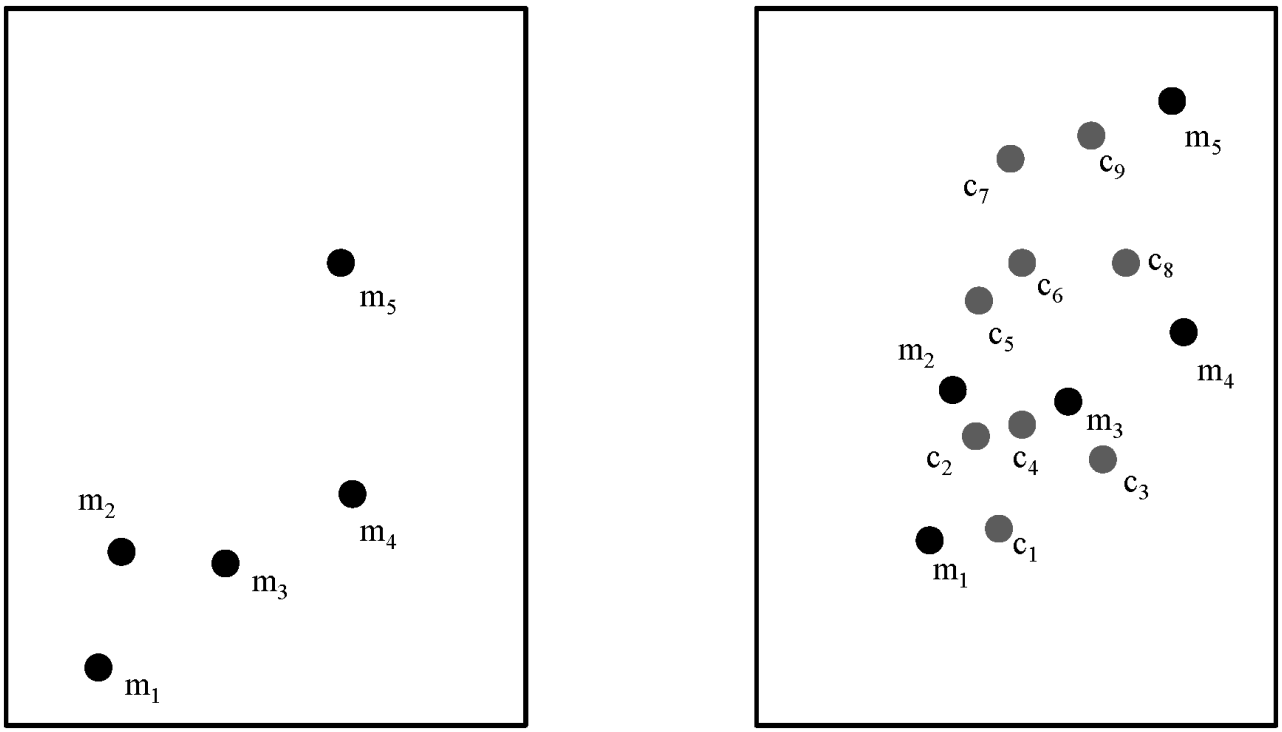

Abbildung 54: BioNN: Beispiel Anfragefinger (links) und Referenzfinger (rechts)

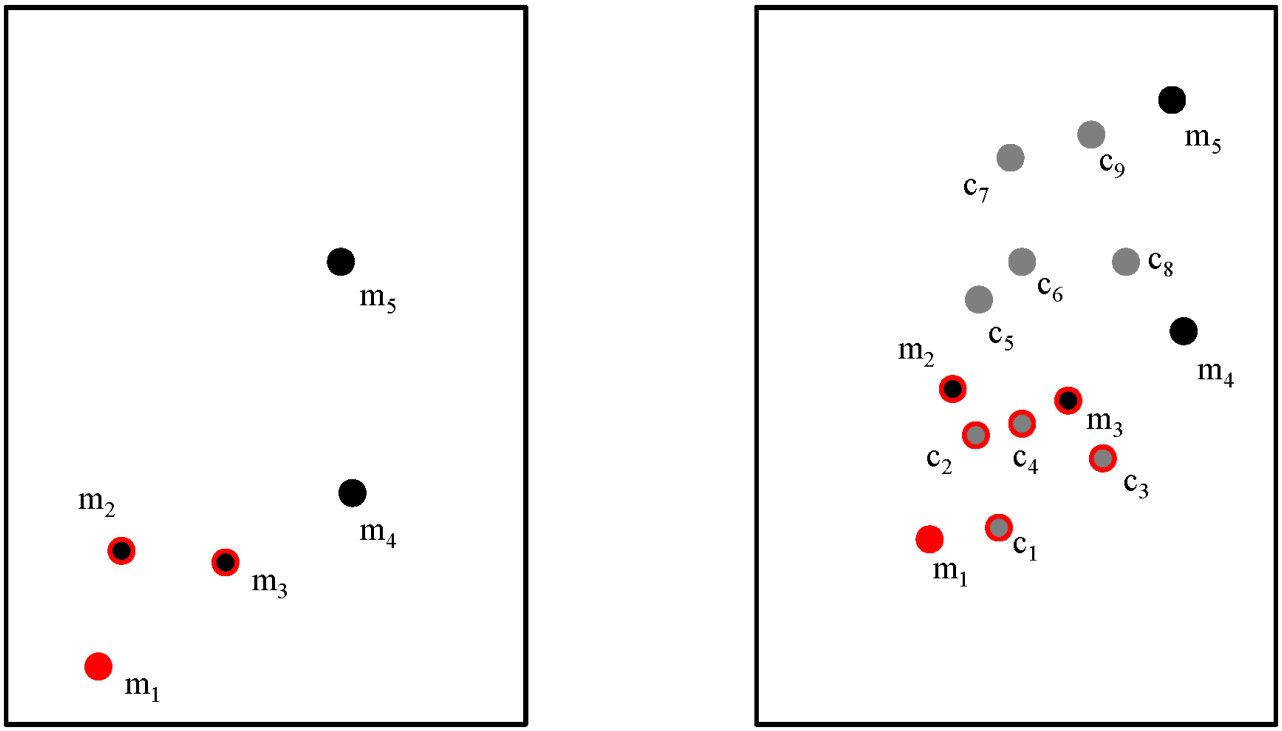

Abbildung 55: BioNN: Suche nach einem ähnlichen Dreieck M1/M2/M3 des Anfragefingers (links) im Referenzfinger (rechts)

In Abbildung 56 ist der Ablauf des Verfahrens BioNN schematisch beschrieben. Die Methode `filter` erhält drei Parameter. Zum einen alle Minutien der angefragten Person und zum anderen einen Wert δ, der den Toleranzwert für die Seiten des Dreiecks spezifiziert, und einen Wert k, der die maximale Anzahl nächster Nachbarn, über die die Suche in Datenbankobjekten erstreckt, spezifiziert. Zu Beginn des Algorithmus wird eine Datenstruktur `candidates` erstellt, die für jede Personen aus der Datenbank die Anzahl der übereinstimmenden Dreiecke zur angefragten Person speichert. Diese Datenstruktur wird mit jeweils 0 Treffern je Person initialisiert. Aus den Minutien der Anfrageperson wird die Datenstruktur `queryStructure` generiert, die für jede Minutie die Referenz auf die ersten und zweiten Nachbarn sowie deren Distanz enthält. Die Datenstruktur `databaseStructure` enthält alle Minutien der Referenzdatenbank sowie die Distanzen der k nächsten Nachbarn.

Im nächsten Schritt wird für jede Minutie der angefragten Person, das ein Dreieck mit deren 1NN und 2NN bildet für jede Minutie aus der `databaseStructure` überprüft ob diese Minutie ein ähnliches Dreieck mit zwei deren k nächsten Nachbarn bildet. Dabei wird für jede Seite des Anfragedreiecks überprüft, ob diese Seite unter Berücksichtigung des Toleranzwertes δ, auch unter den k nächsten Nachbarn der Minutie aus der Datenbank gefunden wird. Wird ein ähnliches Dreieck gefunden, wird zunächst ermittelt von welcher Person dieses Dreieck stammt und anschließend die Anzahl der Treffer für die entsprechende Person erhöht. Nachdem alle Minutien der angefragten Person überprüft wurden, enthält `candidates` die Anzahl übereinstimmender Dreiecke für jede Person der Referenzdaten. Abschließend wird diese Liste nach Anzahl der Treffer sortiert und als Ergebnis zurückgeliefert.

```
Algorithmus filter_BioNN(minutiae_query, δ, k)

Eingabe:  Minutien der Anfrageperson
          Toleranzwert für Distanz
          Anzahl nächster Nachbarn

candidates        = [(p1,0), (p2,0), …, (pn,0)]
queryStructure(minutiae_query)
                  = [{m1,dist(m1,1NN),dist(m1,2NN)},…,{mq,dist(mq,1NN), dist(mq,2NN)}]
databaseStructure = [{m1,dist(m1,1NN),…,dist(m1,kNN)},…,{mDB,dist(mDB,1NN),…,dist(mDB,kNN)}]

FOR EACH Minutia mq IN queryStructure DO{
    FOR EACH Minutia mDB IN databaseStructure DO{
        FOR EACH a≤k DO{
            FOR EACH b≤k DO{
             IF mq.fingertype = mDB.fingertype
             AND dist(mDB.aNN , mDB.bNN)-δ ≤ dist(mq.1NN, mq.2NN) ≤ dist(mDB.aNN , mDB.bNN)+δ
             AND dist(mDB, mDB.aNN)-δ ≤ dist(mq, mq.1NN) ≤ dist(mDB, mDB.aNN)+δ
             AND dist(mDB, mDB.bNN)-δ ≤ dist(mq, mq.2NN) ≤ dist(mDB, mDB.bNN)+δ DO{
                 candidates.increment(mDB.p)
                 BREAK;
              }
            }
        }
    }
}

Ausgabe: Kandidatenliste sortiert nach Treffer-Werten
```

Abbildung 56: Algorithmischer Ablauf des Verfahrens BioNN

6.3 Verfahrensevaluierung

Um die vorgeschlagenen Verfahren zur Template Protection in Identifikationssystemen geeignet evaluieren zu können, wurden unterschiedliche Datenquellen herangezogen. Anhand realer Daten aus der Datenbank NIST SD14 wurden die optimalen Parameter für jedes Verfahren bestimmt. Um die Robustheit der Identifikationslösungen zu evaluieren, wurden diese mit Hilfe gezielt manipulierter Daten getestet.

6.3.1 Vorbereitung der Datenbasis

Da die Identifikationslösungen nur die räumliche Lage, also die X- und Y-Koordinate der Minutien heranziehen, mussten diese Informationen aus den Bildern der Datenbank NIST SD14 extrahiert und geeignet vorverarbeitet werden. Zudem wurden synthetische Fingerabdrücke generiert, bei denen die Position der Minutien auf unterschiedliche Weise manipuliert wurde.

6.3.1.1 NIST SD14

Die NIST SD14 enthält Fingerabdrucksbilder für 2.700 Personen. Für jede Person ist eine Bild je Finger gespeichert. Damit besteht die Datenbank aus 27.000 unterschiedlichen Bildern. Für jeden Fingerabdruck stehen zwei Aufnahmen zur Verfügung. Eine Aufnahme dient als biometrische Referenz in der Datenbasis, die zweite Aufnahme kann für die Anfrage (biometrische Probe)

genutzt werden. Bei genauerer Betrachtung der gespeicherten Bilder, stellte sich jedoch heraus, dass innerhalb der Datenbank einige Aufnahmen enthalten sind, die die Qualitätsanforderungen für eine signifikante Evaluierung nicht erfüllen. Abbildung 57 zeigt exemplarisch drei problematische Fälle. Der linke Fingerabdruck ist durch handschriftliche Bemerkungen verunreinigt. Auf der mittleren Abbildung befinden sich Teile eines weiteren Abdrucks, der die Ergebnisse des korrekten Abdrucks verfälschen würde. Auf der rechten Seite ist zu sehen, dass der Finger im unteren Bereich nicht vollständig aufgenommen wurde.

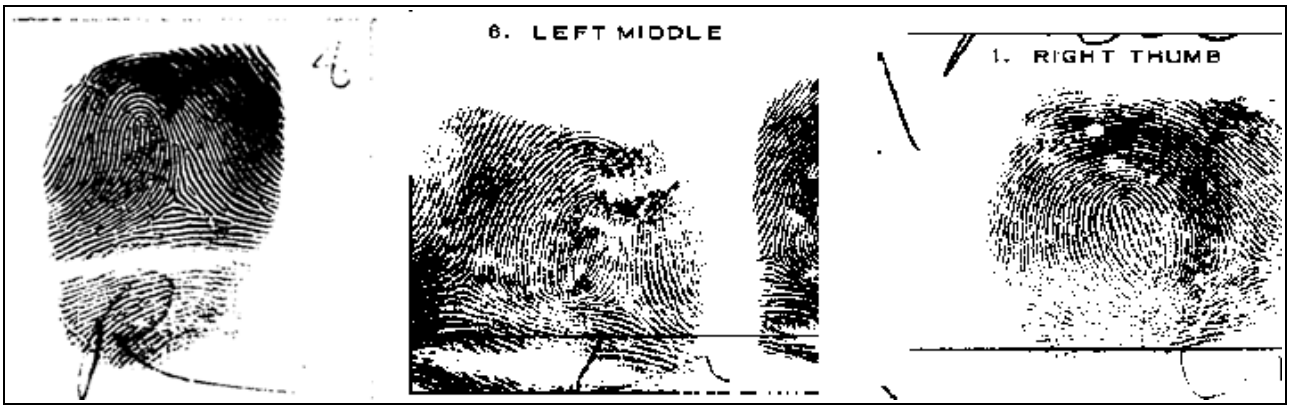

Abbildung 57: Beispiele für Fingerbilder niedriger Qualität

Aus diesem Grund haben wir den NIST Fingerprint-Comparison-Algorithmus BOZORTH3 [nbis] für jeden aufgenommenen Finger der Datenbank SD14 angewendet. Dieser Algorithmus überprüft Minutien-basiert, ob zwei Fingerabdrücke hinsichtlich eines gewissen Schwellwerts übereinstimmen. Um folglich BOZORTH3 anwenden zu können, haben wir in einem ersten Schritt mittels des NIST Feature-Extraktions-Algorithmus mindtct [nbis] alle Minutien extrahiert. Anschließend haben wir für jedes der 27.000 Bildpaare (bestehend aus Anfrage- und Referenzaufnahme) den Comparison-Algorithmus ausgeführt, und alle Abdrücke verworfen die nicht mindestens einen Schwellwert von 40 erreicht haben. Diesen Wert haben wir direkt aus der Originaldokumentation des BOZORTH3 übernommen, da er erfahrungsgemäß eine „wahre“ Übereinstimmung indiziert. Insgesamt erfüllen 17.723 Bildpaare diese Eigenschaft und eignen sich daher für den Validierungsprozess unserer Verfahren. Folglich existieren in unserem Datenbestand Personen, für die nicht für alle zehn Finger ein Bild vorhanden ist. Bei 78 Personen erfüllte kein Fingerabdruck die Anforderung, dass die Aufnahme die für die Anfrage verwendet werden soll in ausreichendem Maße mit der Referenzaufnahme übereinstimmt. Folglich reduzierte sich die Anzahl der Personen in der Datenbank von 2.700 auf 2.622.

Um verlässliche Ergebnisse zu erhalten, wurden Minutien verworfen, sofern sie den festgelegten Anforderungen bezüglich lokaler Bildeigenschaften nicht entsprachen. Diese Entscheidung trafen wir mithilfe des von mindtct ausgegebenen Qualitätsmaßes für jede extrahierte Minutie. Letztendlich wurden für jeden Fingerabdruck die 50 besten Minutien zur Weiterverarbeitung behalten.

Anschließend werden die Referenzdaten entsprechend dem Fuzzy-Vault-Verfahren [JS02] verschleiert, das in Abschnitt 3 detailliert erklärt wird. Die Grundidee basiert auf einer Verschleierung der Daten durch zusätzliche Punkte (Chaff-Points), wodurch die eigentliche Information einem potentiellen Angreifer nicht zugänglich gemacht wird. Aus diesem Grund erzeugten wir für jeweils 50 „wahre“ Minutien 200 Chaff-Points auf der Gesamtfläche des Bildes.

Dabei wurde ein Mindestabstand d_{min} = 18 zwischen Chaff-Points untereinander und zwischen Chaff-Points und Minutien eingehalten. Dieser Mindestabstand lässt sich auch in realen Minutiendaten beobachten und eignet sich daher für eine realistische Verschleierung der Daten. In den Referenzdaten sind somit für 17.723 Fingerabdrücke jeweils 50 „echte“ Minutien (insgesamt 886.150) und 200 Chaff-Points (insgesamt 3.544.600) gespeichert. In der Summe besteht die Referenzdatenbank folglich aus 4.430.750 Einträgen.

6.3.1.2 Synthetische Daten zur Evaluierung der Robustheit

Die Testergebnisse anhand der Datenbank SD14 dienen in erster Linie einer Evaluierung unserer Verfahren unter realen Bedingungen. Zudem interessierten wir uns darüber hinaus für die Stabilität von der Verfahren gegenüber bestimmten Effekten. Zu diesem Zweck generierten wir ausgehend von der realen Datenbank SD14 synthetische Daten, die wir gezielt manipulierten um spezielle Eigenschaften zu untersuchen.

Die Referenzdaten dieser Tests entsprechen den Referenzdaten der Datenbank SD14, die entsprechend Abschnitt 6.3.1.1 aufgebaut wurden. Um die Stabilität der Verfahren gegenüber den zu untersuchenden Effekten beurteilen zu können, wurden jeweils die Minutien von 200 zufällig ausgewählten Personen der Referenzdaten, für die die Information für alle zehn Finger in der Datenbank gespeichert ist, gezielt wie folgt manipuliert und als Anfrage verwendet.

Rotation der Daten

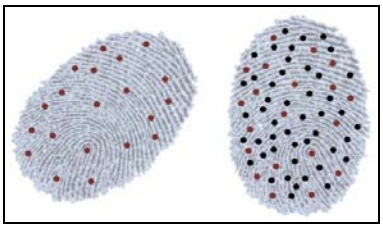

Abbildung 58: Rotierter Abdruck

Dieser Datensatz dient dazu, die Auswirkung unterschiedlicher Rotationen auf die Robustheit unserer Verfahren zu untersuchen. Abbildung 58 illustriert diese Problematik anhand eines kleinen Beispiels. Rechts ist der verschleierte Fingerabdruck der Referenzdatenbank zu sehen. Der Fingerabdruck der Anfrage auf der linken Seite enthält die gleichen Minutien wie der Referenzabdruck. Da der Finger für die Anfrage jedoch rotiert aufgenommen wurde, müssen die Verfahren mit gedrehten Koordinaten der Minutien umgehen können. Zu diesem Zweck generierten wir Daten, deren Minutien einen Rotationswinkel α im Intervall [2°, 4°, 6°, ..., 20°] besitzen. Formal wurde dieser Schritt wie folgt durchgeführt. Die X- und Y-Koordinaten der Minutien des ursprünglichen Abdrucks wurden im ersten Schritt hin zum Ursprung verschoben. Dann wurden durch Anwendung einer Drehmatrix die Koordinaten um den Winkel α gegen den Uhrzeigersinn gedreht. Damit ergeben sich folgende neue Koordinaten:

- $X_{rotiert} = \cos(\alpha) * X - \sin(\alpha) * Y$
- $Y_{rotiert} = \sin(\alpha) * X + \cos(\alpha) * Y$

Abschließend wurden diese Koordinaten wieder um den gleichen Betrag vom Ursprung weg verschoben. Minutien, die durch die Rotation den Bildbereich verlassen, wurden bei den Tests nicht berücksichtigt.

Verschiebung der Daten

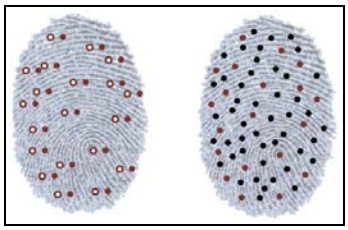

Abbildung 59: Verschobener Abdruck

Abbildung 59 zeigt einen Fingerabdruck, der im Vergleich zum Abdruck der Referenzdatenbank nach rechts verschoben wurde. Die weißen Kreise in der linken Abbildung repräsentieren die ursprüngliche Position der Minutie im Fingerabdruck der Anfrage. Die roten Kreise geben die nach rechts verschobenen Positionen wieder. Um die Robustheit aller Verfahren gegenüber verschobenen Daten zu testen, haben wir die Daten dahingehend manipuliert, dass die räumlichen Koordinaten der Minutien um jeweils 5, 10, 15, …, 50 Pixel verschoben wurden. Wie bereits für die Rotation der Daten beschrieben, wurden auch in diesem Fall, verschobene Minutien, deren Koordinaten außerhalb des Bildbereichs lagen, für die Tests ausgeschlossen.

Verlust und Hinzunahme einzelner Minutien

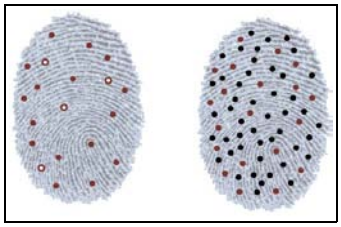

Abbildung 60: Fehlende Minutien

Beispielsweise durch Ungenauigkeiten während des Scan-Vorgangs kann es dazu kommen, dass Minutien für den Abdruck in den Referenzdaten aufgenommen wurden, für die Anfrage jedoch nicht berücksichtigt wurden und umgekehrt. In Abbildung 60 enthält die Aufnahme der Referenzdaten auf der rechten Seite Minutien, die für die Anfrage nicht zur Verfügung stehen. Diese Minutien sind auf der linken Seite weiß gekennzeichnet. Abbildung 61 visualisiert im Gegensatz dazu den Fall, dass auf dem Abdruck der Anfrageminutien enthalten sind, die in den Referenzdaten nicht zu finden sind. Diese zusätzlichen Minutien sind auf der linken Seite in grüner Farbe dargestellt. Um zu untersuchen, inwieweit die Verfahren sensibel gegenüber zusätzlichen bzw. fehlenden Minutien reagieren, wurden den Anfragefingern im Vergleich zu den Referenzdaten 4, 8, 12, … 40 Minutien im entsprechenden Bildbereich zusätzlich zufällig hinzugefügt bzw. bestehende Minutien gelöscht.

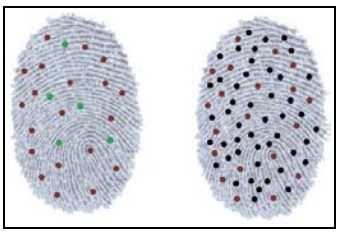

Abbildung 61: Zusätzliche Minutien

6.3.2 Evaluierung der Identifikationslösungen

In einem ersten Schritt werden für jedes Verfahren die optimalen Parameter anhand der Datenbank NIST SD 14 bestimmt. Die Optimierung bezieht sich auf eine Minimierung der Laufzeit bei gleichzeitig optimaler Effektivität. Dabei wird die Laufzeit, die für die Bestimmung der sortierten Kandidatenliste benötigt wird bzw. den Zeitaufwand, den der Aufbau der Indexstruktur bei BioSimJoin* erfordert, berücksichtigt. Anschließend evaluieren wir die Robustheit unserer Verfahren gegenüber gedrehten oder verschobenen Daten und untersuchen, inwieweit fehlende bzw. hinzukommende Minutien die Ergebnisse beeinflussen. Alle Ergebnisse sind stets über alle Personen der entsprechenden Datenmenge gemittelt. Das bedeutet, dass in allen Experimenten jede Person ein Mal als Anfrage verwendet wird. Die Resultate entsprechen somit jeweils repräsentativen Durchschnittswerten.

Die Identifikation einer Anfrageperson wird innerhalb der Experimente für die Verfahren BioSimJoin, BioSimJoin* und BioNN mittels drei Finger durchgeführt, um eine höhere Selektivität der Verfahren zu erzielen. Das bedeutet, dass diese Verfahren parallel für drei Abdrücke arbeiten. Da jedoch nicht für alle Personen die Informationen über alle Fingertypen vorliegen, wird die Auswahl der drei Finger anhand einer bestimmten Priorisierung vorgenommen, dargestellt in Abbildung 62. Demnach werden also die beiden Zeigefinger in Kombination mit dem linken Mittelfinger mit höchster Präferenz für den Identifikationsvorgang ausgewählt. Die beiden kleinen Finger werden nur im Ausnahmefall herangezogen, da diese die geringste Information tragen.

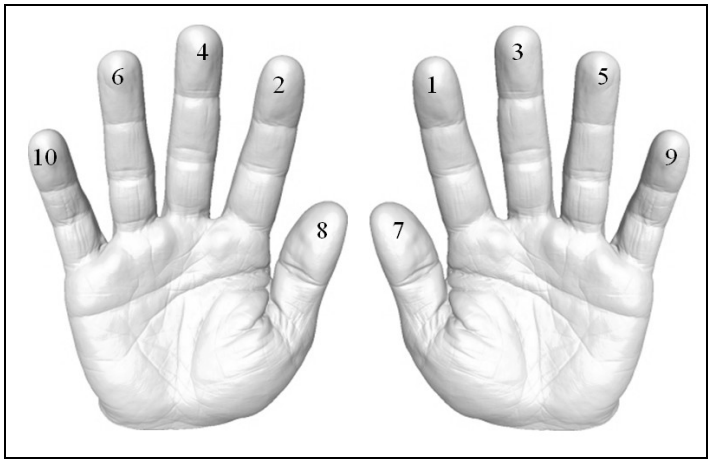

Abbildung 62: Priorisierung der Fingertypen

Die Zeitmessungen aller Experimente wurden für die Verfahren BioSimJoin, BioSimJoin* und BioNN auf einem Dell PowerEdge 6800 (Baujahr 2006) durchgeführt. Dieser ist mit 4 * 2 Intel Dual Core Xeon 7120 M CPUs, also insgesamt acht Kernen mit je 3,0 GHz Taktfrequenz und 4 MB L3 Cache bzw. 32 GB RAM ausgestattet. Als Betriebssystem ist Open-Suse 10.2 (64-Bit) mit Linux-Kernel 2.6.18 installiert.

Für die Verfahren GeoMatch und Matrix-Comperator wurden entsprechende Zeitmessungen auf einem PC mit folgender Software- und Hardware - Ausstattung durchgeführt:
CPU: 2x Intel XEON E5345 4 x 2.33 Ghz, **Speicher:** 16 GB, **Betriebssystem:** Windows XP (64 Bit), **Java:** Java JDK 6.0

6.3.2.1 GeoMatch

Für den Ansatz des GeoMatch (vergleiche Abschnitt 6.2.1) wird lediglich ein Finger für die Identifikation verwendet. Die Zeitmessungen beziehen sich lediglich auf die Zeit die für die Rechenoperationen benötigt werden (I/O-Zeit ist nicht inbegriffen). Für die angegebene Rankingposition entspricht 1 der optimalen Positionierung und |DB| der schlechtesten Positionierung. Um aussagekräftige Werte für Laufzeit und Rankingposition zu erhalten, wurden die Ergebnisse aller Experimente über 200 zufällige Anfragen gemittelt. Falls nicht anders angegeben umfasst die Datenbank ebenfalls 200 zufällig ausgewählte Referenzen (stets inklusive aller Anfragetemplates).

Beschränkung der Dreiecke durch Beschränkung der Seitenlängen:

Wie in Kapitel 6.2.1 erläutert wurde, werden nicht alle möglichen Dreiecke für den Template-Vergleich benötigt und müssen für eine effiziente Suche eingeschränkt werden. Dies erfolgt hier über eine untere, sowie obere Schranke für die Seitenlängen eines Dreiecks. Um einen möglichst genauen Vergleich und dabei eine möglichst effiziente Berechnung zu ermöglichen werden für ein Datenbanktemplate und jede Belegung für untere und obere Schranke zwei Werte betrachtet: Anzahl der verlorenen Minutien (Minutien, welche in keinem reinen Minutiendreieck mehr vorkommen) (Abbildung 63,65) und Anzahl der Dreiecke insgesamt (Abbildung 64, 66). Alle vier Graphen wurden über 40 Anfragen gemittelt. Die zuvor genannten Bedingungen der Genauigkeit und Effizienz stehen im Allgemeinen im Widerspruch und es gilt einen günstigen Trade-off

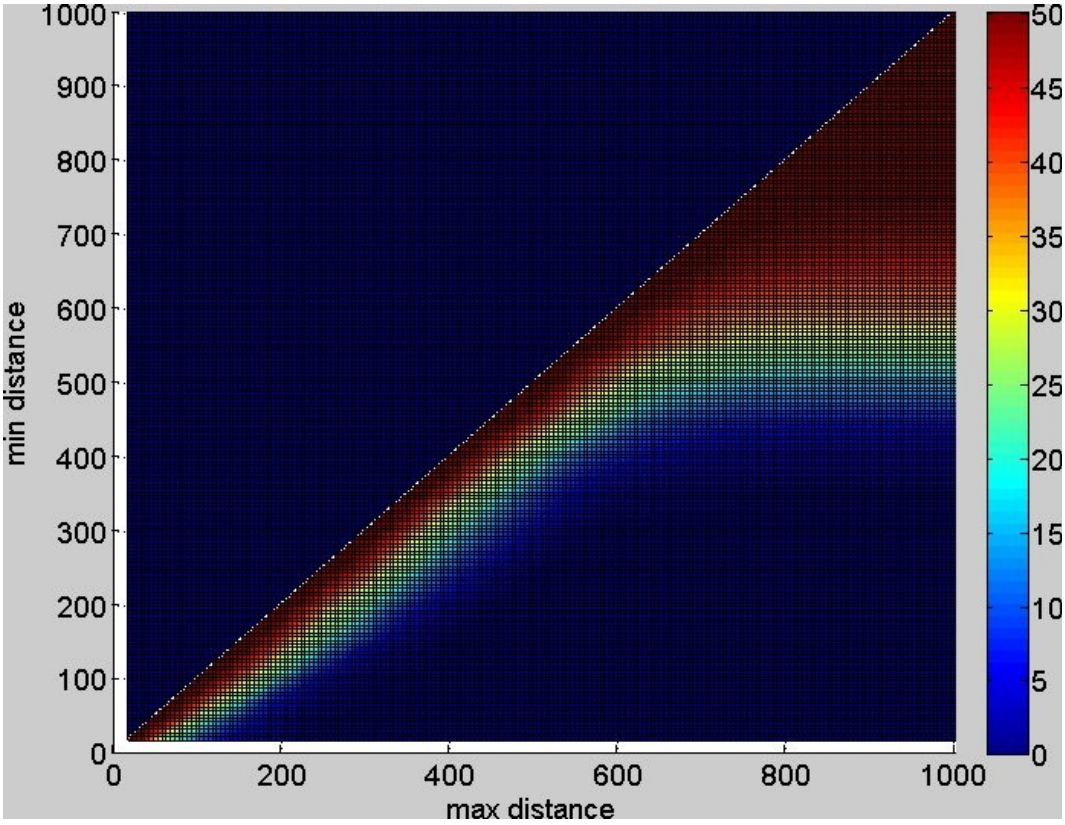

Abbildung 63: Realdaten der NIST SD14: durchschnittliche Anzahl verlorener Minutien in einem Datenbanktemplate bei Beschränkung der Seitenlängen aller Dreiecke

zwischen ihnen zu bestimmen. Im Falle der Realdatenbank zeigt sich bei einer Einschränkung auf Dreiecke lediglich mit Seiten der Länge zwischen 18 und 150 Pixeln als geeignet. Hier ist der Informationsverlust durch verlorene Minutien akzeptabel bei gleichzeitig starker Verringerung der Gesamtzahl aller Dreiecke.

Eine entsprechende Auswertung auf Daten, bei denen sowohl Minutien als auch Chaff-Points gleichverteilt vorliegen, erweist sich eine andere Einschränkung (wie z.B. 100 bis 200 Pixel) als sinnvoll (Abbildungen 65,66). Da die Verteilung der Minutien stark vom verwendeten Minutien-Extraktionsverfahren abhängt, sind die hier vorgeführten Analysen nicht allgemein gültig, sondern sollte für abweichend generierte/extrahierte Daten erneut durchgeführt werden. Die gewählten Schwellwerte für die Seitenlängen wurden für die folgenden Experimente eher konservativ gewählt. Das heißt das nur ein sehr geringer Verlust an tatsächlichen Minutien zu erwarten ist. Da es sich bei dem Ansatz GeoMatch lediglich um ein approximativen Vergleich zweier Templates handeln soll, ist allerdings durchaus auch ein höherer Verlust zu vertreten. Je größer die Seitenbeschränkungen, desto höher die Anzahl der verlorenen Minutien, desto approximativer ist der Vergleich. Gleichzeitig sinkt allerdings auch die Anzahl der relevanten Dreiecke eines Datenbankpatterns, wodurch sich die Vergleichszeit verkürzt.

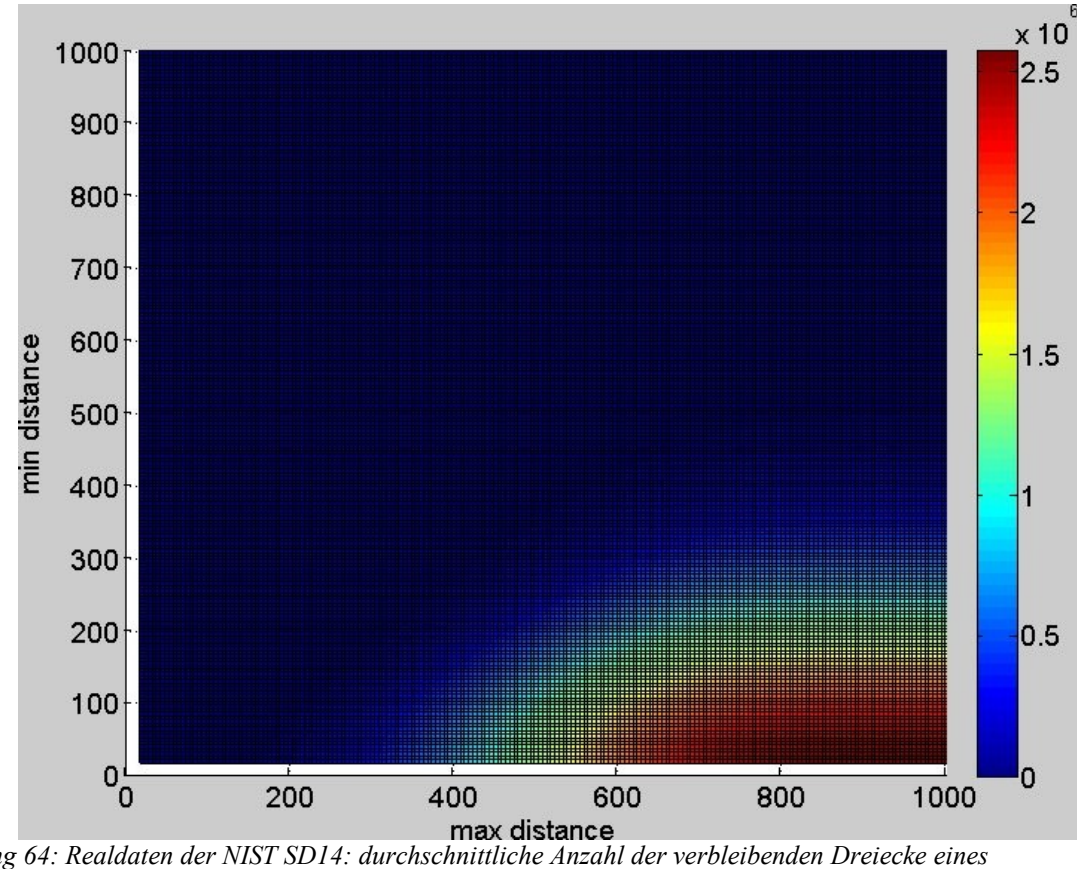

Abbildung 64: Realdaten der NIST SD14: durchschnittliche Anzahl der verbleibenden Dreiecke eines Datenbanktemplates bei Beschränkung der Seitenlängen aller Dreiecke

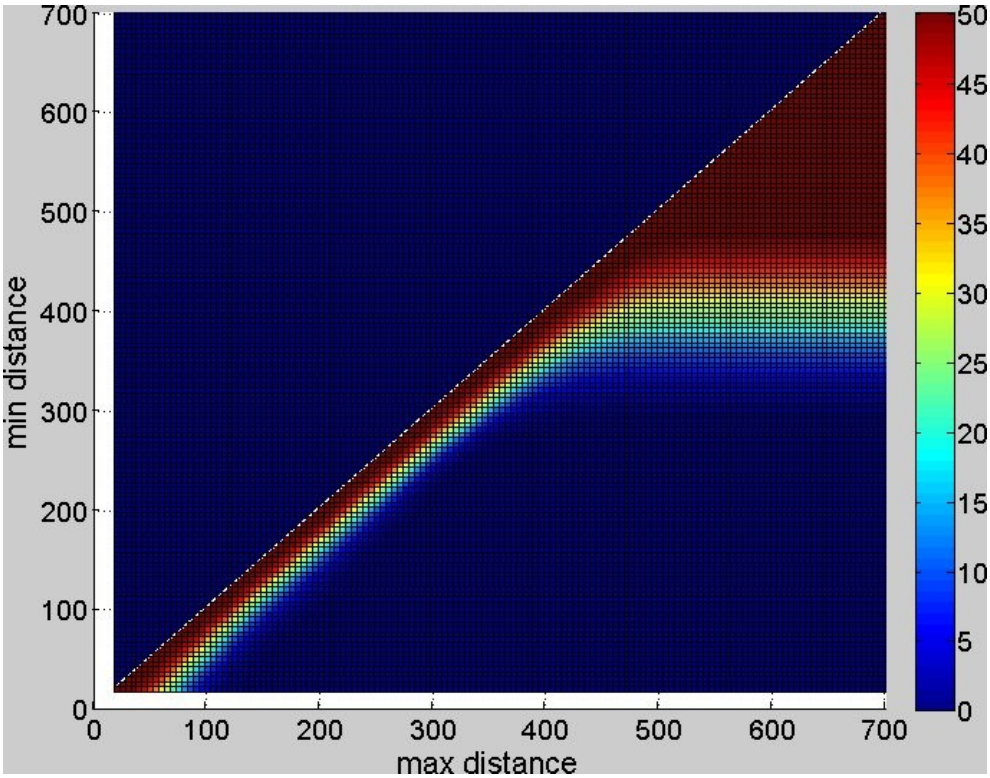

Abbildung 65: Synthetische Daten: durchschnittliche Anzahl verlorener Minutien in einem Datenbanktemplate bei Beschränkung der Seitenlängen aller Dreiecke

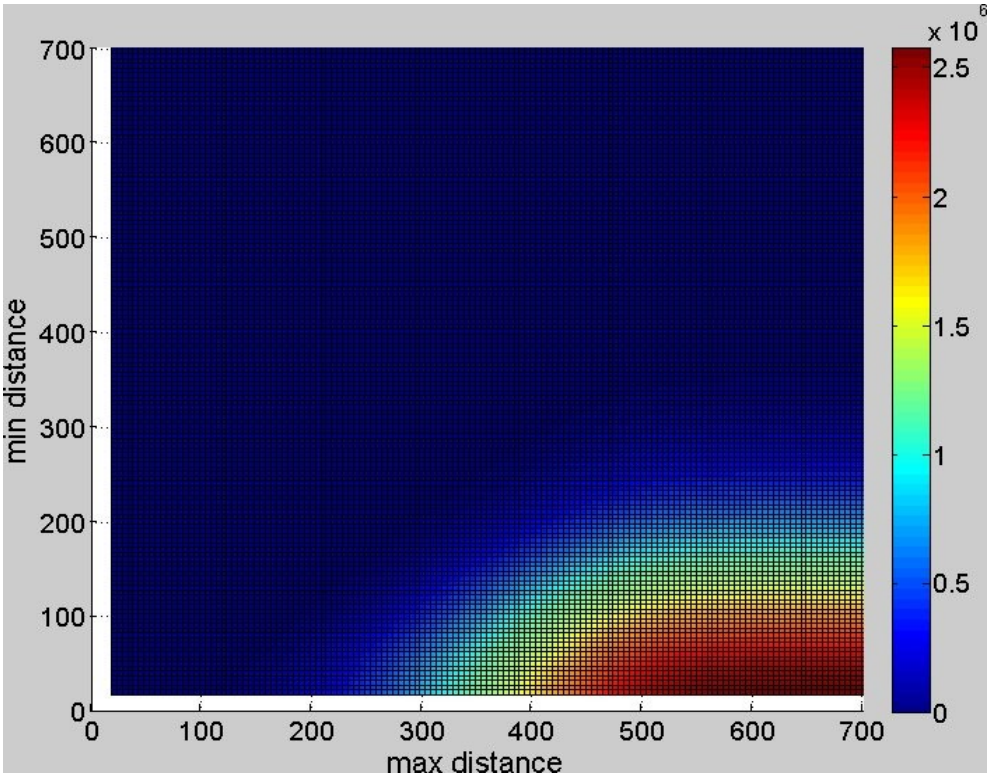

Abbildung 66: Synthetische Daten: durchschnittliche Anzahl der verbleibenden Dreiecke eines Datenbanktemplates bei Beschränkung der Seitenlängen aller Dreiecke

Schätzung des optimalen Toleranzparamters δ

In dieser Versuchsreihe wird die Auswirkung des Toleranzfehlers δ auf die Qualität des Rankings untersucht. Da zwei Minutien ab einem Abstand von 18 Pixeln als unterschiedlich angesehen werden können, werden für die Fehlertoleranz des paarweisen Abstands (einer Seite eines Dreiecks) Werte zwischen 0 und 36 untersucht.

Das Ergebnis dieser Untersuchung ist in Abbildung 67 dargestellt. Es hier deutlich zu erkennen, dass die Laufzeit sehr stark von der Fehlertoleranz δ beeinflusst wird. Es liegt sogar eine exponentielleAbhängigkeit vor. Mit einer höheren Fehlertoleranz erweist sich für jedes Dreieck des Anfragepatterns eine immer höhere Anzahl von Dreiecken des Datenbankpatterns als ähnlich, sodass hier nahezu alle Dreiecke durchsucht werden müssen und keine Abbruchkriterien mehr greifen. Als Konsequenz schmälert eine wachsende Fehlertoleranz die Effizienz des Verfahrens, sodass ein kleiner Wert für δ zu bevorzugen ist.

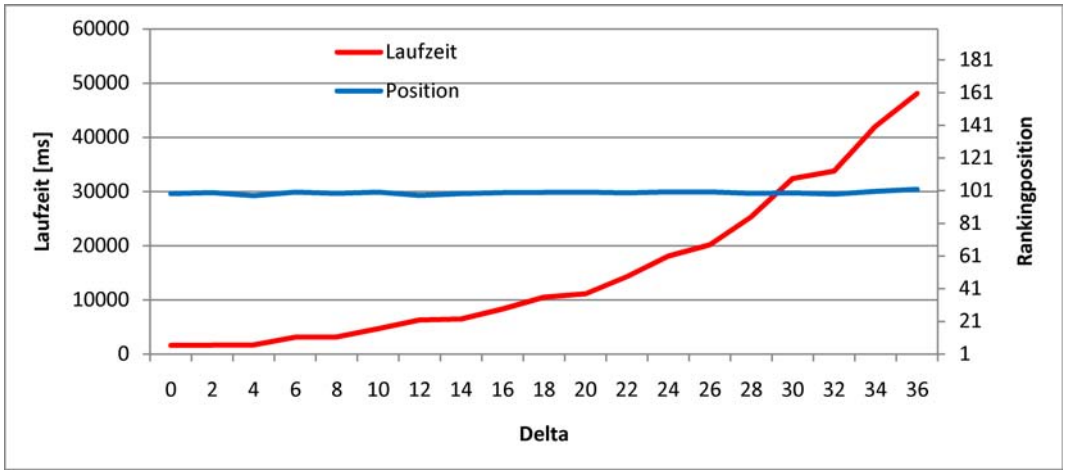

Abbildung 67: Realdaten NIST SD14: Einfluss der Fehlertoleranz delta auf Laufzeit und Rankingposition

Für den Einfluss eines variierenden δ auf die Rankingposition, zeigt sich für die Datenbank NIST SD14 kaum Auswirkung. Die Position des der passenden Datenbankreferenz ist gemittelt stets ca. in der Mitte des Rankings. Da sich somit die Wahl von δ lediglich auf die Laufzeitbetrachtung beschränkt, wodurch ein möglichst kleines δ zu bevorzugen wäre, wurden die Auswirkungen der Deltavariation auf die Rankingposition auch bei synthetischen Daten untersucht(Abbildung 68). Dazu wurden auf einer Datenbank der Größe 300 die Anfragen aller Objekte gemittelt und die Rankingposition für den einfachen Dreiecksvergleich (blau) (Dreiecksvergleich lediglich anhand der Kantenlängen, ohne Berücksichtigung der Winkel) sowie den verbesserten Dreiecksvergleich mit zusätzlicher Winkelbetrachtung (rot) betrachtet. Es zeigt sich das mit steigendem δ tatsächlich die Rankingqualität sinkt, sodass kleinere Werte für δ zu bevorzugen sind. Anschaulich wird hier noch einmal experimentell der Effektivitätsgewinn durch zusätzliche Einbeziehung der globalen Orientierung der Dreiecke demonstriert. Durch den zu großen Freiheitsgrad eines einfachen Seitenlängenvergleichs der Dreiecke, gibt es für die endgültige Bewertung zu viele falschen Akzeptanzwerte. Es ist klar ersichtlich, dass die Winkelberücksichtigung ein wesentlicher Faktor für die Selektivität dieses Ansatzes darstellt. Für den alternativen Ansatz ist hier eine deutlich klarere Tendenz für die steigende Fehlertoleranz zu erkennen.

Da sich ein kleiner Fehlertoleranzwert positiv auf Laufzeit und Rankingposition auswirken, wird für die folgenden Experimente eine Wert von $\delta = 1$ gewählt.

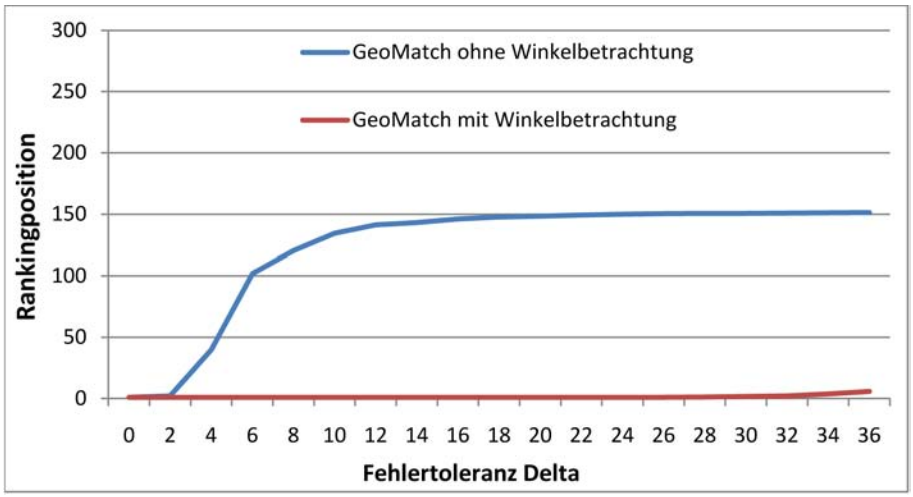

Abbildung 68: Synthetische Daten: Einfluss der Fehlertoleranz delta auf die Rankingposition

Skalierung der Datenbankgröße

In dieser Versuchsreihe wird das Laufverhalten sowie die Selektivität des Ansatzes GeoMatch für steigende Datenbankgröße untersucht. Für dieses Experiment wurde die Datenbankgröße zwischen 1 und 17.000 Subjekten variiert. Für das Verfahren wurde $\delta = 1$ gesetzt. Abbildung 69 zeigt, dass die Laufzeit linear mit der Datenbankgröße wächst und sich die Rankingposition im Mittel stets im selben Bruchteil des Rankings befindet.

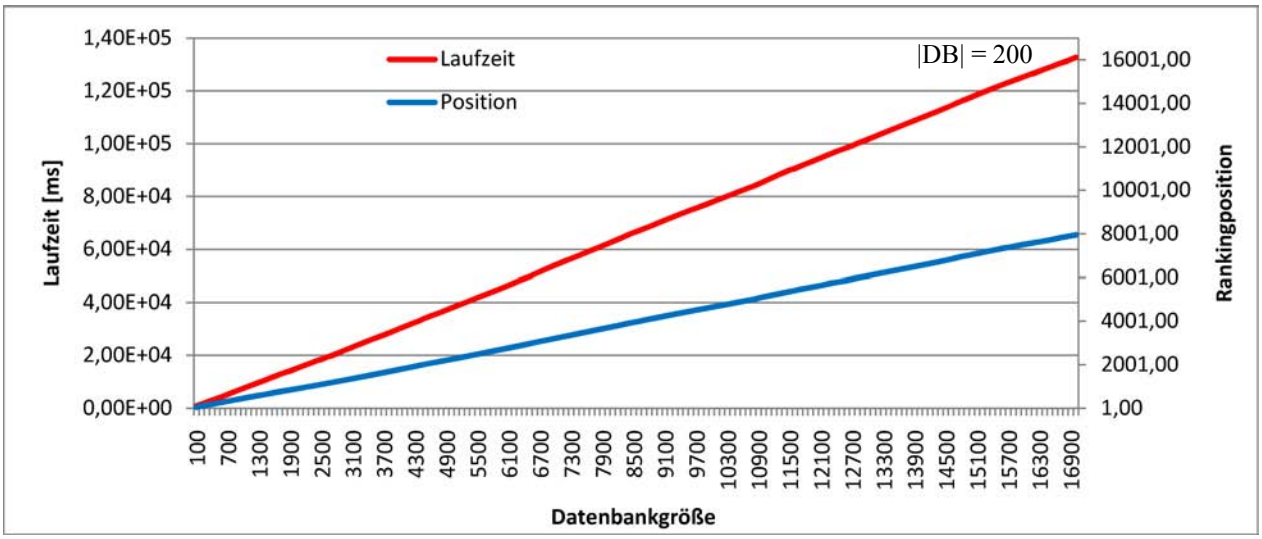

Abbildung 69: Realdaten NIST SD14: Einfluss der Datenbankgröße auf Laufzeit und Rankingposition.

Evaluierung der Robustheit bzgl. verschiedener Datenungenauigkeiten

In den folgenden Experimenten wird die Robustheit von GeoMatch bzgl. verschiedener Datenungenauigkeiten wie z.B. Rotation, Dehnung, Stauchung, lokale Minutienfehler, sowie fehlende oder neue Minutien untersucht. Es werden 200 Anfragen auf 200 Datenbanktemplates gemittelt.

Robustheit gegenüber Rotationen

Im ersten Experiment wird untersucht, inwiefern GeoMatch robust gegenüber globalen Rotationen des Anfragetemplates im Vergleich zum Datenbanktemplate ist. Da in diesem Ansatz keinerlei globale Lagepositionen, sondern lediglich globale Orientierungsinformationen verwendet werden, ist eine solche Robustheit zu erwarten! Betrachtet werden Rotationen von 0° bis 360°. Wie vermutet belegt Abbildung 70, dass GeoMatch robust gegenüber solch globalen Rotationen des Anfragetemplates, bzw. des Datenbanktemplates ist. Die gesuchte Datenbankreferenz bekommt stets die ersten Rankingpositionen.

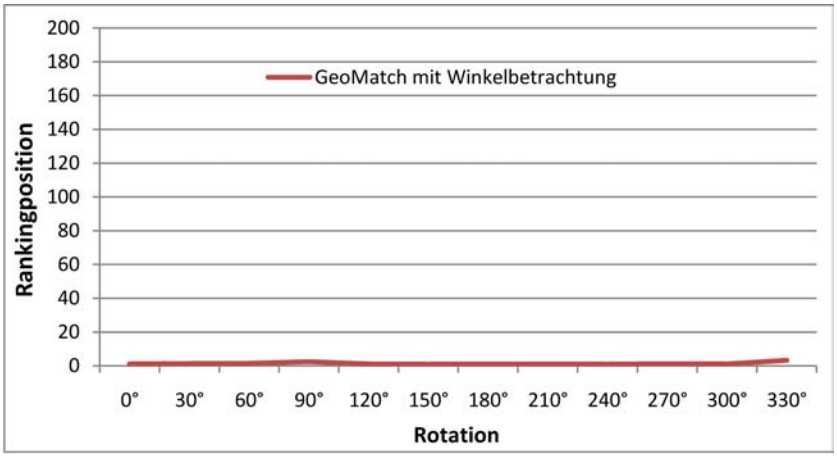

Abbildung 70: Synthetische Daten: Auswirkung einer Rotation des Anfragetemplates auf die Rankingposition

Robustheit gegenüber globalen Verschiebungen

Im diesem Experiment wird untersucht, inwiefern GeoMatch robust gegenüber globalen Verschiebungen des Anfragetemplates im Vergleich zum Datenbanktemplate ist. Aus den gleichen Gründen wie zuvor für die globale Rotation, ist ein Einfluss auf das Rankingverhalten hier nicht zu Erwarten. Es werden Verschiebungen sowohl in X- als auch in Y-Richtung von -40 bis +40 Pixel untersucht. Auch hier bestätigt Abbildung 71 die Annahmen. Die blaue Linie beschreibt die Ergebnisse für Verschiebung in X-Richtung, die rote Linie beschreibt gleiche Verschiebung in Y-Richtung. Globale Verschiebungen des Anfrage-, bzw. Datenbanktemplates wirken sich nicht auf das Rankingverhalten von GeoMatch aus.

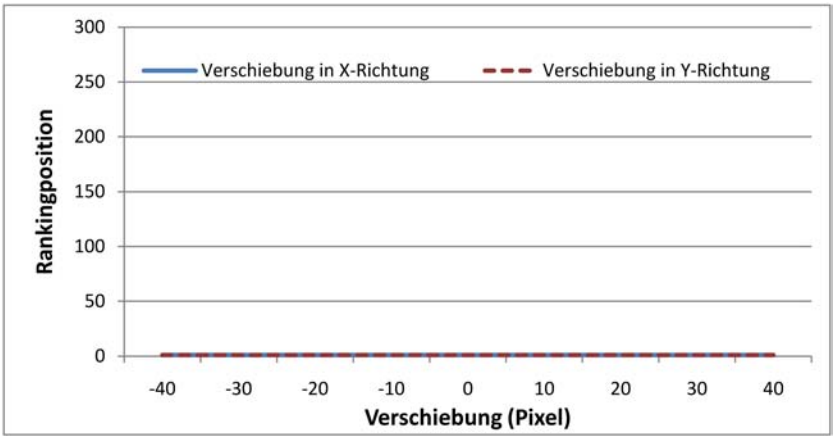

Abbildung 71: Synthetische Daten: Auswirkung auf die Rankingposition des Referenzfingers bei globalen Verschiebung des Anfragetemplates in X- und Y-Richtung

Robustheit gegenüber fehlenden oder zusätzlichen Minutien

Im diesem Experiment wird das Rankingverhalten von GeoMatch für fehlende oder zusätzliche Minutien innerhalb des Anfragetemplates untersucht. Zu diesem Zweck werden von den insgesamt 50 vorhandenen Minutien des Anfragetemplates bis zu 40 Minutien entfernt (x-Achse < 0) oder hinzugefügt (x-Achse > 0).

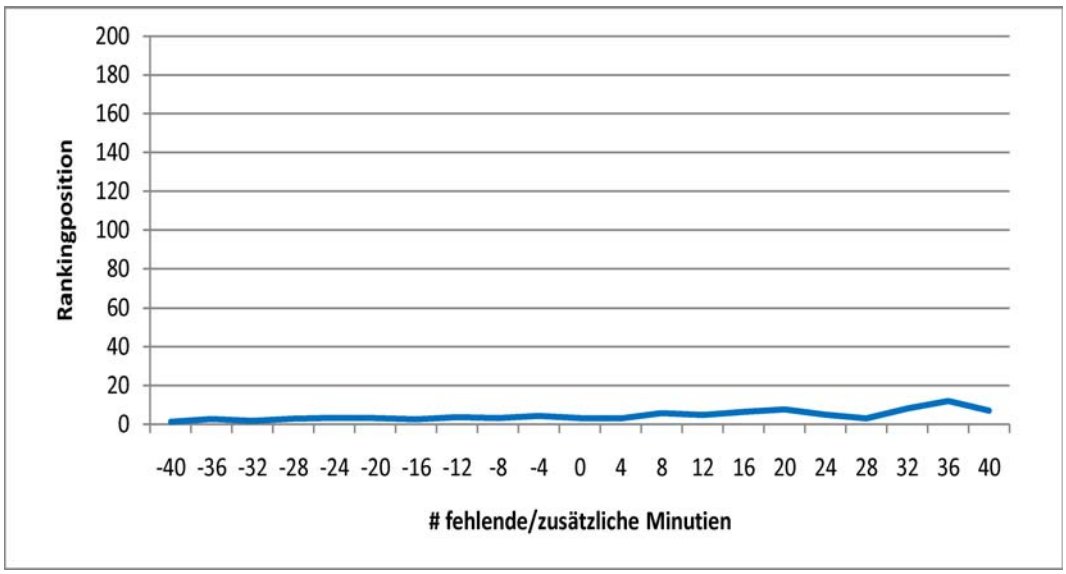

Abbildung 72: Synthetische Daten: Auswirkung auf die Rankingposition des Referenzfingers bei fehlenden (x<0) sowie zusätzlichen (x>0) Minutien innerhalb des Anfragetemplates

Die Ergebnisse aus Abbildung 72 zeigen das die Rankingposition nahezu unverändert bleibt. Die verbleibenden Minutien bzw. die ursprünglichen Minutien dominieren den Vergleich zwischen Anfrage- und Datenbanktemplates. Erst bei zu vielen neuen Minutien, verschlechtert sich die Erkennungsleistung merklich, da es dann zu falschen Übereinstimmungen der Minutien in fremden Datenbanktemplates kommt. Fallen lediglich Minutien weg, so wird der Referenzfinger unter anderem auf Grund der geringen Fehlertoleranz dennoch gut erkannt.

Durchschnittlicher Speicherverbrauch eines Datenbanktemplates

Der durchschnittliche Speicherbedarf eines Datenbanktemplates für das Verfahren GeoMatch ist abhängig von der Anzahl *m* der zu Grunde liegenden Koordinaten (Minutien und Chaff-Points), der Wahl für die *n*-Ecke, sowie die gewählte Beschränkung der Seitenlängen. Der größte Speicheraufwand entsteht, falls keine Seitenlängenbeschränkung angewendet wird und alle $\binom{m}{n}$ vielen Dreiecke mit je 28 byte (3 Seitenlängen (Double: 8 byte) und ein Winkel (Integer: 4 byte)) zu speichern sind. In den vorliegenden Experimenten wurden Daten mit je 50 Minutien und 200 Chaff-Points verwendet, sowie eine Seitenlängenbeschränkung auf 18 bis 150 Pixel festgelegt. Für die vorliegenden Experimente hat ein Datenbanktemplate im Durchschnitt 22.770,39 Dreiecke und benötigt somit durchschnittlich 0,64 MB Speicherplatz.

6.3.2.2 Matrix-Comparator

Für den Ansatz des Matrix-Comparators (vergleiche Abschnitt 6.2.2) wird lediglich ein Finger für die Identifikation verwendet. Die Zeitmessungen beziehen sich lediglich auf die Zeit die für die Rechenoperationen benötigt werden (I/O-Zeit ist nicht inbegriffen). Für die angegebene Rankingposition entspricht 1 der optimalen Positionierung und |DB| der schlechtesten Positionierung. Um aussagekräftige Werte für Laufzeit und Rankingposition zu erhalten, wurden die Ergebnisse aller Experimente über 100 zufällige Anfragen gemittelt. Falls nicht anders angegeben umfasst die Datenbank ebenfalls 100 zufällig ausgewählte Referenzen (stets inklusive aller Anfragetemplates).

Schätzung des opimalen Toleranzparamters δ

In dieser Versuchsreihe wird die Auswirkung des Toleranzfehlers δ auf die Position des Referenztemplates im Ergebnisranking untersucht. Da zwei Minutien ab einem Abstand von 18 Pixeln als unterschiedlich angesehen werden können, werden für die Fehlertoleranz des paarweisen Abstands in der Distanzmatrix Werte zwischen 0 und 26 untersucht.

Die Ergebnisse in Abbildung 73 zeigen, dass mit zunehmender Fehlertoleranz die für einen Vergleich benötigte Zeit rapide abnimmt. Dies ist damit zu erklären, dass mit höherer Fehlertoleranz schneller eine Übereinstimmung zwischen Zellen des Anfragepatterns und des Datenbankpatterns festgestellt werden können, ohne dass die vollständige Distanzmatrix des Datenbankpatterns untersucht werden muss. Je größer also δ gewählt wird, desto effizienter arbeitet das Verfahren.

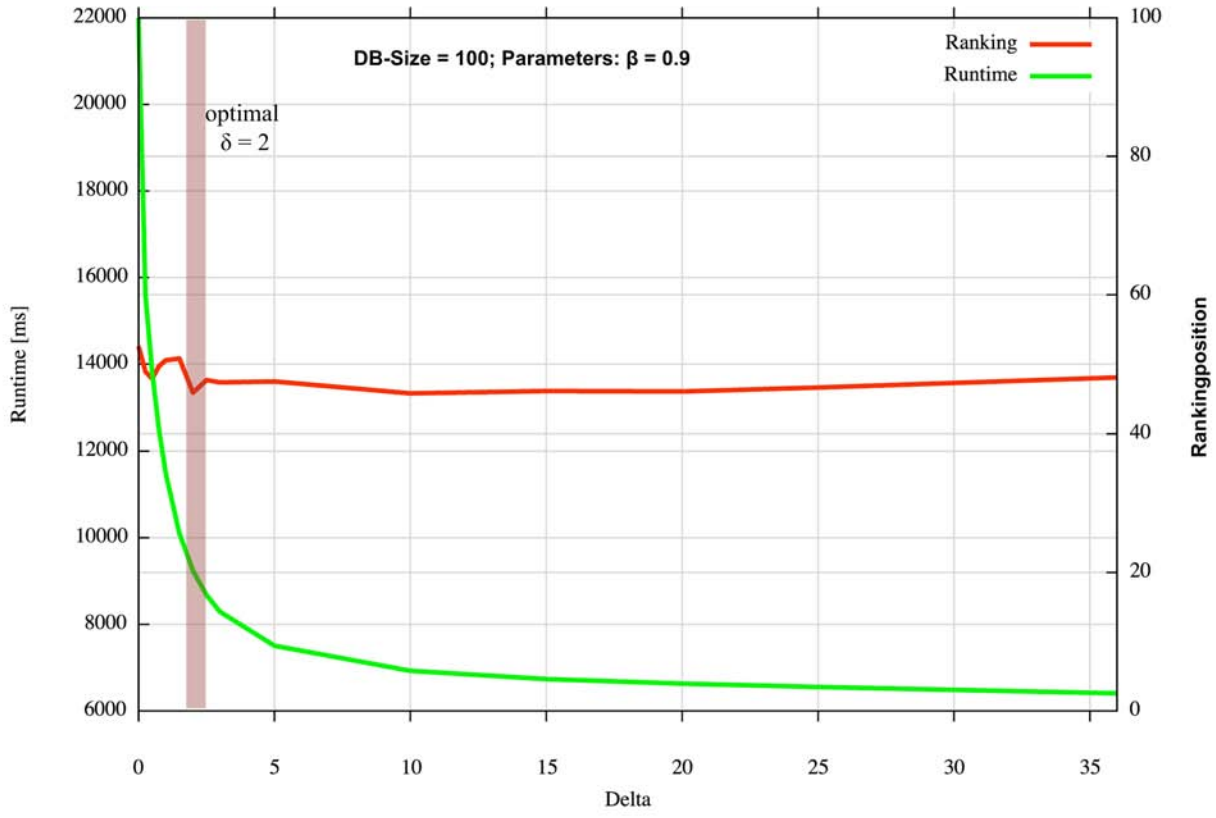

Abbildung 73: Realdaten NIST SD14: Einfluss der Fehlertoleranz delta auf die Rankingposition sowie die Laufzeit

Für die Rankingposition zeigte sich auf der NISTSD14 Datenbank bei Variation von δ kaum eine Veränderung. Es ist lediglich eine kleine Verbesserung bei $\delta = 2$ festzustellen. Da mit steigender Fehlertoleranz immer stärkere Unähnlichkeiten zwischen Anfrage- und Datenbanktemplate toleriert werden, sollte δ für eine ausreichende Genauigkeit des Verfahrens nicht zu groß gewählt werden. Entsprechende Experimente auf synthetischen Daten zeigten den erwarteten Verlust der Rankinggenauigkeit mit zunehmender Fehlertoleranz δ .

Insgesamt empfehlen wir eine Fehlertoleranz von $\delta = 2$, da hier ein Günstiger Trade-off zwischen Effizienz und Genauigkeit besteht. Diese Fehlertoleranz wird für die folgenden Experimente übernommen.

Schätzung des opimalen Paramters β

In dieser Versuchsreihe wird die Auswirkung des Schwellwertes β auf die Position des Referenztemplates im Ergebnisranking untersucht.

Die Ergebnisse in Abbildung 74 zeigen, dass es mit variierendem Schwellwertes β , wie zu erwarten, nur sehr geringe Variationen bzgl. der Laufzeit gibt; hier im Bereich von 10 ms. Für die Rankingposition zeigen sich hier ebenfalls keine starken Abhängigkeiten für die Datenbank NISTSD14. Allerdings lässt sich im Bereich von $\beta = 0.6$ eine kleinere Verbesserung der Rankingposition erkennen. Im den folgenden Experimenten wird daher ein Schwellwert von $\beta = 0.6$ angesetzt. Das bedeutet, eine Minutie m_i im Datenbanktemplate muss mindestens 60 % der Nachbardistanzen einer Minutie des Anfragetemplates m_j aufweisen, damit m_i und m_j als übereinstimmend erkannt werden.

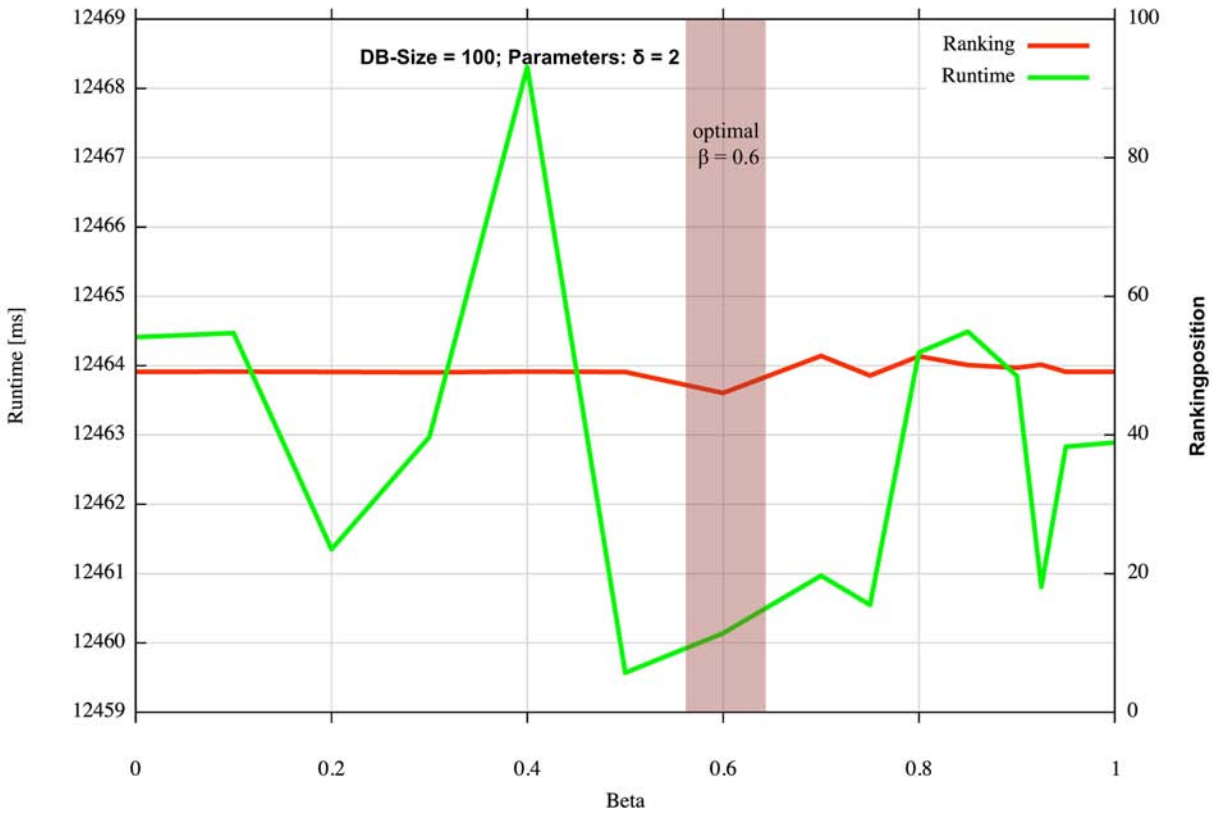

Abbildung 74: Realdaten NIST SD14: Einfluss des Schwellwertes beta auf die Rankingposition

Skalierung der Datenbankgröße

Für das folgende Experiment wird das Laufverhalten sowie die Selektivität des Ansatzes Matrix-Comparator für eine steigende Datenbankgröße untersucht. Für dieses Experiment wurde die Datenbankgröße zwischen 1 und 100 Personen variiert. Für das Verfahren wurde $\delta = 2$ und $\beta = 0{,}6$ gesetzt. Abbildung 75 zeigt, dass die Laufzeit linear mit der Datenbankgröße wächst und sich die Rankingposition im Mittel stets im selben Bruchteil des Rankings befindet.

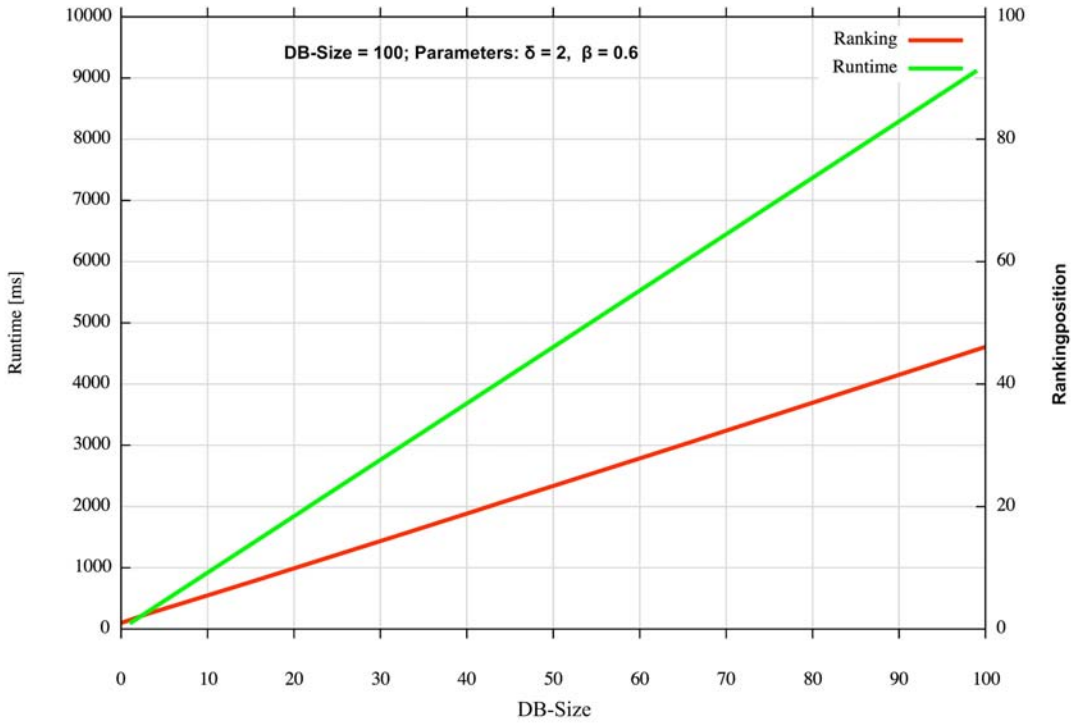

Abbildung 75: Realdaten NIST SD14: Selektivität sowie Effizienz des Matrix-Comparator Ansatzes bei steigender Datenbankgröße

Evaluierung der Robustheit bzgl. verschiedener Datenungenauigkeiten

In den folgenden Experimenten wird die Robustheit des Ansatzes Matrix-Comparator bzgl. verschiedener Datenungenauigkeiten wie z.B. Rotation, Dehnung, Stauchung, lokale Minutienfehler, sowie fehlende oder zusätzliche Minutien des Anfragetamplates untersucht.

Robustheit gegenüber Rotationen

Im ersten Experiment wird untersucht, inwiefern der Ansatz Matrix-Comparator robust gegenüber globalen Rotationen des Anfragetemplates im Vergleich zum Datenbanktemplate ist. Da in diesem Ansatz keinerlei globale Lagepositionen, sondern lediglich Distanzinformationen der Minutienpaare verwendet werden, ist kein Einfluss einer globalen Rotation zu erwarten. Betrachtet werden Rotationen von 0° bis 20°. Wie vermutet belegt Abbildung 76, dass der Ansatz Matrix-Comparator robust gegenüber solch globalen Rotationen des Anfragetemplates, bzw. des Datenbanktemplates ist. Die gesuchte Datenbankperson landet stets auf der besten Position (Position 1) des Rankings.

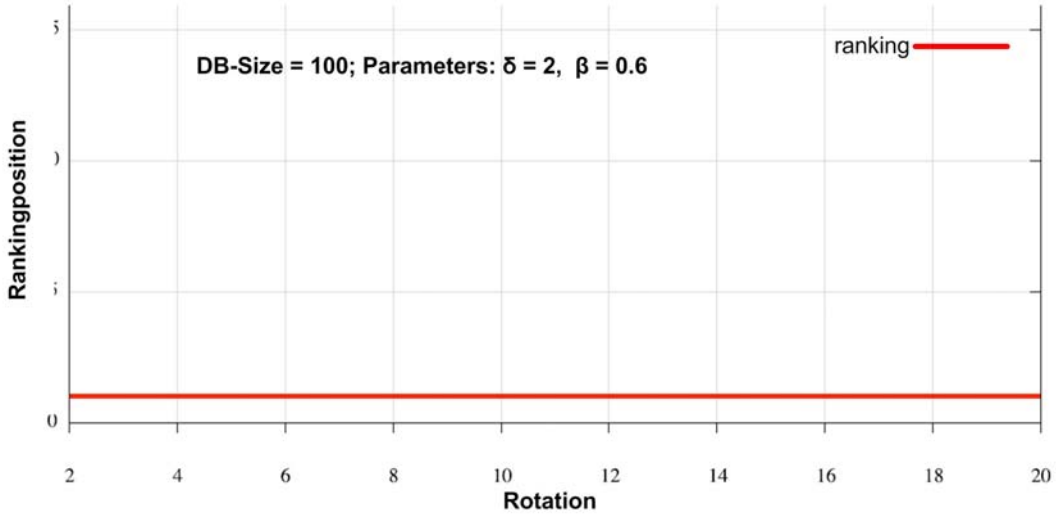

Abbildung 76: Synthetische Daten: Auswirkung einer Rotation des Anfragetemplates auf die Rankingposition

Robustheit gegenüber globalen Verschiebungen

Im diesem Experiment wird die Robustheit des Ansatzes gegenüber globalen Verschiebungen des Anfragetemplates im Vergleich zum Datenbanktemplate ist. Aus den gleichen Gründen wie zuvor für die globale Rotation, ist ein Einfluss auf das Rankingverhalten hier nicht zu erwarten. Es werden Verschiebungen sowohl in X- als auch in Y-Richtung von -40 bis +40 Pixel untersucht. Auch hier bestätigt Abbildung 77 die Annahmen. Die rote Linie beschreibt die Ergebnisse für eine Verschiebung in X-Richtung, die blaue Linie beschreibt gleiche Verschiebungen in Y-Richtung. Globale Verschiebungen des Anfrage-, bzw. Datenbanktemplates wirken sich nicht auf das Rankingverhalten von GeoMatch aus.

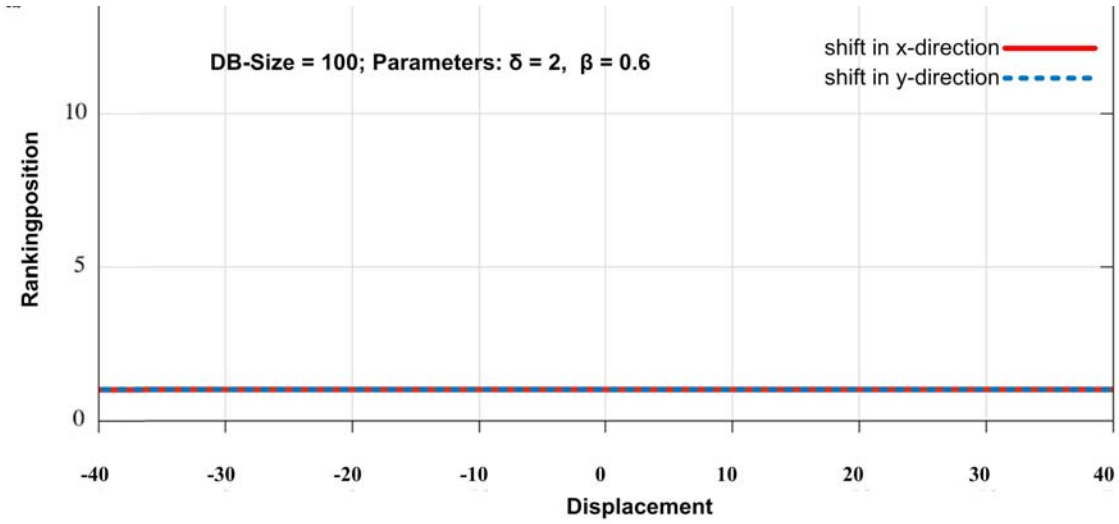

Abbildung 77: Synthetische Daten: Auswirkung auf die Rankingposition des Referenzfingers bei globalen Verschiebung des Anfragetemplates in X- und Y-Richtung

Robustheit gegenüber fehlenden oder zusätzlichen Minutien

Im diesem Experiment wird die Robustheit des Ansatzes gegenüber fehlenden sowie zusätzlichen Minutien innerhalb des Anfragetemplates im Vergleich zum Datenbanktemplate getestet. Dazu werden von den insgesamt 50 vorhandenen Minutien des Anfragetemplates bis zu 40 Minutien entfernt (x-Achse < 0) oder hinzugefügt (x-Achse > 0).

Die Ergebnisse aus Abbildung 78 zeigen das die Rankingposition nahezu unverändert bleibt bei bis zu 28 fehlenden und bis zu 12 zusätzlichen Minutien innerhalb des Anfragetempaltes. Hier dominieren die verbleibenden Minutien bzw. die ursprünglichen Minutien den Vergleich zwischen Anfrage- und Datenbanktemplates. Erst wenn mehr als 28 Minutien, also mehr als 50%, des Anfragetemplates fehlen, verschlechtert sich die Erkennungsleistung merklich. Für mehr als 12 hinzukommende Minutien verschlechtert sich die Erkennungsleistung ebenfalls, allerdings nicht im selben Ausmaß wie für fehlende Minutien.

Deutlich zu erkennen ist hier auch der starke Einfluss der Minutienanzahl auf die Laufzeit eines Vergleichs.

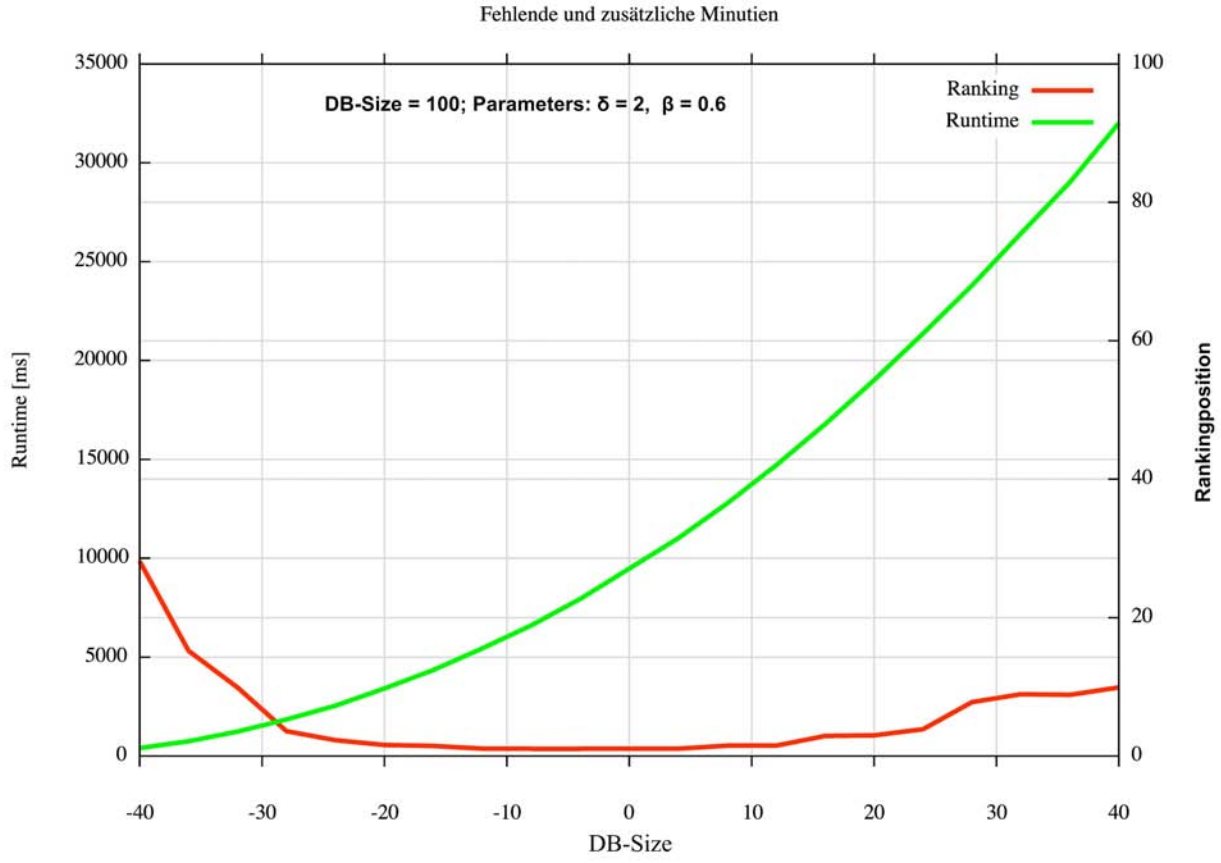

Abbildung 78: Synthetische Daten: Auswirkung auf die Rankingposition des Referenzfingers bei neuen und fehlenden Minutien innerhalb des Anfragetemplates

Durchschnittlicher Speicherverbrauch eines Datenbanktemplates

Der durchschnittliche Speicherbedarf eines Datenbanktemplates für das Verfahren Matrix-Comparator ist abhängig von der Anzahl der zu Grunde liegenden Koordinaten (Minutien und Chaff-Points). Da für diesen Ansatz die Distanzmatrix aller Koordinaten benötigt wird, ist die Platzkomplexität hier quadratisch in der Anzahl der Koordinaten (vgl. Abschnitt 6.2.2.4). In den vorliegenden Experimenten wurden Daten mit je 50 Minutien und 200 Chaff-Points verwendet, sodass insgesamt 250 x 250 Double-Werte (8 byte), insgesamt also 0,5 MB, zu speichern sind. Durch Ausnutzen der Symmetrieeigenschaft der Distanzmatrix ist jedoch ein effektiver Speicherverbrauch von 0,25 MB ausreichend.

6.3.2.3 BioSimJoin

Schätzung des optimalen Radius $\tilde{r}_{opt}$

Für das Verfahren BioSimJoin ist nur der Parameter *r*, also der Radius der Bereichsanfrage zu optimieren. Da BioSimJoin die Identifikation der Kandidatenliste ohne Indexunterstützung bewerkstelligt, und daher verhältnismäßig mehr Zeit benötigt als BioSimJoin*, wird der optimale Radius $\tilde{r}_{opt}$ auf einem Sample der Datenbank SD 14, bestehend aus 100 Personen, das einer Datenbankgröße von insgesamt 182.500 Einträgen (Chaff-Points/Minutien) entspricht.

Die rote Kurve in Abbildung 79 zeigt bis zu einem Radius-Wert von 40 die durchschnittlich benötigte Laufzeit, um die Kandidatenliste zu ermitteln. Die linke Ordinate ist hierfür im Bereich 1.200 bis 1.400 ms skaliert. Ein höherer Radius-Wert wirkt sich auf die Laufzeit allgemein negativ aus, da damit eine höhere Zahl von zu überprüfenden Kandidaten einhergeht. Die Schwankungen, die im vorderen Bereich zu beobachten sind, werden durch das Anlegen von Datenstrukturen, Objekten oder Referenzen verursacht. Die blaue Kurve stellt die durchschnittliche Position der Anfrageperson innerhalb der Kandidatenliste dar. Ein zu kleiner Radius bewirkt hier, dass sich die Anfrageperson durchschnittlich erst an späterer Stelle in der Kandidatenliste befindet. Diese Beobachtung lässt sich durch eine mangelnde Toleranz gegenüber gedrehten bzw. verschobenen Daten erklären, weil durch einen zu kleinen Radius verhältnismäßig viel Information verloren geht. Um einen bestmöglichen Kompromiss zwischen geringer Laufzeit und hoher Effektivität zu erzielen, sollte somit der Radius $\tilde{r}_{opt}$ im Bereich zwischen 12 und 16 gewählt werden, und wird für die folgenden Experimente $\tilde{r}_{opt}$ =13 gewählt.

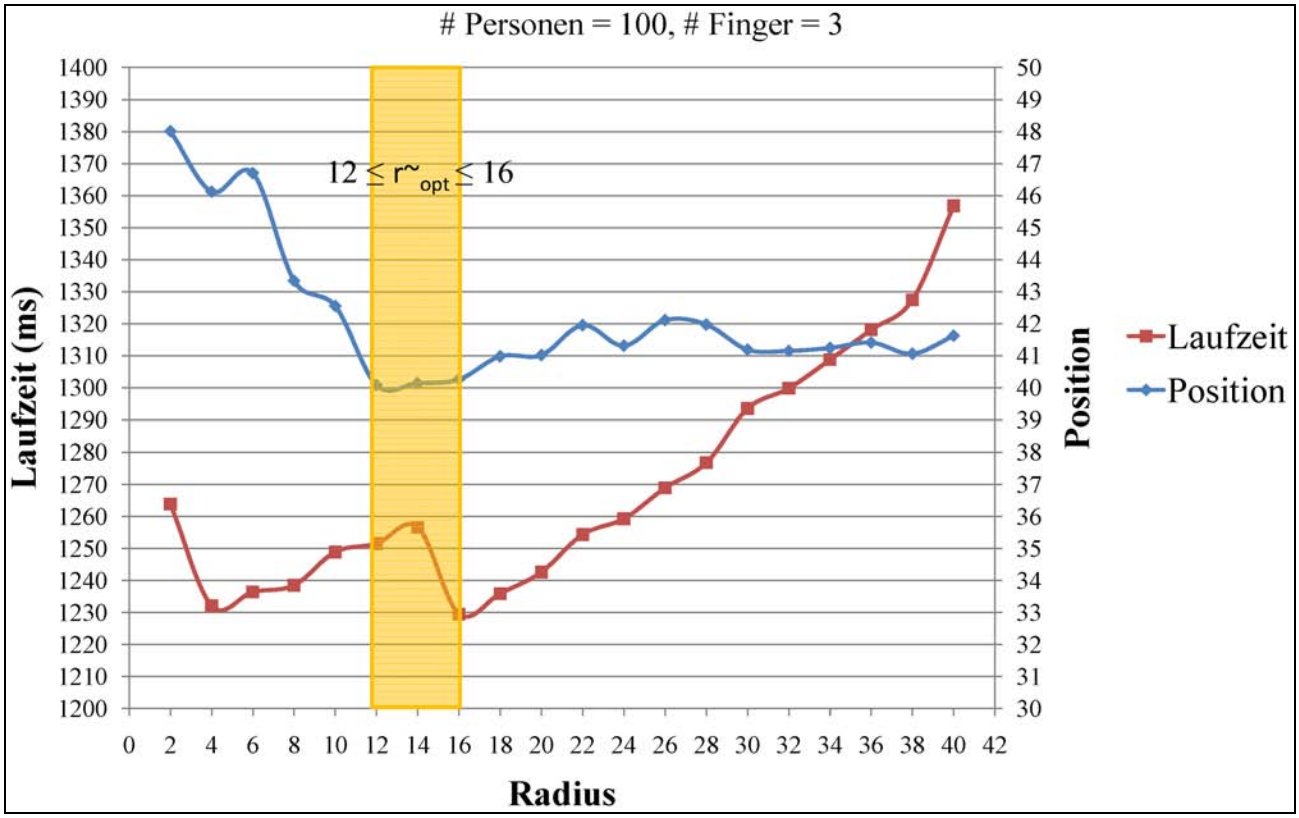

Abbildung 79: BioSimJoin: Laufzeit und Effektivität für unterschiedliche Radien

6.3.2.4 BioSimJoin*

Das Verfahren BioSimJoin* gestaltet die Bereichsanfrage des Baseline-Verfahrens BioSimJoin dahingehend effizienter, dass die X- und Y-Koordinaten aller Minutien in einem R-Baum gespeichert sind, und dieser für jeden Fingertyp initial nur ein Mal erstellt wird und nicht bei jeder Anfrage erneut aufgebaut werden muss. An dieser Stelle werden wir die Zeit untersuchen, die benötigt wird, um die Indexstruktur aufzubauen. Ziel dieser Experimente ist es, eine optimale Seitenkapazität zu ermitteln. Aufbauend auf diesen Ergebnissen, wird anschließend die Effizienz der Datenbanksuche mit Indexunterstützung untersucht und überprüft, inwiefern die Festlegung der Seitenkapazität den in Bezug auf die Laufzeit optimalen Radius der Bereichsanfrage beeinflusst. Abschließend evaluieren wir Effizienz und Effektivität von BioSimJoin* mit den zuvor ermittelten optimalen Parametern für Seitenkapazität und Radius der Bereichsanfrage.

Evaluierung des Aufbaus der Indexstruktur

Wir untersuchen nun, inwieweit sich unterschiedliche Seitenkapazitäten auf die Komplexität des Indexaufbaus auswirken. Der Radius wird dabei auf den Wert $\tilde{r}_{opt}$ =13 gesetzt, der anhand des Verfahrens BioSimJoin ermittelt wurde. Abbildung 80 stellt den Zeitaufwand, der für die Erstellung des R-Baums benötigt wird, für unterschiedliche Seitenkapazitäten dar. Sowohl Abszisse, als auch Ordinate sind dabei für eine erleichterte Interpretierbarkeit der Ergebnisse logarithmisch skaliert.

Dieses Experiment wurde zudem in Abhängigkeit der Datenbankgröße durchgeführt. Wir haben den benötigten Zeitbedarf für Datenbanken gemessen, die aus 4, 8, 16 und 32 Personen bestehen. Das entspricht einer tatsächlichen Datenbankgröße von 8.250 (33 Fingerabdrücke), 16.500 (66 Finger-

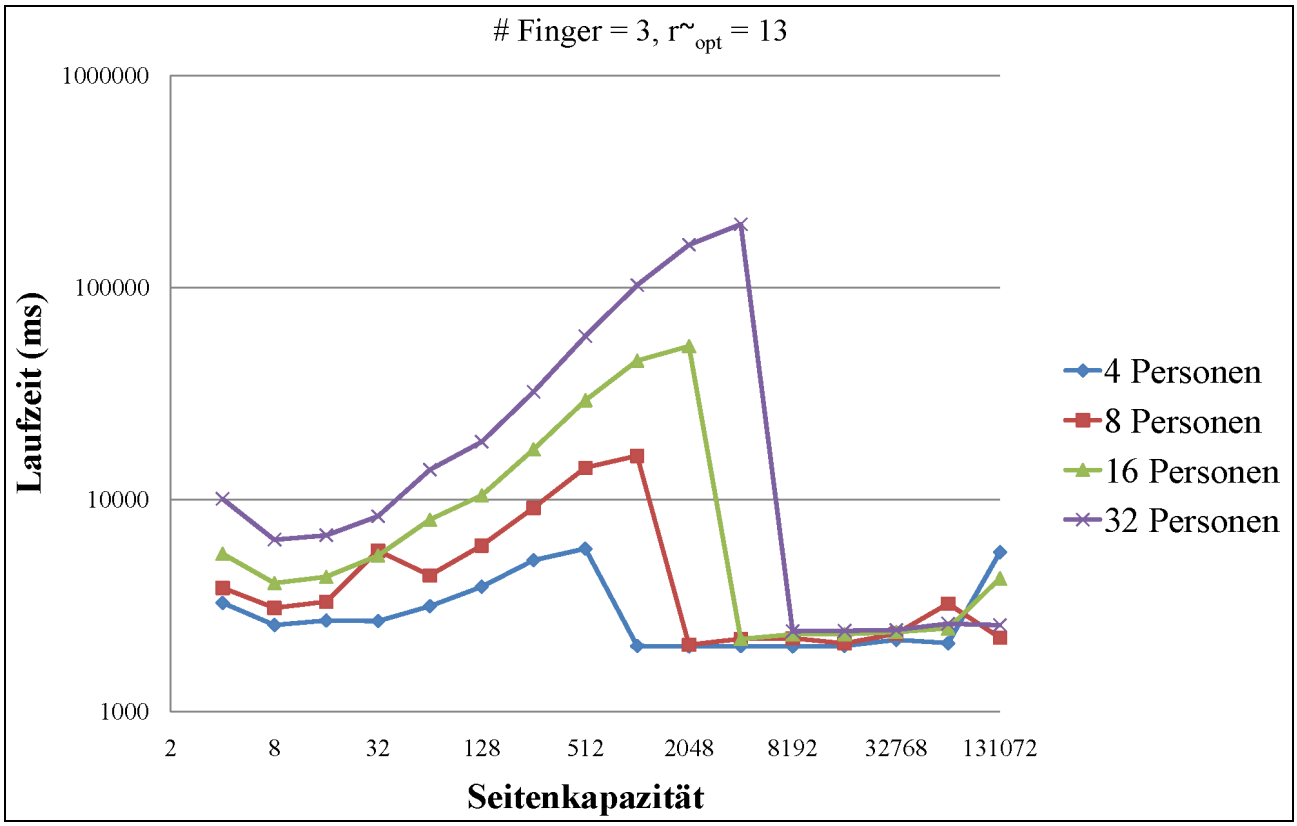

Abbildung 80: BioSimJoin: Indexaufbau für unterschiedliche Seitenkapazitäten*

abdrücke), 31.500 (126 Fingerabdrücke) und 61.250 (245 Fingerabdrücke) Einträgen für Minutien bzw. Chaff-Points.

Die Grafik zeigt, dass der R-Baum effizient aufgebaut werden kann, solang eine Seite nicht mehr als 32 Einträge speichert. Die Schwankungen in den vorderen Bereichen der Kurven sind auch hier durch Implementierungs-Overhead zu erklären. Vor allem beim Aufbau der Indexstruktur müssen zu Beginn zahlreiche Programmiersprachen-abhängige Datenstrukturen, Objekte und Referenzen erstellt und verwaltet werden. Ein ähnlicher Effekt ist am rechten Rand der Grafik zu beobachten. Ein höherer Zeitaufwand begründet sich durch Verwaltungsarbeiten für Speicherbedarf, der durch erhöhte Seitenkapazitäten zustande kommt. Im mittleren Teil der Grafik bewirkt eine steigende Seitenkapazität, dass die Initialisierung insgesamt mehr Zeit erfordert. Das liegt daran, dass beim Aufteilen einer Datenseite der Zugriff auf die Daten innerhalb einer Seite auf einem sequentiellen Scan basiert. Dieser ist umso ineffizienter, je mehr Daten in einer Seite gespeichert sind. Sobald die Kapazität groß genug ist, um alle Einträge in einer einzigen Seiten zu speichern, nimmt die Laufzeit der Indexierung rapide ab. Wenn sich allerdings alle Daten in nur einer Seite, also der Wurzel des R-Baums befinden, ist keinerlei Indexunterstützung mehr gegeben. Demzufolge ist ein positives Leistungsverhalten zu erwarten, wenn eine möglichst geringe Seitenkapazität gewählt wird, die den anfänglichen Implementierungs-Overhead kompensiert.

Ermittlung der optimalen Kapazität c_{opt}

Die vorangegangenen Tests ergaben, dass eine möglichst kleine Seitenkapazität unter Berücksichtigung anfänglicher Initialisierungsaufgaben optimal geeignet ist, um einen effizienten Indexaufbau zu garantieren. Gleiches lässt sich auch bei der Laufzeitanalyse (vgl. Abbildung 81) der eigentlichen Suche beobachten.

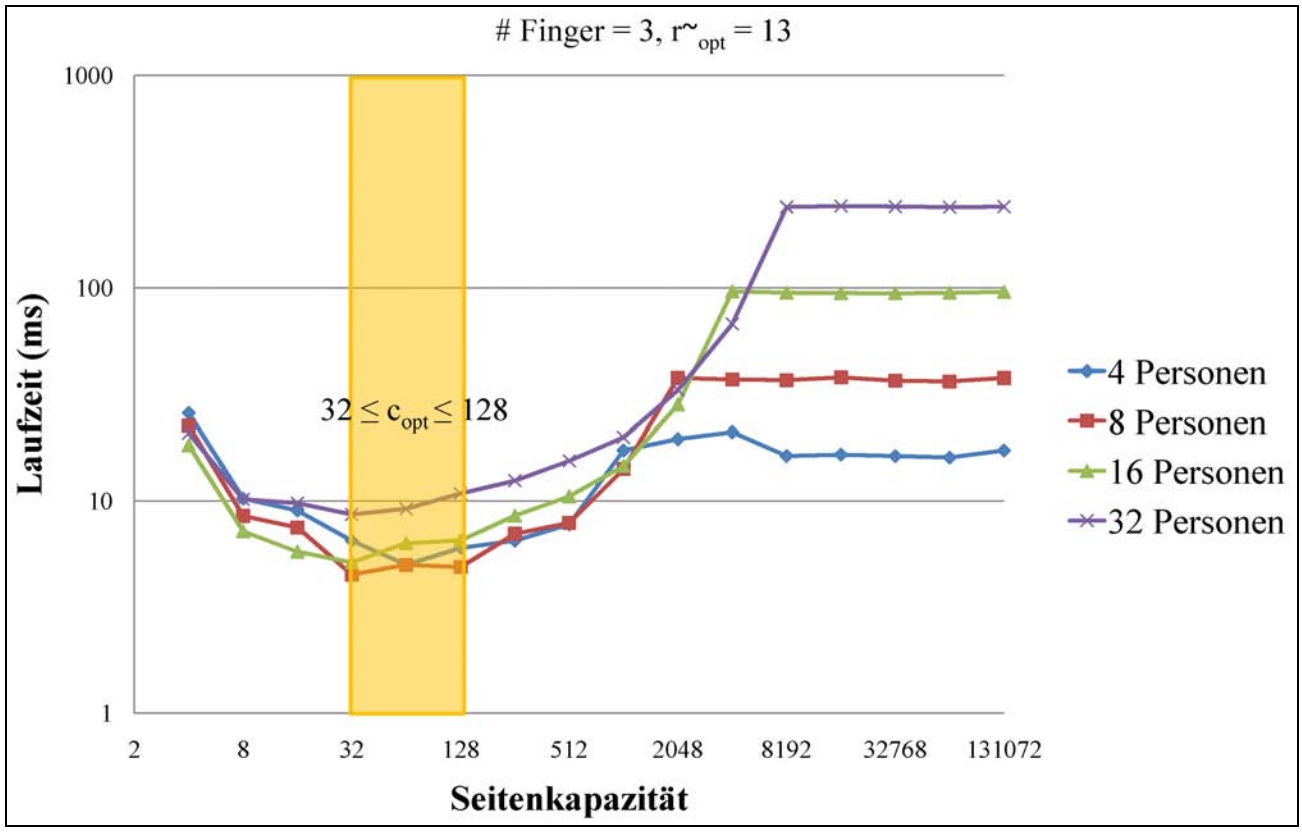

Abbildung 81: BioSimJoin: Suchzeit für unterschiedliche Seitenkapazitäten*

Eine steigende Kapazität bewirkt, dass die Seiten immer mehr Daten aufnehmen können. Dadurch verringert sich der Verzweigungsgrad und damit die Höhe des resultierenden R-Baumes. Das hat zur Folge, dass für die Bereichsanfragen innerhalb des Verfahrens BioSimJoin* häufig nur ein kleiner Teil der Daten ausgeschlossen werden kann. Da der Zugriff auf die Elemente innerhalb einer Seite sequentiell erfolgt, nimmt die Suchzeit mit steigender Seitenkapazität zu. Im Extremfall sind alle Daten in der Wurzel gespeichert. Ab diesem Punkt verlaufen die Kurven auf konstant hohem Niveau. Diese Experimente lassen vermuten, dass eine Kapazität $32 \leq c_{opt} < 128$, unabhängig von der Größe der Datenbank, sowohl für die Indexierung als auch die darauf aufbauende Suche optimal sind. Die folgenden Tests werden untersuchen, inwieweit dieses Ergebnis mit dem zweiten Parameter des Verfahrens BioSimJoin*, dem Radius der Bereichsanfrage, in Beziehung steht.

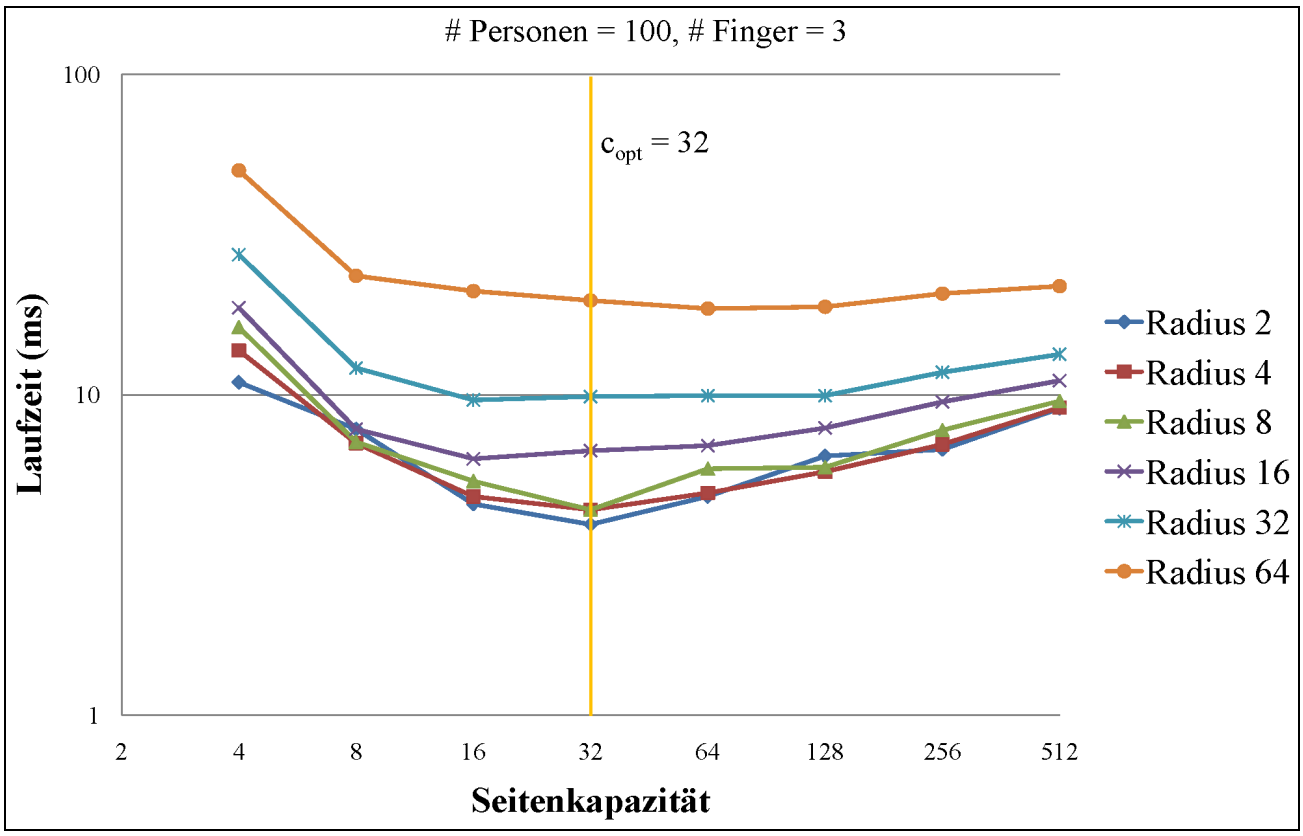

Abbildung 82: BioSimJoin: Zusammenhang zwischen Seitenkapazität und Radius*

Bis jetzt wurden alle Experimente mit dem näherungsweise optimalen Radius $\tilde{r}_{opt} = 13$, der auf Basis des Baseline-Verfahrens BioSimJoin bestimmt wurde, durchgeführt. Nun soll überprüft werden, welcher Zusammenhang zwischen diesem Parameter und der Kapazität einer Datenseite besteht. Die im Rahmen dieser Tests verwendete Datenbank umfasst 61.250 Einträge für insgesamt 245 Fingerabdrücke, die von 16 unterschiedlichen Personen stammen. Die Seitenkapazität wurde im relevanten Bereich bis zu einem maximalen Wert von 512 variiert. Abbildung 82 zeigt diese Werte auf der Abszisse logarithmisch zur Basis 2. Die Ordinate skaliert die benötigte Laufzeit ebenfalls in logarithmischer Form. Jede Kurve repräsentiert die Komplexität des Verfahrens BioSimJoin* unter Verwendung des entsprechenden Radius-Wertes und der variierenden Seitenkapazität. Dabei ist anzunehmen, dass sich eine Seitenkapazität $c_{opt} = 32$, unabhängig von der Parametrisierung des Radius für eine effiziente Suche eignet.

Im folgenden soll nun der relevante Wertebereich exakt analysiert werden. Die blaue Kurve in Abbildung 83 gibt die Laufzeit für unterschiedliche Seitenkapazitäten im relevanten Bereich [0, ..., 105] an. Die orange Linie stellt den sechsstufig gewichteten Mittelwert der Laufzeit für unterschiedliche Seitenkapazitäten im Intervall [1, …, 100] dar. Dieser Schätzer erlaubt es, die tatsächliche Tendenz innerhalb der blauen Kurve zu erkennen, indem untypische Zeitmessungen geeignet kompensiert werden. Jeder Punkt des Schätzers t_i berechnet sich durch die gewichtete Summe der sechs angrenzenden Messungen in beide Richtungen und definiert sich formal wie folgt:

$$t_i = \frac{1*t_{i-6}+2*t_{i-5}+3*t_{1-4}+4*t_{1-5}+6*t_{i-6}+7*t_i+6*t_{i+1}+5*t_{i+2}+4*t_{i+3}+3*t_{i+4}+2*t_{i+5}+1*t_{i+6}}{49}$$

Damit ist zu beobachten, dass die blaue Kurve ein klares Optimum beinhaltet. Dieser feingranulare Test zeigt, dass das Verfahren BioSimJoin* bei einer Seitenkapazität von $c_{opt} = 50$ und einem Radius $\tilde{r}_{opt} = 13$ mit höchster Effizienz ausgeführt wird. Die durchschnittliche Suche innerhalb einer Datenbank mit 182.500 Einträgen dauert in diesem Fall lediglich etwa 28ms.

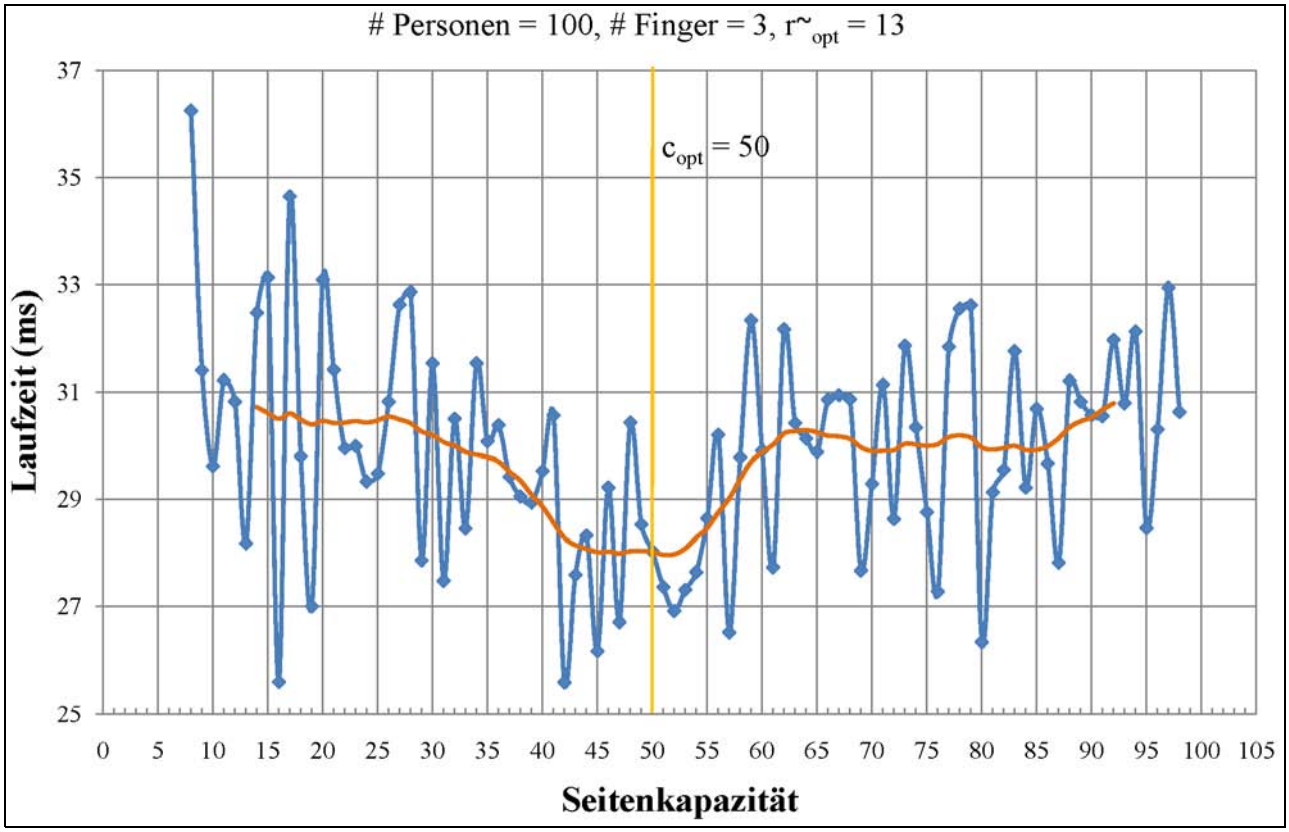

Abbildung 83: BioSimJoin: Exakte Evaluierung der Seitenkapazität*

Ermittlung des optimalen Radius r_{opt}

Nachdem nun eine optimale Parametrisierung der Seitenkapazität experimentell ermittelt wurde, soll an dieser Stelle der exakte Radius der Bereichsanfrage r_{opt} geeignet gewählt werden. Erste Experimente mittels des Verfahrens BioSimJoin, das ausschließlich die Parametrisierung des Radius erfordert, zeigten dass ein gutes Laufzeitverhalten für Radius-Werte zwischen 12 und 16

erzielt werden kann. In Anlehnung an die zuvor gewonnenen Resultate wurden die Experimente an dieser Stelle im Wertebereich [0 … 40] durchgeführt.

Wie auch schon bei den Experimenten zu dem Verfahren BioSimJoin ersichtlich war, wirken sich hohe Radius-Werte auch negativ auf die Laufzeit des indexierten Verfahrens BioSimJoin* aus. Das begründet sich daraus, dass auch in diesem Fall ein höherer Radius zu einer größeren Zahl von Kandidaten führt, die während des Ablaufs des Algorithmus betrachtet werden müssen. Dieser Effekt wird mithilfe der roten Kurve in Abbildung 84 verdeutlicht. Demzufolge wäre also ein möglichst kleiner Radius geeignet, um das Laufzeitverhalten zu optimieren.

Kleine Radius-Werte sind allerdings nicht in der Lage Rotationen oder Verschiebungen in den Daten zu kompensieren. Daher befindet sich die angefragte Person bei einem Radius von 1 durchschnittlich nur an Position 65 innerhalb einer Kandidatenliste, die insgesamt aus 100 Personen besteht. Die optimale Position der angefragten Person ist bei einem Radius von $r_{opt} = 13$ zu erzielen. Die Suche dauert in diesem Fall im Durchschnitt 45ms.

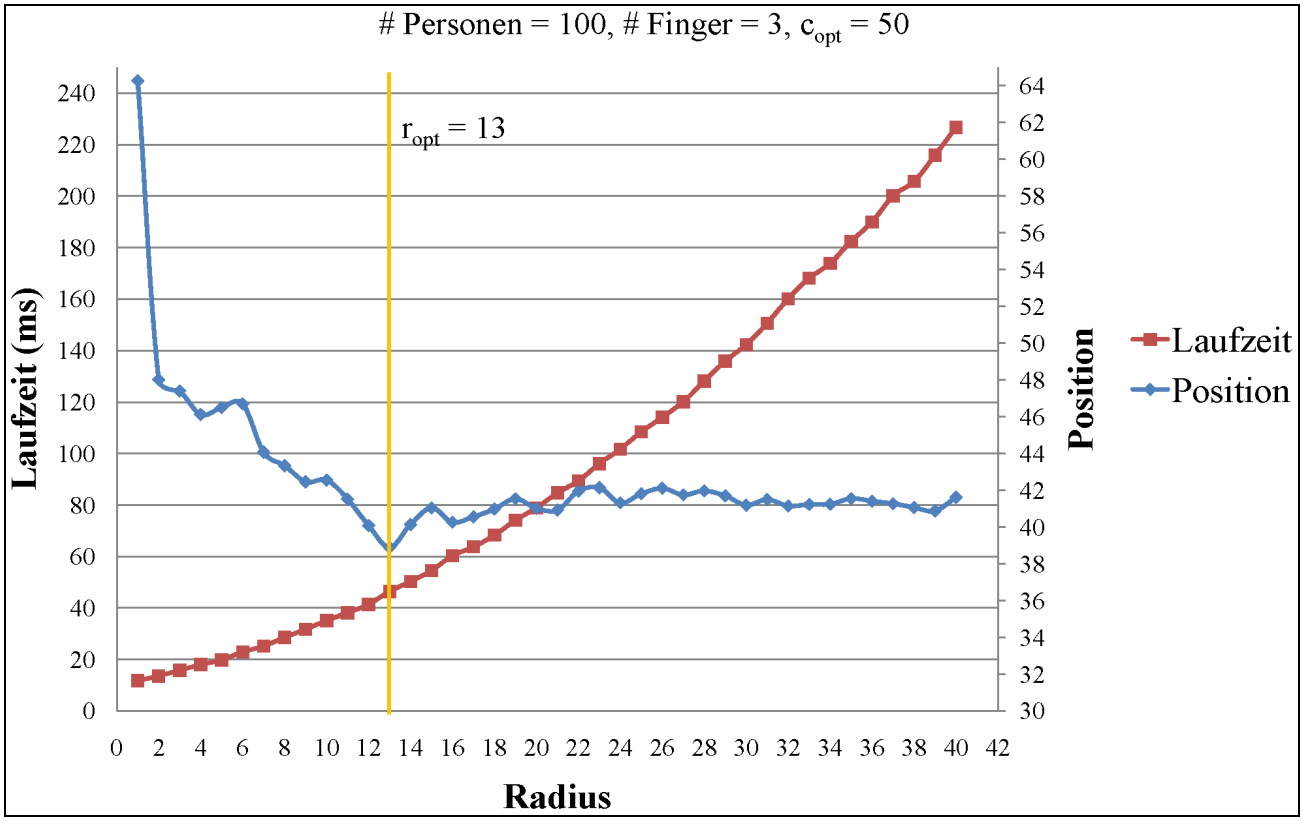

Abbildung 84: BioSimJoin: Laufzeit und Effektivität für unterschiedliche Radien*

Evaluierung der Datenbankgröße

Die bis hier durchgeführten Experimente dienten dazu, die optimale Parametrisierung der Verfahren BioSimJoin und BioSimJoin* zu ermitteln. An dieser Stelle soll nun untersucht werden, welchen Einfluss die Größe der Datenbank auf Laufzeit und Stabilität der Index-unterstützten Suche durch BioSimJoin* mit der Parametrisierung $r_{opt} = 13$ und $c_{opt} = 50$ hat. Dazu haben wir basierend auf der Datenbank SD 14 unterschiedliche Datenbestände erstellt, die in Tabelle 15 spezifiziert sind.

Personen	2	4	8	16	32	64	128	256	512	1.024	2.048	2.622
Finger-abdrücke	18	33	66	126	245	490	901	1.765	3.521	7.206	13.839	17.723
Minutien bzw. Chaff-Points	4.500	8.250	16.500	31.500	61.250	122.500	225.250	441.250	880.250	1.801.500	3.459.750	4.430.750

Tabelle 16: Spezifikation unterschiedlich großer Datensätze basierend auf der Datenbank SD 14

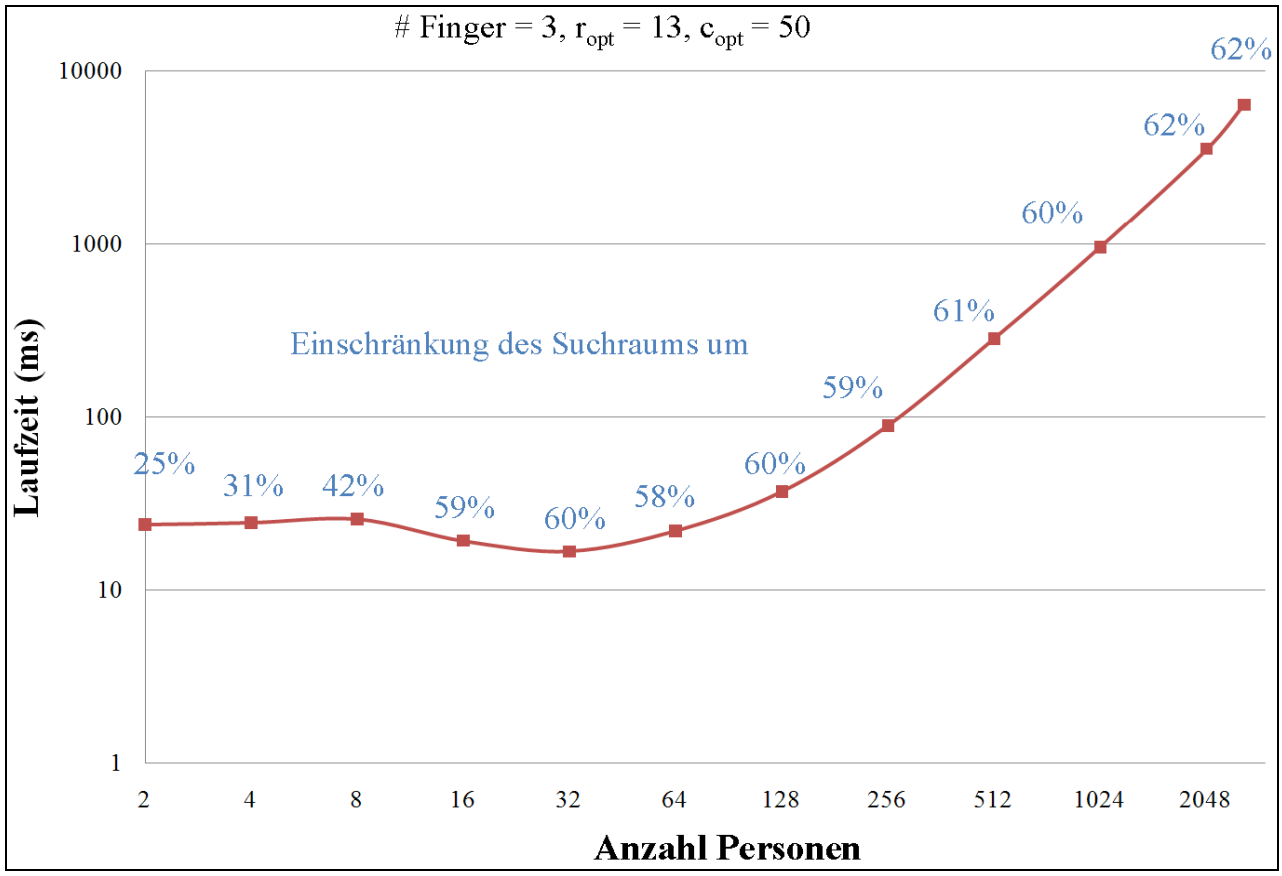

Abbildung 85: BioSimJoin: Laufzeit und Effektivität für unterschiedlich große Datenbanken*

Abbildung 85 zeigt klar, dass die durchschnittliche Laufzeit pro Suche mit wachsender Datenbankgröße tendenziell zunimmt, da in diesem Fall eine größere Zahl von Referenzdaten durchsucht werden muss. Dieser Anstieg ist vor allem bei sehr großen Datenbanken, die in der Praxis häufig zu finden sind, zu erkennen. Die Datenbankgröße und Laufzeit sind auch hier für eine erleichterte Interpretierbarkeit logarithmisiert aufgetragen. Zusätzlich zeigt die Grafik, dass der relevante Suchraum durch das Verfahren stark reduziert werden kann. Im Durchschnitt können ca. 60% der relevanten Kandidaten erfolgreich ausgeschlossen werden, wodurch eine starke Effizienzsteigerung der Identifikation erzielt wird.

Evaluierung von BioSimJoin* gegenüber gezielt manipulierten Daten

Im Folgenden wird nun die in den vorherigen Experimenten bestimmte optimale Parametrisierung r_{opt} = 13 und c_{opt}= 50 verwendet um die Robustheit von BioSimJoin* anhand synthetisch manipulierter gedrehter und verschobener Anfragefinger, sowie Anfragefinger bei denen im Vergleich zum Referenzfinger Minutien fehlen bzw. hinzukommen, zu evaluieren. In allen Abbildungen ist auf der Ordinate die tatsächliche Position der angefragten Person innerhalb der Kandidatenliste gemittelt über jeweils 200 Anfragen angegeben. Auf der Abszisse sind jeweils die unterschiedlichen Grade der jeweiligen Manipulation aufgetragen. Für die Anfragen werden für jede Person jeweils drei Finger in der in Abschnitt Evaluierung der Identifikationslösungen angegebenen Priorisierung verwendet.

Rotierte Daten

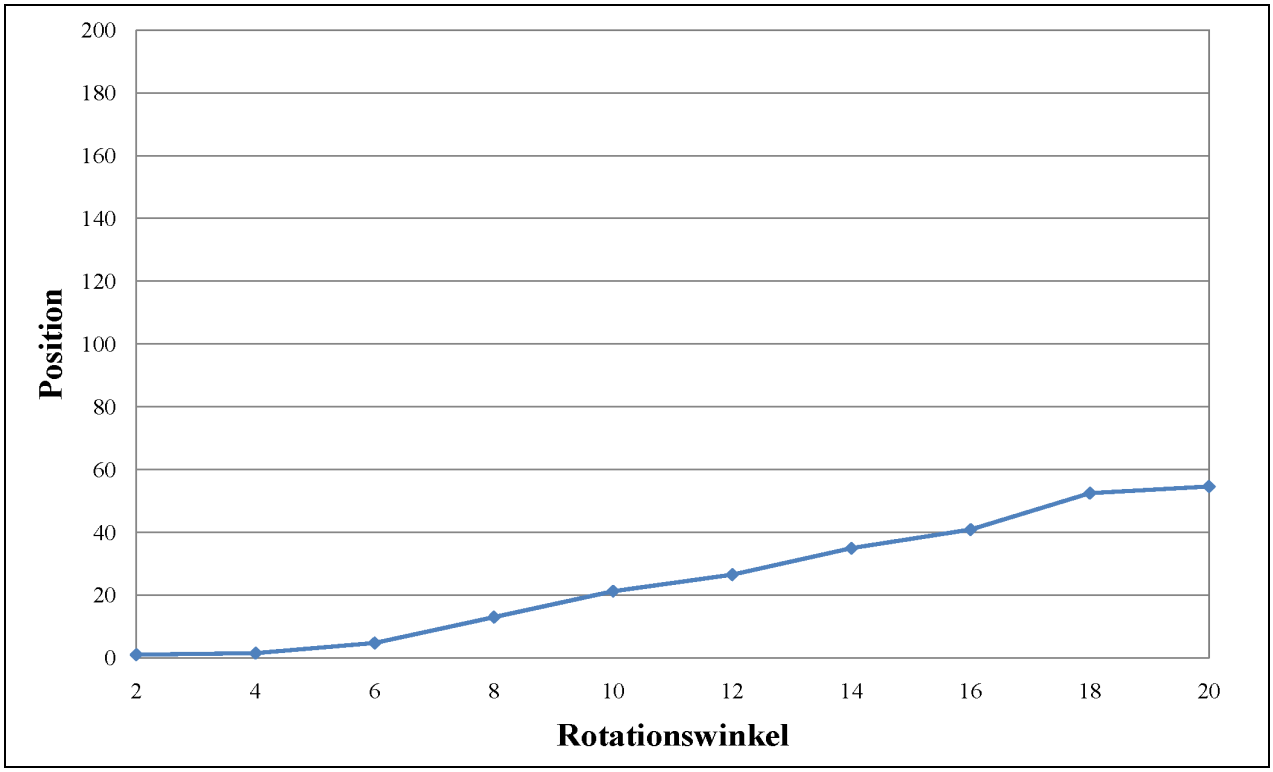

Abbildung 86: BioSimJoin: Robustheit gegenüber rotierten Daten*

Abbildung 86 zeigt, dass die Effektivität von BioSimJoin* mit zunehmender Drehung im Intervall [2,...,20] des Anfragefingers leicht abnimmt, da mit zunehmender Drehung des Anfragefingers, die Minutien des Referenzfingers nicht mehr optimal in den Radius der einzelnen Anfrageminutien fallen. Allerdings liegt der gesuchte Anfragefinger bei einer starken Drehung von 20° im Mittel in der Kandidatenliste immer noch auf einem guten Platz 54. Dies ist sehr positiv zu bewerten, da algorithmisch die Drehung bei BioSimJoin* nicht explizit berücksichtigt wird. Zudem ist zu erwähnen, dass die Fingerabdrücke in der Datenbank nicht gegeneinander vorausgerichtet sind. Eine solche Vorausrichtung der Daten könnte somit zusätzlich positiven Einfluss auf die Ergebnisse haben.

Verschobene Daten

Um zu testen, inwieweit sich eine Verschiebung der Anfragefinger im Vergleich zum Referenzfinger auf das Ergebnis von BioSimJoin* auswirkt, wird im Folgenden die Position von zunehmend stark verschobenen Anfragefinger in der Kandidatenliste untersucht.

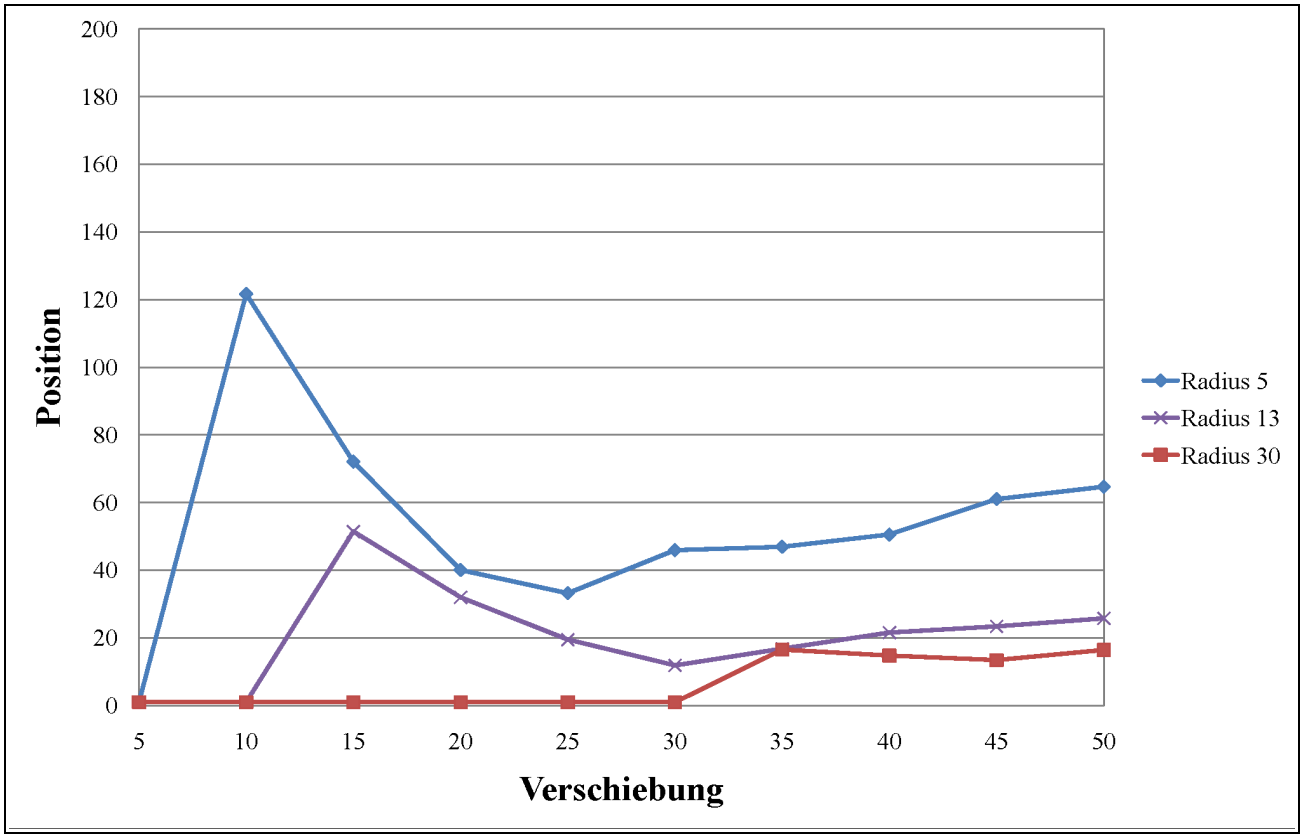

Abbildung 87: BioSimJoin: Robustheit gegenüber verschobenen Daten*

Abbildung 87 zeigt, dass bei einer Verschiebung der Anfragefinger um einen Wert kleiner dem Radius *r* der Bereichsanfrage, die gesuchten Personen innerhalb der Kandidatenliste stabil auf dem ersten Platz bleiben. Bei einer Verschiebung, die die Größe von *r* übersteigt, werden die Minutien des Referenzfingers nicht mehr gefunden. Liegt der Wert der Verschiebung zudem noch unter dem Mindestabstand von Minutien und Chaff-Points (vgl. Abschnitt Vorbereitung der Datenbasis), fallen auch keine Chaff-Points der zu identifizierenden Person in die Bereichsanfrage. Dies erklärt den sprunghaften Anstieg der Position der angefragten Person innerhalb der Kandidatenliste. Übersteigt die Verschiebung den Mindestabstand deutlich, verbessert sich das Ergebnis auf den ersten Blick wieder, da nun vermehrt auch Chaff-Points oder andere Minutien des entsprechenden Referenzfingers in die Bereichsanfrage fallen. Diese Verbesserungen sind jedoch nur zufällig und stellen keine signifikante Übereinstimmung dar.

Fehlende oder zusätzliche Minutien

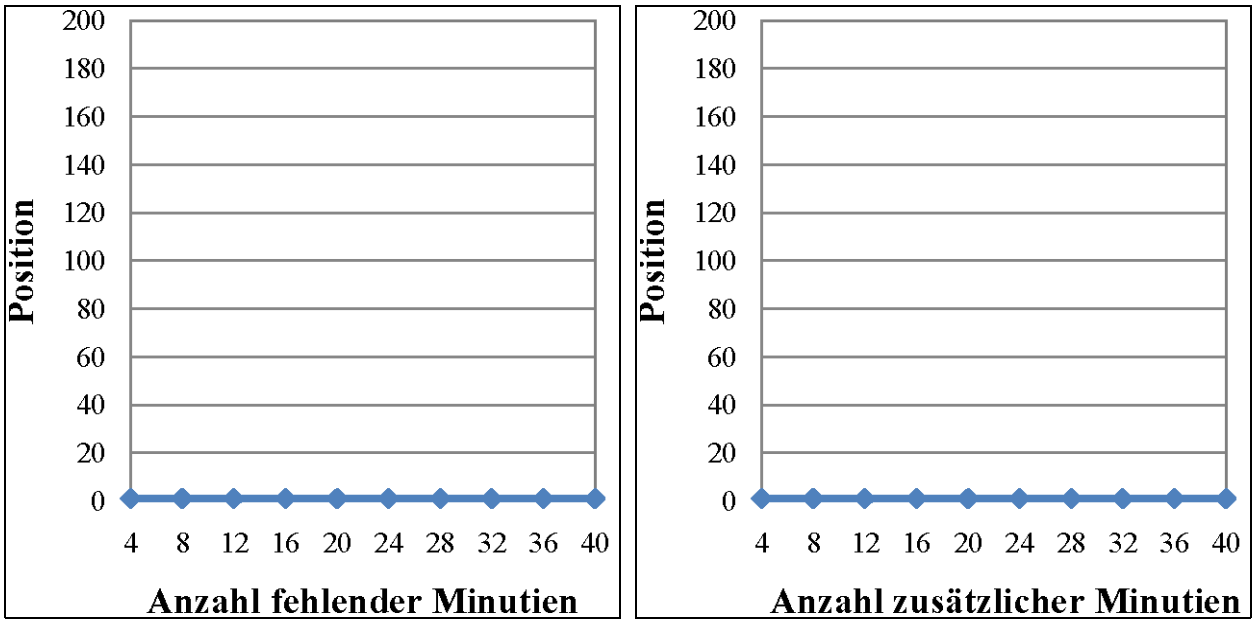

Abbildung 88: BioSimJoin: Robustheit gegenüber Daten mit fehlenden/ zusätzlichen Minutien*

Abbildung 88 zeigt, dass eine Weg- bzw. Hinzunahme von bis zu 40 Minutien bei der Anfrage keine Auswirkung auf das Ergebnis von BioSimJoin* haben. In beiden Fällen ist die Anfrageperson in der Kandidatenliste konstant auf dem ersten Platz zu finden. Sogar bei Weg- bzw. Hinzunahme von 40 Minutien, also fast 50% der Minutien erzielt BioSimJoin* ein perfektes Ergebnis. Dies ist dadurch zu erklären, dass sich bei der Wegnahme von Minutien im Anfragefinger lediglich die Anzahl zu betrachtender Anfrageradien reduziert. In jedem verbleibenden Anfrageradius wird jedoch eine passende Referenzminutie gefunden. Die Reihenfolge der Personen innerhalb der Kandidatenliste ändert sich dadurch nicht.

Auch das Hinzufügen von Minutien zum Anfragefinger hat offensichtlich keine Auswirkung auf die Effektivität von BioSimJoin*. In diesem Fall erhöht sich die Anzahl an Anfrageradien. In den hinzugekommenen Anfrageradien können sich nun Chaff-Points des Referenzabdrucks befinden, aber auch Minutien und Chaff-Points von Finger anderer Personen. Da aber in den ursprünglich vorhandenen Anfrageradien jeweils eine perfekte Übereinstimmung im Referenzfingerabdruck der Datenbank zu finden ist, ändert sich die Liste der Kandidaten nicht.

6.3.2.5 BioNN

Ermittlung der optimalen Anzahl nächster Nachbarn k_{opt} und des optimalen Toleranzparameters δ_{opt}

Ziel dieser Experimente ist festzustellen, wie sich die Parametrisierung des Verfahrens BioNN auf Effektivität und Laufzeitverhalten bei den Suchanfragen auswirkt. BioNN hat zwei Parameter, die Anzahl nächster Nachbarn k und den Toleranzparameter bezüglich Entfernungen zwischen den Minutien δ, die gegeneinander kreuzvalidiert werden müssen. Als Datenbasis für diese Experimente wird eine nach dem Zufallsprinzip erzeugte Auswahl der Datenbank SD14 herangezogen, bestehend aus 600 Personen. Die Identifikation einer Person erfolgt anhand von drei Fingern, wie in Abschnitt Evaluierung der Identifikationslösungen erläutert. Wenn für eine Person in der erzeugten Auswahl weniger als drei Finger abgelegt sind, so wurde diese Person entfernt und durch eine zufällige Person mit mindestens drei abgelegten Fingern aus der Datenbank SD14 ersetzt.

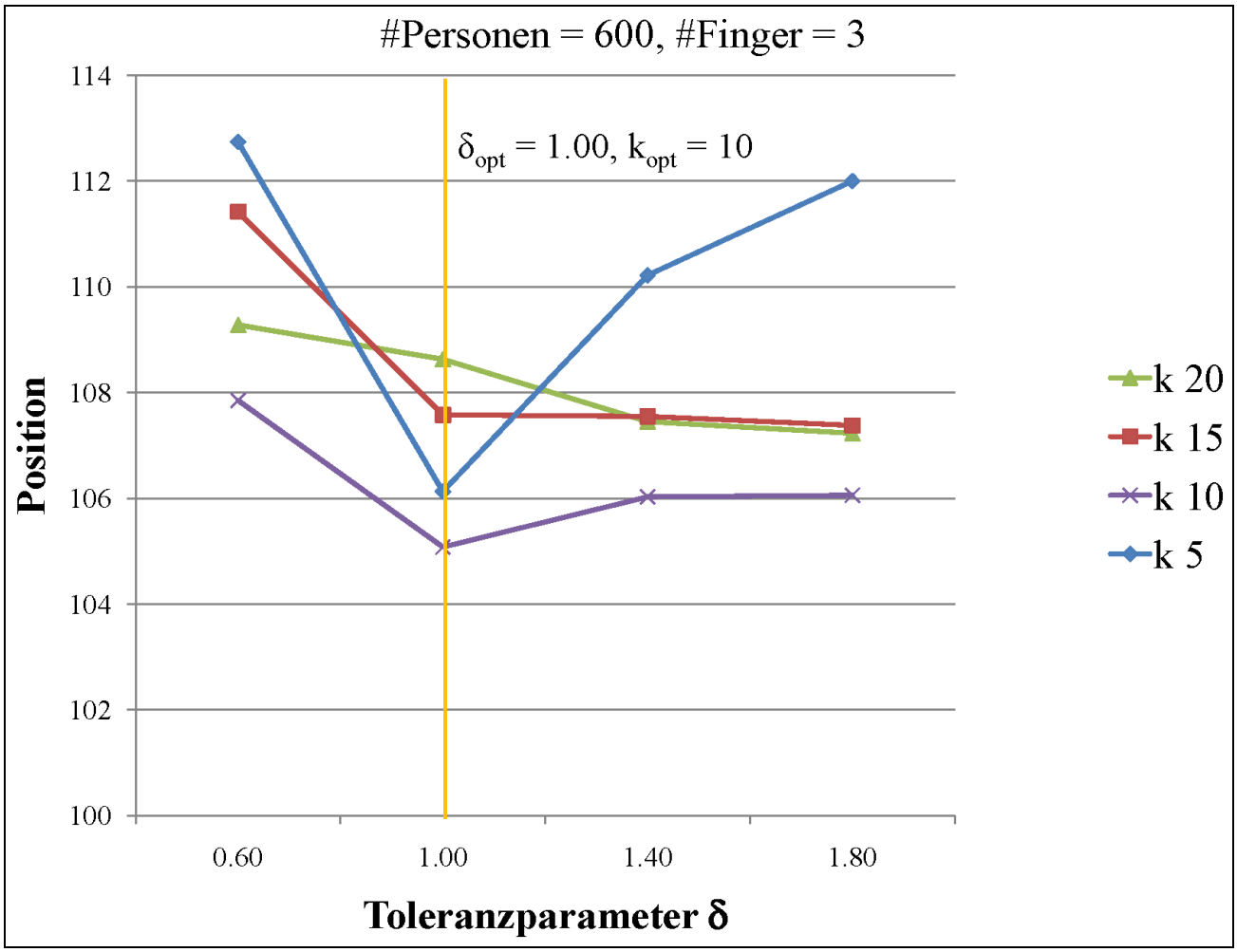

Abbildung 89: BioNN: Effektivität für unterschiedliche Parametrisierungen von k und δ

Abbildung 89 zeigt, wie sich die Wahl der Parameter δ und k auf die Effektivität des Verfahrens auswirkt. Als Maß für die Effektivität des Verfahrens dient die mittlere Position, auf der die

jeweilige Anfrageperson in der sortierten Kandidatenliste des Verfahrens bei 600 Anfragen vorkommt. Diese Maßzahl ist auf der Ordinate aufgetragen. Auf der Abszisse sind die Werte des Parameters δ im Intervall [0.60, 1.00, 1.40, 1.80] dargestellt. Unterschiedliche Werte des Parameters k im Intervall [5, 10, 15, 20] sind durch vier Kurven in unterschiedlichen Farben repräsentiert.

Je nach der Wahl der Parameter δ und k kommt die gesuchte Person im Durchschnitt auf der Stelle 105 bis 113 von 600 in der Sortierung des Verfahrens vor. BioNN reduziert also die Suchmenge für die Identifikation der Person um durchschnittlich ca. 82%.

Beste Effektivität zeigt das Verfahren bei k=10. Dies ist durch das Mischverhältnis zwischen Chaff-Points und echten Minutien in den Referenzdaten zu erklären, da unter den 10 nächsten Nachbarn einer Minutie zwei echte Minutien und acht Chaff-Points zu erwarten sind. Bei k=10 und δ=1,00 erreicht das Verfahren seinen höchsten Effektivitätswert. Daher werden diese Parameterwerte für weitere Experimente mit synthetischen Daten verwendet.

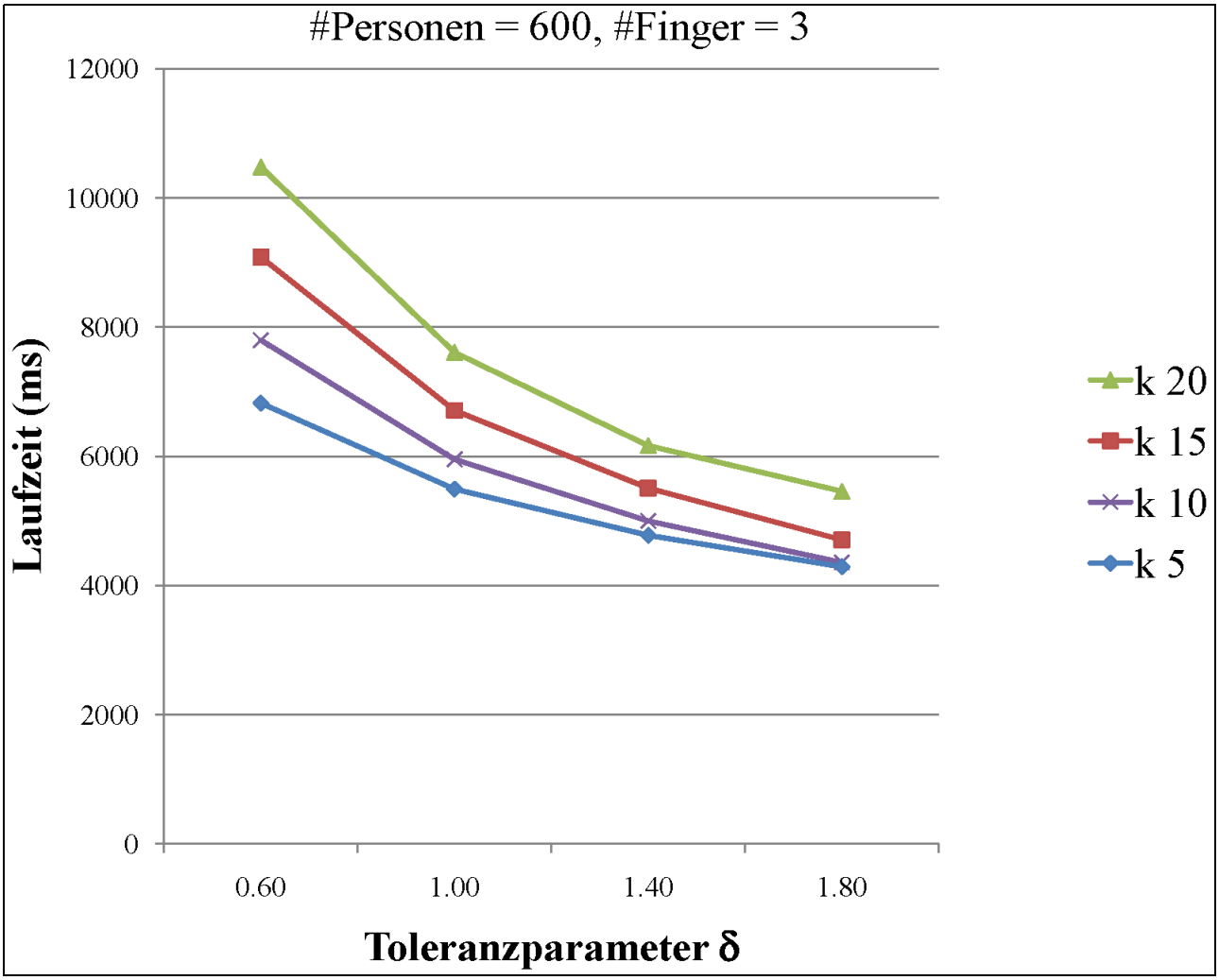

Abbildung 90: BioNN: Laufzeit für unterschiedliche Parametrisierungen von k und δ

Bei zu großen Werten des Toleranzparameters δ schätzt das Verfahren bei allen Datenbankpersonen im Schnitt mehr Minutien als ähnlich ein, was dazu führt, dass die Anzahl der Treffer bei allen Datenbankpersonen steigt und die Identifikation der Referenzperson erschwert wird. Bei zu kleinen Werten des Toleranzparameters werden weniger Minutien als ähnlich erkannt, wodurch sich die Anzahl der Treffer bei allen Personen der Referenzdaten verringert und ebenfalls die Identifikation der Referenzperson erschwert wird. Im Allgemeinen wird durch diese Ergebnisse gezeigt, dass BioNN, unabhängig von der Parametrisierung, eine hohe Effektivität aufweist.

Abbildung 90 zeigt, wie sich unterschiedliche Parameterwerte von δ und k auf die durchschnittliche Laufzeit einer Suchanfrage auswirken. Die Laufzeit in Millisekunden (ms) ist auf der Ordinate aufgetragen. Auf der Abszisse sind die Werte des Parameters δ im Intervall [0,60, 1,00, 1,40, 1,80] aufgetragen. Unterschiedliche Werte des Parameters k im Intervall [5, 10, 15, 20] sind durch vier Kurven in unterschiedlichen Farben repräsentiert. Die durchschnittliche Laufzeit hängt stark von k und δ ab und liegt bei 600 Personen zwischen etwa 4.200 ms und 10.500 ms. Die Laufzeit ist steigend in k und fallend in δ.

Evaluierung von BioNN gegenüber gezielt manipulierten Daten

Diese Experimente dienen dazu, die Robustheit von BioNN gegenüber gedrehten und verschobenen Daten, sowie Daten mit fehlenden oder hinzukommenden Minutien zu untersuchen. Als Datenbasis für diese Experimente dienen die 200 Personen aus der Datenbank SD14, die bereits für die Tests des Verfahrens BioSimJoin* verwendet wurden. Als Anfragepersonen wird die jeweilige Manipulation der Datenbasis verwendet. Es laufen hiermit pro Experiment 200 Anfragen auf einer Datengrundlage von 200 Datenbankpersonen ab. Die Parameter δ und k werden auf die in den Experimenten mit den Realdaten ermittelten Optimalwerte 1,0 und 10 gesetzt. In jedem Experiment wird die durchschnittliche Position der Anfrageperson in der Kandidatenliste angegeben.

Rotierte Daten

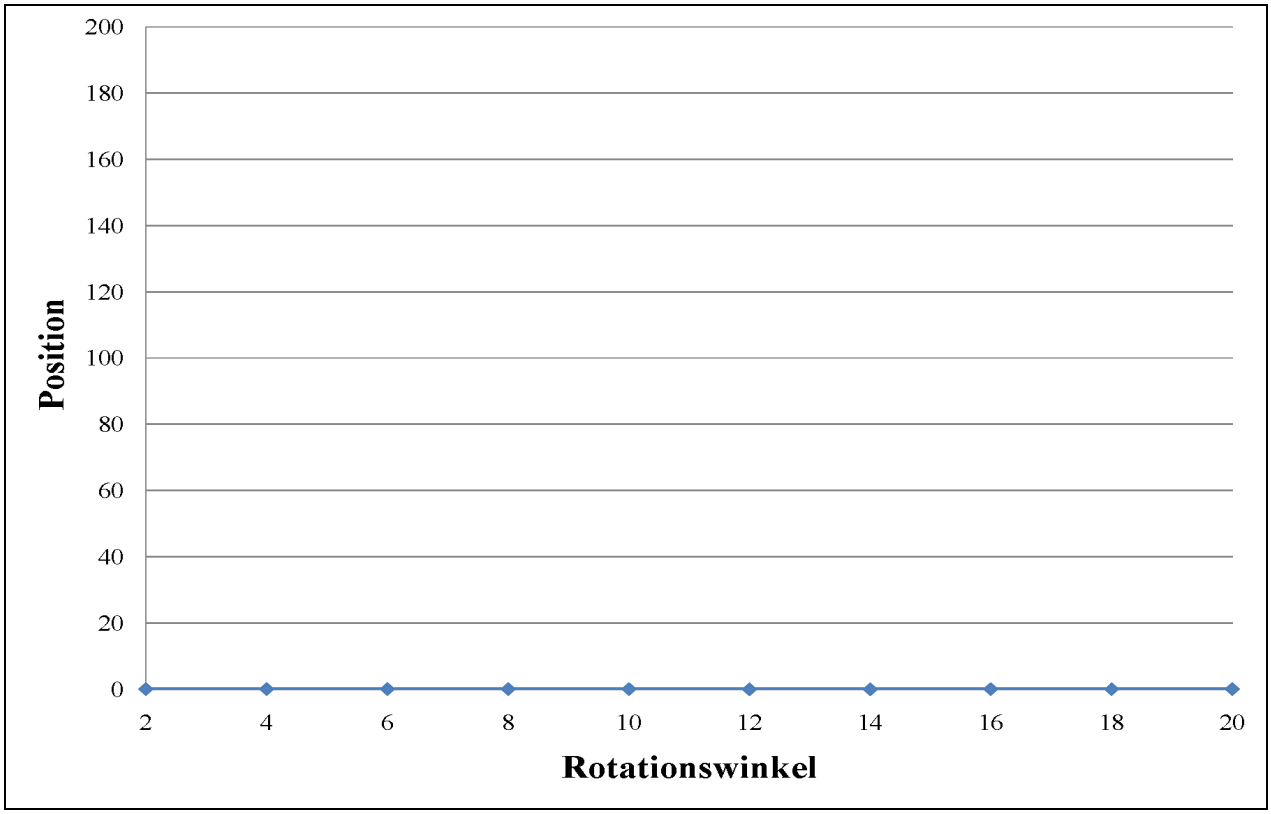

Abbildung 91: BioNN: Robustheit gegenüber rotierten Daten

In Abbildung 91 wird die Position der Anfrageperson für verschiedene Rotationswinkel im Intervall [2,...,20] dargestellt. Die gesuchte Datenbankperson kommt dabei immer auf der ersten Stelle in der Kandidatenliste des Algorithmus vor.

Verschobene Daten

In Abbildung 92 wird die mittlere Position der Anfrageperson bei Verschiebung der Minutien der Daten als Manipulation der Daten dargestellt. Auf der Abszisse ist die Verschiebung jeder Anfrageperson gemessen in Pixel gegenüber der Referenzperson in der Datenbank eingetragen. Die gesuchte Datenbankperson kommt dabei immer auf der ersten Stelle in der Ergebnissortierung des Algorithmus vor.

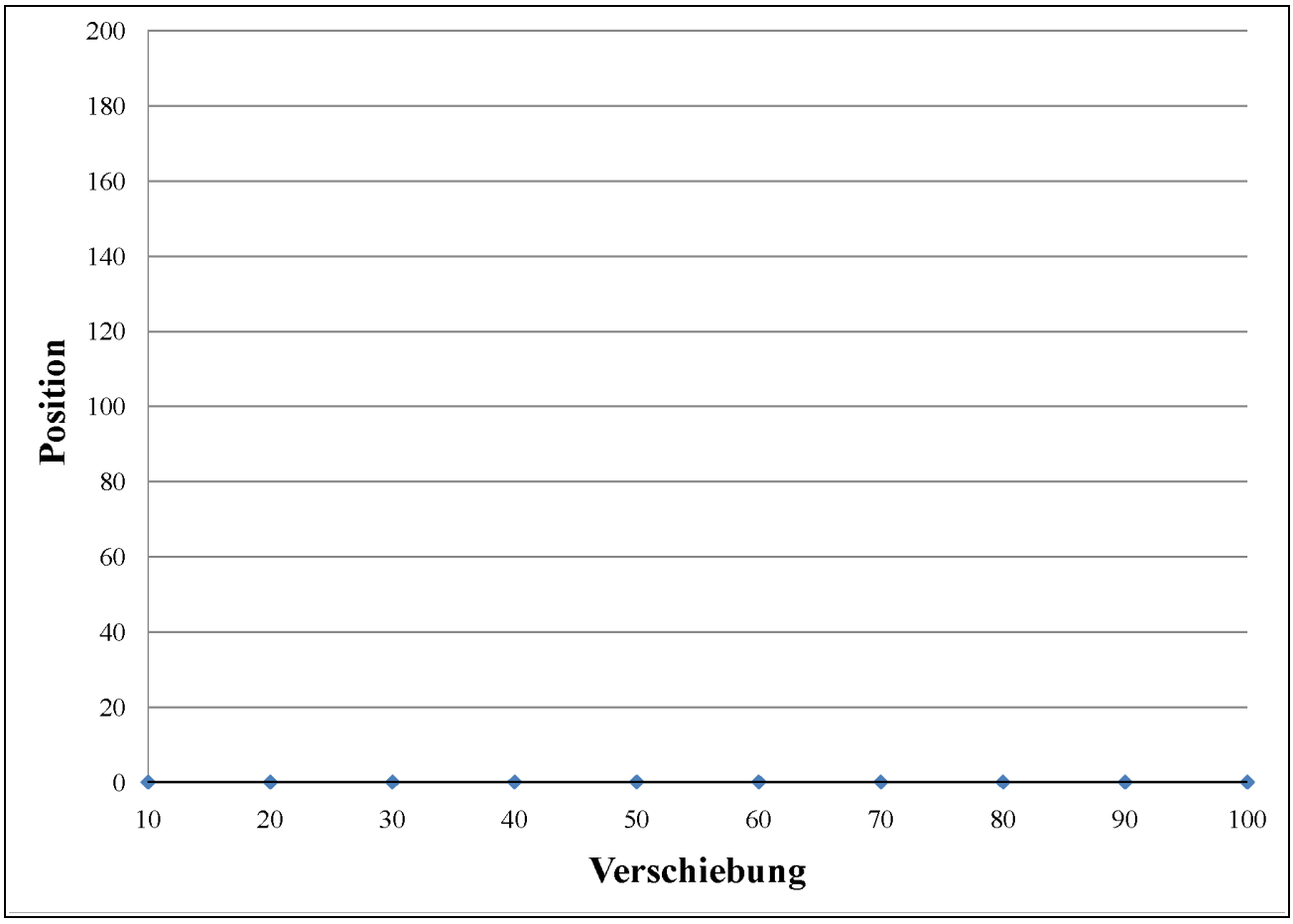

Abbildung 92: BioNN: Robustheit gegenüber verschobenen Daten

Fehlende und zusätzliche Minutien

In Abbildung 93 wird die mittlere Position der Anfrageperson beim Fehlen einiger Minutien im Vergleich zur Referenzdatenbank dargestellt. Auf der Abszisse ist die Anzahl bei jeder angefragten Person fehlender Minutien im Vergleich zur Referenzperson in der Datenbank angetragen. Die gesuchte Datenbankperson kommt dabei bei bis zu 50% fehlenden Minutien gegenüber der Referenzperson beinahe immer auf der ersten Stelle in der Ergebnissortierung des Algorithmus vor.

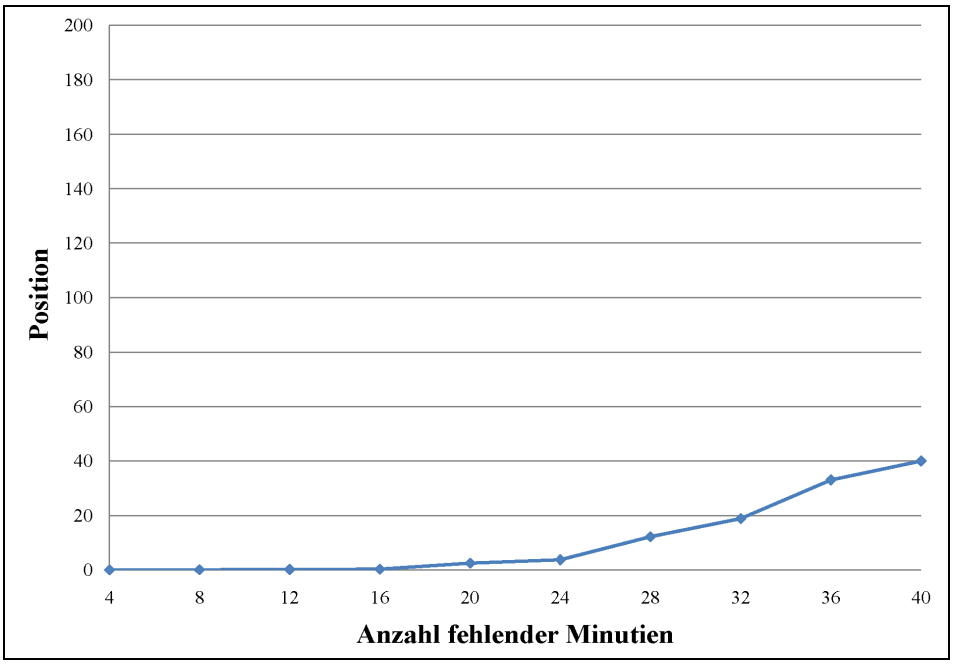

Abbildung 93: BioNN: Robustheit gegenüber Fehlen der Minutien

In Abbildung 94 wird die mittlere Position der Anfrageperson beim Einfügen zusätzlicher Minutien im Vergleich zur Referenzdatenbank dargestellt. Auf der Abszisse ist die Anzahl zusätzlich bei jeder Anfrageperson vorhandener Minutien im Vergleich zur Referenzperson in der Datenbank eingetragen. Die gesuchte Datenbankperson kommt dabei immer auf der ersten Stelle innerhalb des sortierten Rankings vor.

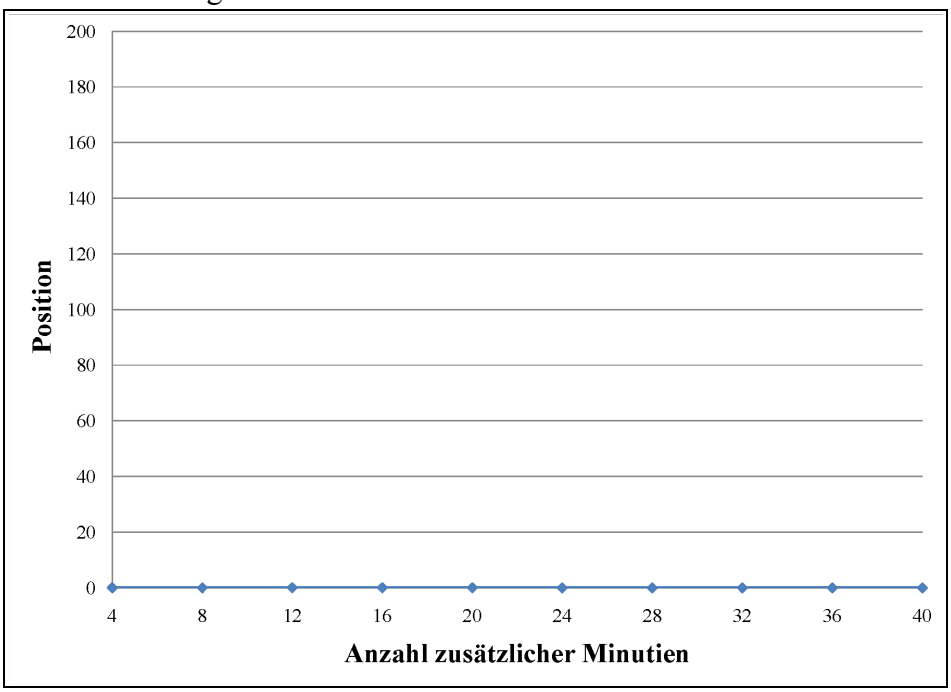

Abbildung 94: BioNN: Robustheit gegenüber Einfügen zusätzlicher Minutien

6.4 Sicherheit

Die Suchverfahren, die in den letzten Abschnitten vorgestellt wurden, verwenden nur die Informationen, die in den Vault Sets aller Benutzer in der Datenbank gespeichert sind. Während der Suchprozesse werden die Streupunkte und Stützpunkte gleichartig betrachtet. Aus Sicherheitsgründen müssen alle Punkte die gleichen statistischen Eigenschaften vorweisen. Sonst haben keine anderen zusätzlichen Informationen auf den Suchprozess Einfluss.

Im Vergleich zum Verifikationsszenario sind mehrere Vault Sets öffentlich zugängig. Da die Vault Sets unterschiedlicher Benutzer unabhängig erzeugt werden und die geheimen Zeichenketten zufällig generiert werden, ist es unmöglich, Informationen über den pseudonymen Identifikator oder die biometrischen Daten eines Benutzers anhand der geschützten Templates anderer Benutzer zu erraten. Die Sicherheit des Fuzzy-Vault-Verfahrens ändert sich hinsichtlich der Komplexität der Polynomrekonstruktion nicht. Die Sicherheit des Fuzzy-Vault-Verfahrens während der Identifikation bezüglich Brute-Force-Angriffen, Verknüpfungsangriffen und Hill-Climbing-Angriffen bleibt gleich wie bei der Verifikation. Die Sicherheitsevaluierung für diese Angriff wurde in 3.3.1.1, 3.3.1.2, 3.3.1.3 beschrieben. Im Vergleich zu den Verifikationsverfahren wird die FAR_{id} im Identifikationsverfahren aufgrund der höheren Anzahl der Vergleiche in jedem Durchlauf erhöht. $FAR_{id} = 1-(1-FAR_{veri})^n$, wobei *n* die Anzahl der Datensätze in der Datenbank ist.

Zusätzlich basieren die oben erklärten Suchverfahren auf der geometrischen Relation zwischen den Minutien-Punkten. Es gibt keine zusätzliche Offenlegung der Fingerabdruck-Informationen.

6.5 Statistische Eigenschaften der Minutien

Im Fuzzy-Vault-Verfahren werden Minutien-Informationen in zahlreichen Streupunkten, den sog. Chaffpoints, versteckt. Die Erkennungsrate des Verfahrens ist von der Zuverlässigkeit des Vergleichs zwischen Streupunkten und Minutien-Information abhängig. In einem Identifikationsszenario stehen Vault Sets unterschiedlicher Benutzter zur Verfügung. Statistische Eigenschaften biometrischer Merkmale spielen eine wichtige Rolle.

Minutien und Core-Punkte sind die wichtigsten charakteristischen Informationen eines Fingerabdrucks. Wie im internationalen Standard „Finger Minutiae Format for Data Interchange“ beschrieben, besteht die Minutien-Information aus der Position (x- und y- Koordinaten), Typ („ridge ending“, „ridge bifurcation“ or „split point“) und dem Winkel θ. Abbildung 95 zeigt die Minutien-Information (x,y,θ) eines Fingerabdrucks.

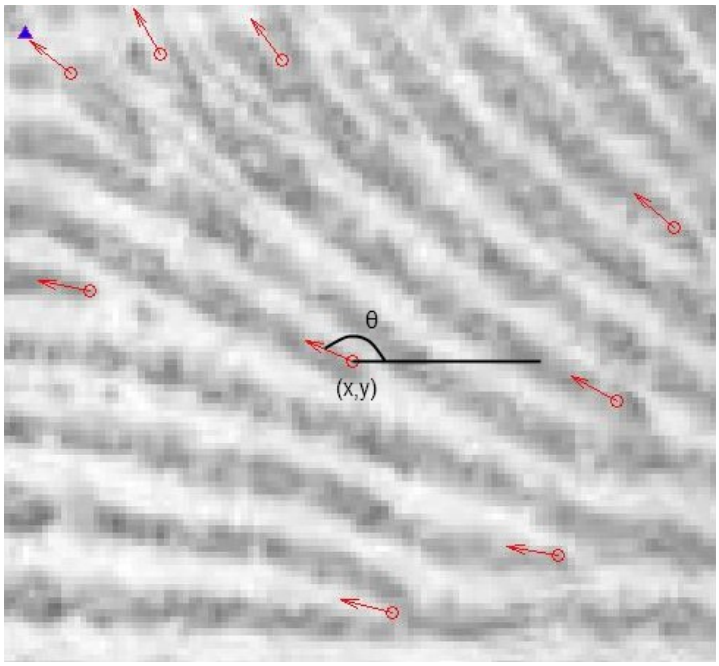

Abbildung 95: Minutien-Information (x, y, θ) eines Fingerabdrucks

Im Folgenden werden die statistischen Eigenschaften der Minutien-Merkmale der Datenbank NIST SD14 [NIST-SD14] analysiert. Die Minutien wurden mit einer Software der Firma Neurotechnology extrahiert (siehe Kapitel 5). In Abbildung 96 werden die extrahierten Minutien sowie der Core-Punkt (grünes Rechteck) auf dem Fingerabdruckbild dargestellt. Die rote Kreise markieren die Position der Minutien und deren entsprechende Richtung werden mit einem roten Pfeil gekennzeichnet. Die blau-gestrichelt markierten Bereiche zeigen die falsch detektierten Minutien; der orange-markierte Kasten zeigt das zweite Fingerglied.

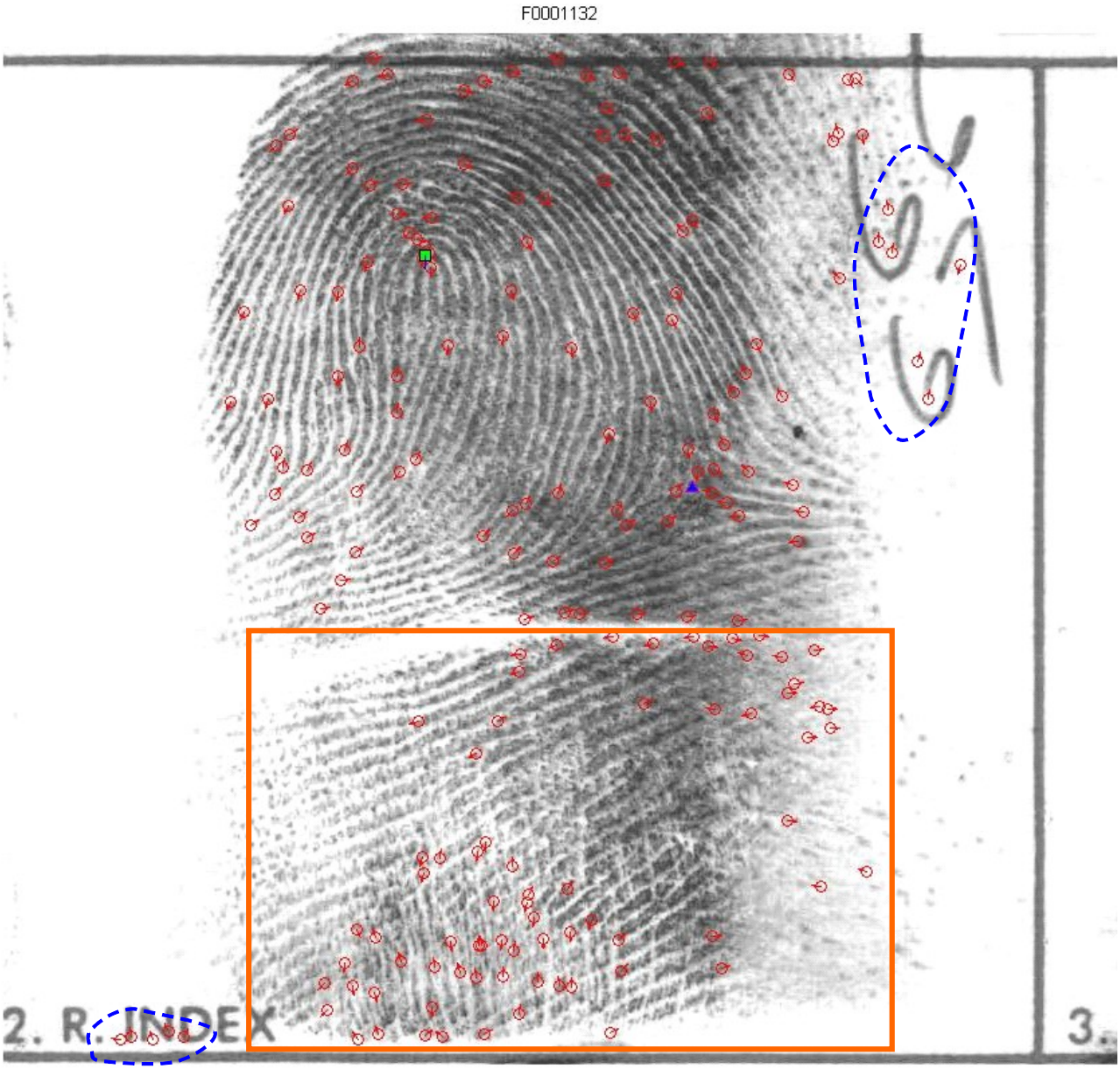

Abbildung 96: Ein Fingerabdruckbild aus der SD14

Die Häufigkeitsverteilung der Minutien für die einzelnen Pixelpositionen werden in Abbildung 97 gezeigt. Die Minutien aller rechten Zeigefinger (Enrolment-Bilder SD14 2 F, Verifikationsbilder SD14 2 S) und linken Zeigefinger (Enrolment-Bilder SD14 7 F, Verifikationsbilder SD14 7 S) ohne Berücksichtigung der Ausrichtung wurden gezählt. Die Intensität zeigt die Anzahl der Minutien an jeder Pixelposition an. Eine helle Farbe deutet auf eine höhere Anzahl von Minutien, eine dunkle Farbe auf eine geringe Anzahl hin. Die meisten Minutien befinden sich in der Mitte der Bilder. Außerdem werden auch Minutien außerhalb des eigentlichen Fingerabdruckbereichs detektiert, da die Bilder der verwendeten Datenbank nur eine geringe Qualität aufweisen. Darüber hinaus sind an den Rändern Linien zu erkennen, da Verunreinigungen der Bilder oder handschriftliche Notizen zu

Fehldeutungen führen (siehe blau markierte Bereich in Abbildung 96). So erklärt sich die höhere Dichte an Minutien in den Ecken rechts oben bzw. links unten.

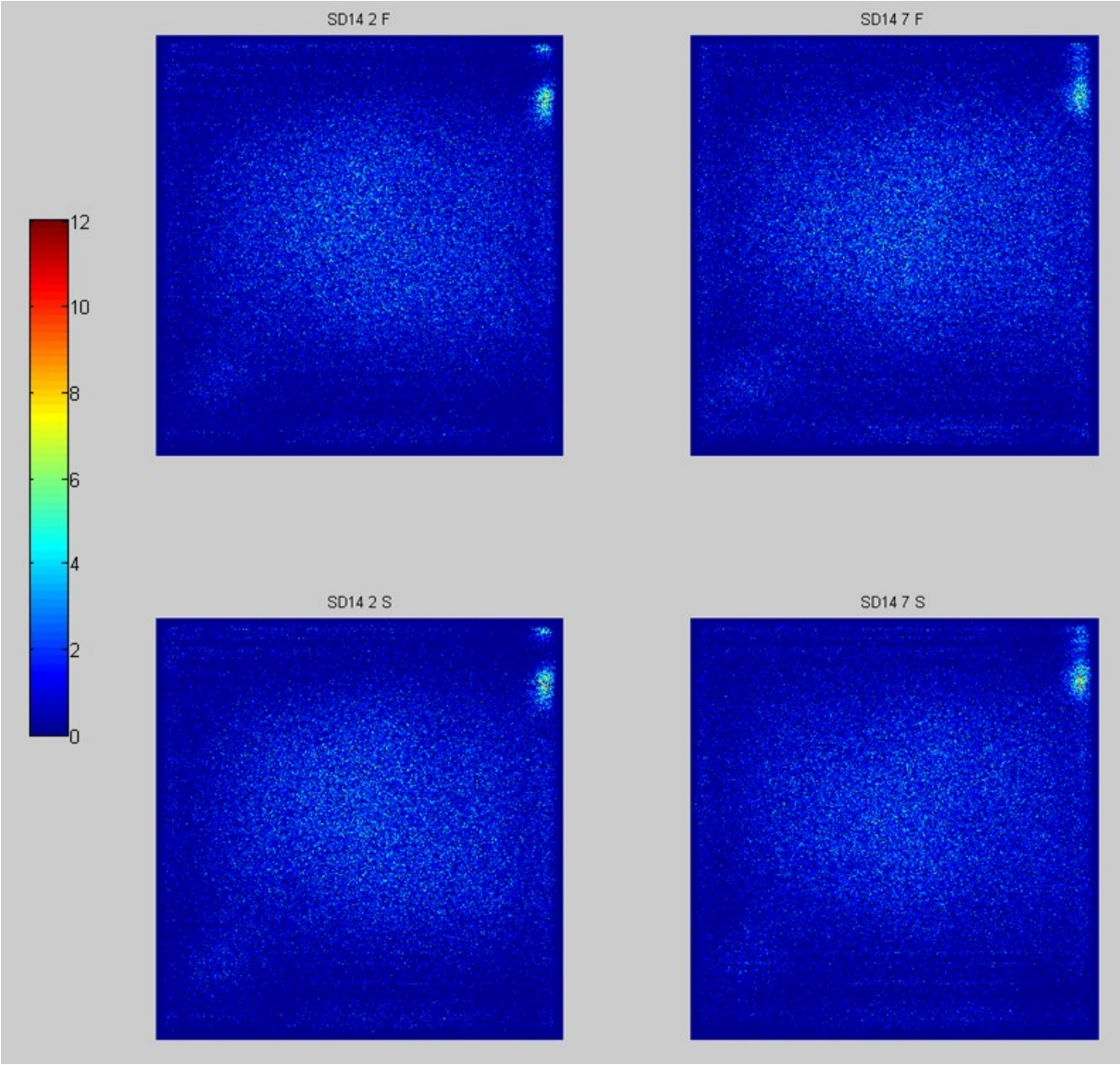

Abbildung 97: Minutien-Verteilung auf den Fingerabdruckbildern

Um nun die statistischen Eigenschaften der Minutien genau zu analysieren, werden die Minutien an deren Core-Punkten ausgerichtet. In den Datenbanken sind die Fingerabdrücke ungefähr vertikal auf den Bildern abgebildet. Der Ursprung des Koordinatensystems wird auf die Core-Punkte gelegt. Für Fingerabdrücke, die mehrere Core-Punkte aufweisen, wird der oberste (im Koordinaten-System) Core-Punkt als Referenzpunkt benutzt. Die Positionen der Minutien werden in Polarkoordinaten konvertiert, um die Eigenschaften der Minutieninformation bezüglich der Core-Punkte zu analysieren. Wenn eine Minutie M die Eigenschaften (x,y,θ) hat, wobei x und y die x- und y-Koordinate der Minutie und θ der Ausrichtungswinkel der Minutie ist, sind (ρ, Φ, θ) die neuen Koordinaten von M, wobei $x = \rho \cdot \cos(\Phi)$ und $y = \rho \cdot \sin(\Phi)$. Φ ist die Winkelkoordinate zwischen $-\pi$ und $+\pi$. Da es nur 2700 Finger pro Fingertyp in der SD14 gibt, ist dies nicht ausreichend, um die Verteilung der Minutien zu analysieren. Stattdessen werden Häufigkeitsverteilungen berechnet. Die Radialkoordinate ρ der Minutien wird in 200 gleichmäßig

verteilte Intervalle im Bereich [0, 1000] unterteilt. Die Winkel Φ der Minutien-Positionen werden in 200er Intervalle im Bereich [-π, +π] aufgeteilt. Dann wird die Anzahl der Minutien in jedem Intervall gezählt. Das Ergebnis ist in Abbildung 98 darstellt.

Alle vier Bilder in Abbildung 98 haben ähnliche Eigenschaften. Die meisten Minutien haben einen Abstand von 50 bis 200 zu den Core-Punkten. Es ist auffällig, dass viele Minutien im Bereich $\Phi \in$ [-2.4, -0.3] liegen. Das liegt daran, dass ein Teil des zweiten Fingerglieds aufgenommen wurde (siehe orange markierter Kasten in Abbildung 96).

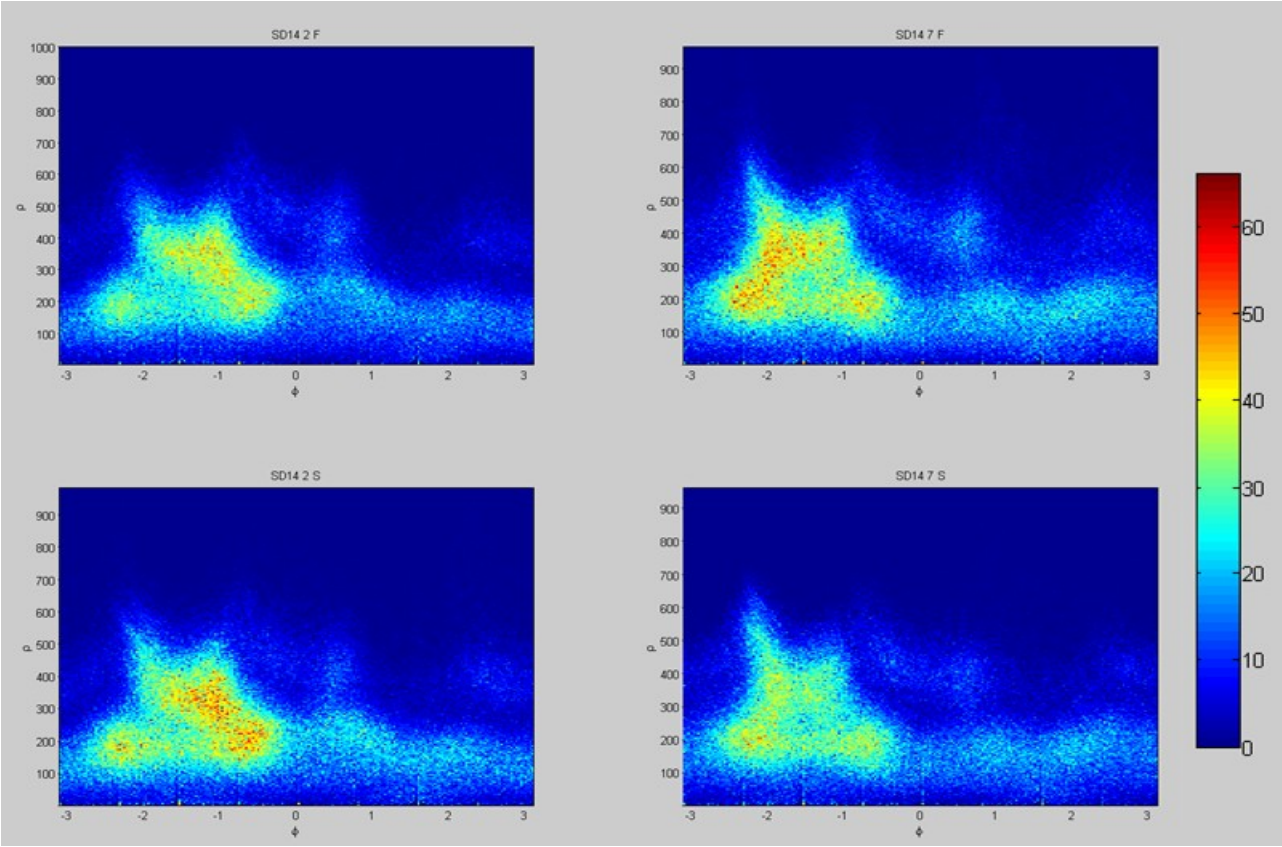

Abbildung 98: Häufigkeitsverteilungen der ausgerichteten Minutien in Bezug auf Abstand ρ und Winkel Φ relativ zum Ursprung (dem Core-Punkt, wo ρ=0, Φ=0)

In ähnlicher Weise wird die Häufigkeitsverteilung bezüglich des Winkels der Minutien-Position Φ und der Minutien-Ausrichtung θ untersucht. Die Winkelkoordinate Φ wird in 500 Intervalle im Bereich [-π, +π] unterteilt. Die Ausrichtungswinkel der Minutien θ werden in 100 Intervalle im Bereich [0, +2π] aufgeteilt. Die Abbildung 99 zeigt die entsprechenden Häufigkeitsverteilungen. Auf allen vier Bildern in Abbildung 99 gibt es ähnliche Muster zu erkennen. Es ist eine deutliche Abhängigkeit zwischen Φ und θ zu erkennen. Beispielsweise auf das Bild rechts-oben in Abbildung 99, in dem Bereich, indem $\Phi \in$ [-1.5,-1] ist, ist θ mit hoher Wahrscheinlichkeit in den Bereichen von [2.5, 3.2] oder [5.5, 2π]. Es ist auffällig, dass sich viele Minutien im Bereich von $\Phi \in$ [-2.4, -0.3] befinden. Dieser Effekt ist dadurch zu erklären, dass vielfach ein Teil des zweiten Fingerglieds mit abgebildet ist (siehe orange markierter Kasten in Abbildung 96) und somit die Abstände zum

Core-Punkt größer sind. Wenn die Minutien-Winkel-Information auch in der Fuzzy-Vault-Methode genutzt wird, dann ist die statistische Abhängigkeit zu berücksichtigen.

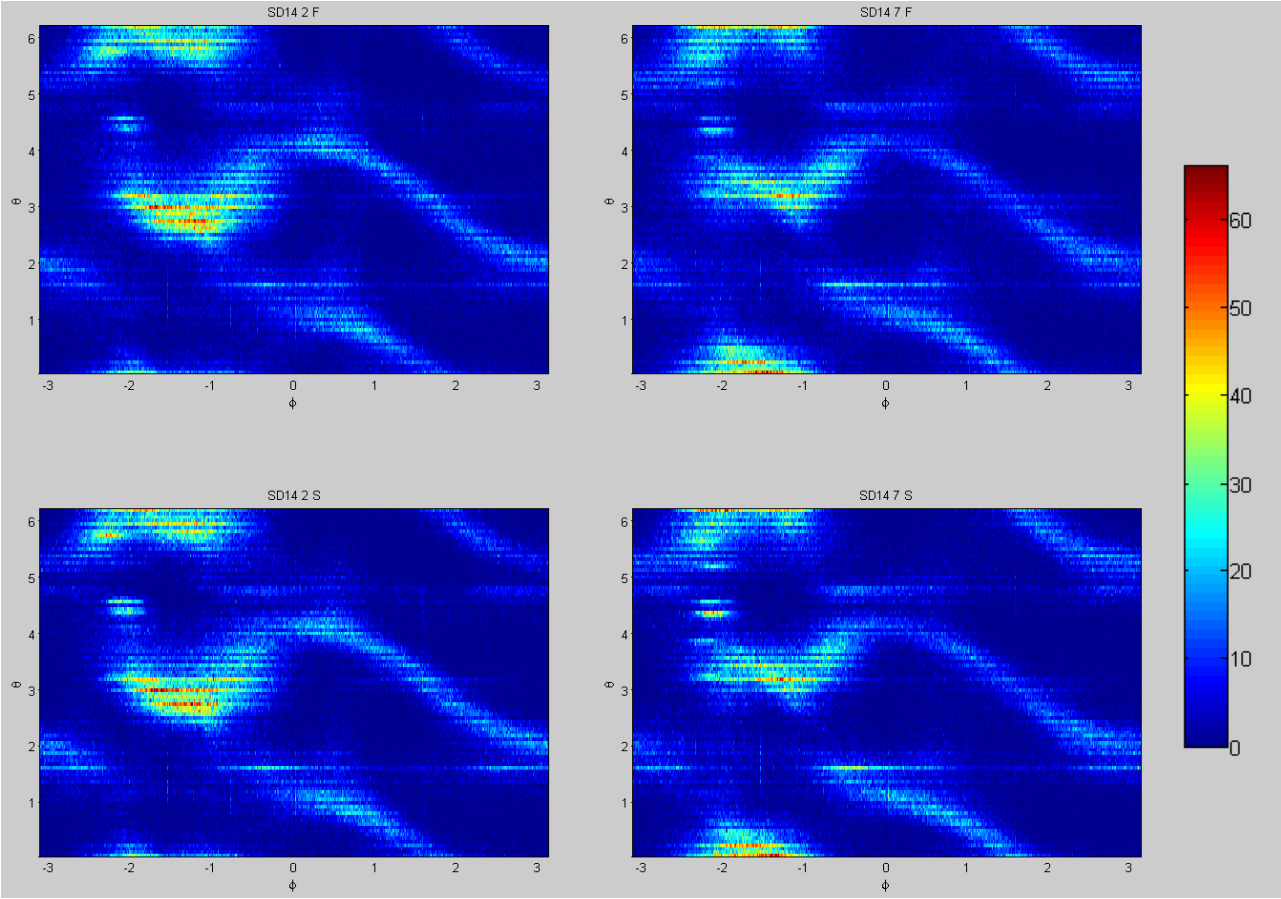

Abbildung 99: Häufigkeitsverteilungen der ausgerichteten Minutien bezüglich Winkel-Koordinate in Bezug zum Ursprung und Minutien-Ausrichtung

Core-Punkte sind auch wichtige charakteristische Referenzpunkte eines Fingerabdrucks. Im Folgenden werden die statistischen Eigenschaften der Core-Punkte analysiert. Die Anzahl der detektierten Core-Punkte der rechten Zeigefinger-Bilder der SD 14 Datenbank werden berechnet. Abbildung 100 zeigt ein Histogramm der Anzahl der detektierten Core-Punkte der Enrolment- und Verifikation-Bilder. Die x-Achse gibt die Anzahl der Core-Punkte wieder, die im Enrolment detektiert wurden. Auf einem Fingerabdruckbild konnten null, ein, zwei oder maximal vier Core-Punkte detektiert werden. Die Höhe jedes Balkens zeigt den Anteil der Enrolment-Fingerabdrücke, die eine bestimmte Anzahl von Core-Punkten aufweisen. Auf 18.8% der Fingerbilder wurde keine Core-Punkte gefunden. Nur ein Core-Punkt wurde auf 54.7% Fingerbilder detektiert. Auf 25.5%, 0.97% und 0.074% der Fingerabdrücke konnten jeweils zwei, drei und vier Core-Punkte gefunden werden.

Die unterschiedlichen Farben innerhalb eines Balkens zeigen die Anzahl der Core-Punkte für die gleichen Benutzer bei der Verifikation. Bei 54.7% der Benutzer, bei denen eine Core-Punkt im Enrolment gefunden wurde, kann nur bei 31.7% der Benutzer wieder einen Core-Punkt detektiert

werden. Dies entspricht 31.7%/54.7%=57.95% der Benutzer, bei denen eine Core-Punkt sowohl im Enrolment als auch wieder bei der Verifikation gefunden werden kann. 11.3% der Benutzer haben einen Core-Punkt im Enrolment und keine bei der Verifikation gefunden. 11.4% der Benutzer hat ein Core-Punkt im Enrolment und zwei während der Verifikation gefunden. Und 0.2% der Benutzer hat ein Core-Punkt im Enrolment und drei bei der Verifikation gefunden. Die Zuverlässigkeit der Core-Punkt-Detektion wird von der Qualität der Fingerabdrücke beeinflusst. Es ist denkbar, dass eine bessere Übereinstimmung der Anzahl von Core-Punkten beim Enrolment und der Verifikation erreicht werden kann, wenn die Qualität der Fingerabdrücke steigt.

In diesem Abschnitt wurden die statistischen Eigenschaften der Fingerabdrücke der Datenbank NIST SD 14 einschließlich Minutien-Position, -Ausrichtungswinkel sowie der Anzahl der Core-Punkte analysiert. Bei der Generierung der Streupunkte im Fuzzy-Vault müssen die statistischen Verteilungen der Minutien-Eigenschaften berücksichtigt werden. Die Streupunkte und unterstützenden Punkte müssen die gleichen statistischen Eigenschaften aufweisen, um so die Unterscheidung der unterstützenden Punkte von den Streupunkten zu erschweren. Deshalb können diese Eigenschaften beim Identifikationsverfahren nicht helfen, um beispielsweise die Streupunkte herauszufiltern. Es ist aber denkbar, dass andere Fingerabdruck-Informationen für eine Vorauswahl bei der Suche dienlich sein können. Solche Informationen müssen allerdings zuverlässig sein und eine gewisse Unterscheidungskraft aufweisen. Außerdem sollten sie keine zusätzlichen Informationen über die Beschaffenheit des Fingerabdrucks freisetzen. Ein Beispiel dafür wären Core-Punkte, da diese nur wenig Information über den Fingerabdruck selbst preisgeben. Wegen ihrer Unzuverlässigkeit sind sie jedoch nicht geeignet für eine Vorauswahl.

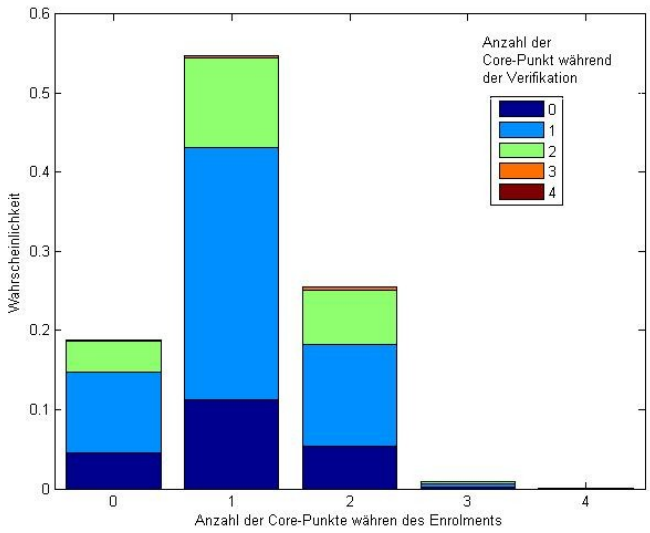

Abbildung 100: Histogramm der Anzahl der detektierten Core-Punkte bei Enrolment und Verifikation

6.6 Fazit und Ausblick

Alle in diesem Kapitel vorgestellten Verfahren GeoMatch, Matrix-Comparator, BioSimJoin, BioSimJoin* sowie BioNN versuchen durch eine approximierte Ähnlichkeitsberechnung eines Anfragetemplates mit Elementen der Datenbank ein Ranking der Datenbanktemplates zu erstellen. Auf diese Weise kann der genaue, aber aufwändige Authentifikationsvergleich zunächst auf Datenbanktemplates mit einer hohen Trefferwahrscheinlichkeit ausgeführt werden, wodurch die Anzahl der insgesamt zu betrachteten Datenbanktemplates sehr stark eingeschränkt wird. Die drei Verfahren GeoMatch, Matrix-Comparator und BioNN versuchen dabei den paarweisen Vergleich zweier Templates durch Approximationen zu beschleunigen. Bei all diesen Verfahren ist ein weiterer Geschwindigkeitsvorteil zu erzielen, sofern eine geringere Approximationsgenauigkeit akzeptabel ist. Da in diesen Verfahren dennoch alle paarweisen Vergleiche zwischen Anfrage- und Datenbanktemplates durchgeführt werden müssen um ein Ranking zu erstellen, ist hier immer eine linear wachsende Laufzeit bei steigender Datenbankgröße zu erwarten. Um eine solche linear steigende Laufzeit zu vermeiden, sind Indexstrukturen unverzichtbar. Für Bereichsanfragen mehrdimensionaler Objekte, wie sie durch die unscharfe Suche auf Minutien erforderlich sind, eignen sich insbesondere R- bzw. R*-Bäume. Ein entsprechender Ansatz, sowie entsprechende experimentelle Beobachtungen, zeigt hier die indexierte Variante von BioSimJoin, der Ansatz BioSimJoin*.

Für den Fall, dass eine Identifikation der Anfrageperson nicht möglich ist, entweder aufgrund einer verfälschten Aufnahme oder wenn die Anfrageperson nicht in den Referenzdaten enthalten ist, wird, unabhängig vom zugrunde liegenden Identifikationsverfahren, ohne weitere Rand-bedingungen das gesamte Ranking und damit die gesamte Datenbank für den Authentifikations-vergleich betrachtet. Es ist daher empfehlenswert ein Schwellwertkriterium zu definieren, durch welches ein vorzeitiger Abbruch der Identifikationslösung möglich ist, falls die durch die vorgestellten Verfahren approximierten Ähnlichkeit zu gering ist.

Zusammenfassend ist das Leistungsverhalten aller fünf Verfahren in Tabelle 17 dargestellt. Es ist jeweils die durchschnittlich benötigte Laufzeit für die Identifikation des Rankings mithilfe des entsprechenden Verfahrens angegeben, wenn eine Person als Anfrageperson an eine Datenbank, die insgesamt Daten für 100 Personen enthält, verwendet wird. Desweiteren enthält die Tabelle jeweils die Position, an der sich die Anfrageperson durchschnittlich innerhalb des Rankings von insgesamt 100 Personen befindet. GeoMatch, BioNN und Matrix-Comparator kompensieren explizit Drehungen und Verschiebungen innerhalb der Daten. Die dadurch erwartete höhere Effektivität konnte allerdings auf dem NIST SD14 Datensatz nicht nachgewiesen werden. In jedem Falle erfordern diese Verfahren jedoch die höchsten Laufzeiten. BioNN erzielt im Vergleich zu GeoMatch und Matrix-Comparator eine gesteigerter Laufzeiteffizienz. Dies ist dadurch zu erklären, dass BioNN als heuristische Umsetzung von GeoMatch betrachtet werden kann, da der Suchraum bei BioNN gegenüber dem Suchraum bei GeoMatch stark eingegrenzt ist. Bei BioNN besteht ein zu untersuchendes Dreieck jeweils aus einer Minutie der Anfrageperson und den beiden Minutien, die dieser Minutie am nächsten liegen. Jedes solche Dreieck ist also eindeutig bestimmt. Bei GeoMatch muss dagegen eine viel größere Zahl von relevanten Dreiecken betrachtet werden, da diese Menge lediglich durch die maximale Seitenlänge beschränkt wird. Dazu kommt, dass BioNN im Gegensatz zu GeoMatch keine Unterscheidung zwischen globalen und lokalen Rotationen vornimmt. Das beste Laufzeitverhalten erzielt das Verfahren BioSimJoin*, da hier explizit der Einsatz einer geeigneten Indexstruktur zum Tragen kommt. In durchschnittlich nur 34 ms wird ein Ranking bestimmt, in dem sich die Anfrageperson durchschnittlich an Position 44,4 befindet. BioSimJoin

ermittelt dasselbe Ranking ohne Indexunterstützung in 1.465 ms. Da weder bei BioSimJoin, noch bei BioSimJoin* Rotationen bzw. Verschiebungen gezielt berücksichtigt werden, ist (v.a. das indexierte Verfahren BioSimJoin*) den anderen Verfahren hinsichtlich Laufzeit überlegen, kann jedoch bezüglich Effektivität keine mit BioNN vergleichbaren Ergebnisse liefern.

In den Identifikationslösungen wurden keine zusätzlichen biometrischen oder personenbezogenen Informationen verwendet. Deswegen ist die Sicherheit der Verfahren sowie die Fähigkeit die Privatsphäre zu schützen vergleichbar mit der der Verifikationsverfahren. Es ist denkbar, andere Information zu benutzen, die einerseits die Suchprozesse beschleunigen und anderseits nicht viele Informationen über biometrische Daten preisgeben.

Außerdem wurden die statistischen Eigenschaften des Fingerabdrücke untersucht. Die Minutien sind im mittleren Bereich der Fingerabdrücke konzentriert. Es gibt eine Abhängigkeit der Minutien-Ausrichtung und -Positionswinkel. Core-Punkte sind die wichtigen Charakteristika der Fingerabdrücke. Aufgrund der schlechten Qualität der Fingerabdruckbilder ist die Core-Punkte-Detektion nicht zuverlässig. Die Core-Punkte sind nicht eignet, um die Suchprozesse zu verbessern.

	GeoMatch	**Matrix-Comparator**	**BioSimJoin**	**BioSimJoin***	**BioNN**
Laufzeit	838 ms	9.120 ms	1.465 ms	34 ms	1.521 ms
Position	48,8	46,07	44,4	44,4	21,9

Tabelle 17: Gegenüberstellung aller Verfahren hinsichtlich Laufzeit und Genauigkeit (DB-Größe: 100 Personen)

7 Standardisierung

Dieses Kapitel beschreibt die im Kontext Template Protection relevante Standardisierung im Überblick. Ein besonderer Schwerpunkt liegt auf dem Standard ISO/IEC 24745 Biometric Template Protection, der wesentlich durch das EU-Projekt TURBINE beeinflusst und durch die bereits in Kapitel 2 dargestellte Referenzarchitektur mitgestaltet wurde. Daher wird in diesem Kapitel zunächst das TURBINE Projekt vorgestellt und dann ein Überblick zur ISO Standardisierung gegeben bevor im folgenden Abschnitt eine Zusammenfassung von ISO/IEC 24745 erfolgt.

7.1 Das EU-Projekt TURBINE

Das TURBINE (TrUsted Revocable Biometric IdeNtitiEs) ist ein integriertes Forschungsprojekt, das von der Europäischen Union im siebten Rahmenprogramm (FP7) über einen Zeitraum von drei Jahren seit 2008 gefördert wird [TUR08]. In diesem Projekt werden Technologien zu innovativen Identity Management Systemen erforscht und entwickelt. Der wesentliche Technologie-Schritt ist dabei die Kombination von Nutzer-Identifikation mittels Fingerbilderkennung mit zuverlässigen Methoden zum Schutz der biometrischen Referenzdaten durch fortgeschrittene kryptografische Verfahren. In diesem Projekt arbeiten unter der Konsortial-Leitung von Sagem Securite verschiedene Unternehmen und akademische Forschungsinstitutionen zusammen. Zu den Konsortialpartnern zählen die Unternehmen Precise Biometrics AG, Philips Research Europe, Cryptolog, Sagem Orga, ARTTIC, 3DSA sowie die Hochschulen K.U. Leuven, University of Twente und das Gjøvik University College.

Die Zielsetzung des Projektes ist einerseits die Erforschung von Template-Protection-Verfahren, die eine Transformation von Fingerprint-Minutientemplates mittels einer Einweg-Funktion ermöglicht, sodass eine Rekonstruktion des ursprünglichen Samples unmöglich wird. Im gleichen Zuge soll die Transformation ermöglichen, unterschiedliche pseudonyme Identifikatoren von seiner biometrischen Charakteristik abzuleiten, so dass Identifikatoren, die in unterschiedlichen Anwendungen für eine betroffene Person gespeichert werden, keinerlei Querbezüge zwischen den Anwendungen erlauben und zu unerwünschten Auswirkungen für die betroffene Person führen könnten. Andererseits umfassen die Aktivitäten im TURBINE Projekt aber auch die Analyse der Protokollsicherheit und die Analyse der Erkennungsleistung der entwickelten Template-Protection-Verfahren. Dies wird in eigenständigen Teilprojekten untersucht. Das TURBINE Konsortium möchte mit seiner Arbeit die Entwicklung des ISO/IEC 24745 positiv beeinflussen und hat dazu in der bisherigen Projektlaufzeit bereits wesentliche Beiträge geliefert [Bre08].

Die Erprobung der entwickelten Technologien erfolgt in zwei Anwendungen: In einer Telematikanwendung wird die biometrische Authentisierung mit einer Health-Professional-Card verbunden und in einer generischen Anwendung eines Apotheken-Szenarios evaluiert. In einer Access-Control-Anwendung wird die physikalische Zugangskontrolle von Mitarbeitern zu Sicherheitsbereichen des Flughafens Thessaloniki evaluiert.

7.2 Überblick zu Standardisierung

Die Standardisierung im Bereich der Informationstechnologie wird von einem Joint Technical Committee (JTC) zwischen der International Organization for Standardization (ISO) und der

International Electrotechnical Commission (IEC) erarbeitet. Mit der Biometriestandardisierung beauftragt wurde das im Jahr 2002 etablierte Subcommittee SC37, das seine Arbeit in sechs Working Groups durchführt, die sich mit den Themen Harmonized Biometric Vocabulary, Biometric Technical Interfaces, Biometric Data Interchange, Biometric Functional Architecture, Biometric Testing and Reporting sowie Cross-Jurisdictional and Societal Aspects beschäftigen. Parallel dazu wird im Subcommittee SC27 die Standardisierung von Sicherheitsverfahren sowie im Subcommittee SC17 die Standardisierung von SmartCards und deren Kommunikationsprotokolle bearbeitet. Die folgende Abbildung zeigt die verschiedenen Gremien, sowie deren formalisierte Verbindungen (Liasions), die untereinander aber auch mit der International Civil Aviation Organization (ICAO) über SC17 eingerichtet wurden.

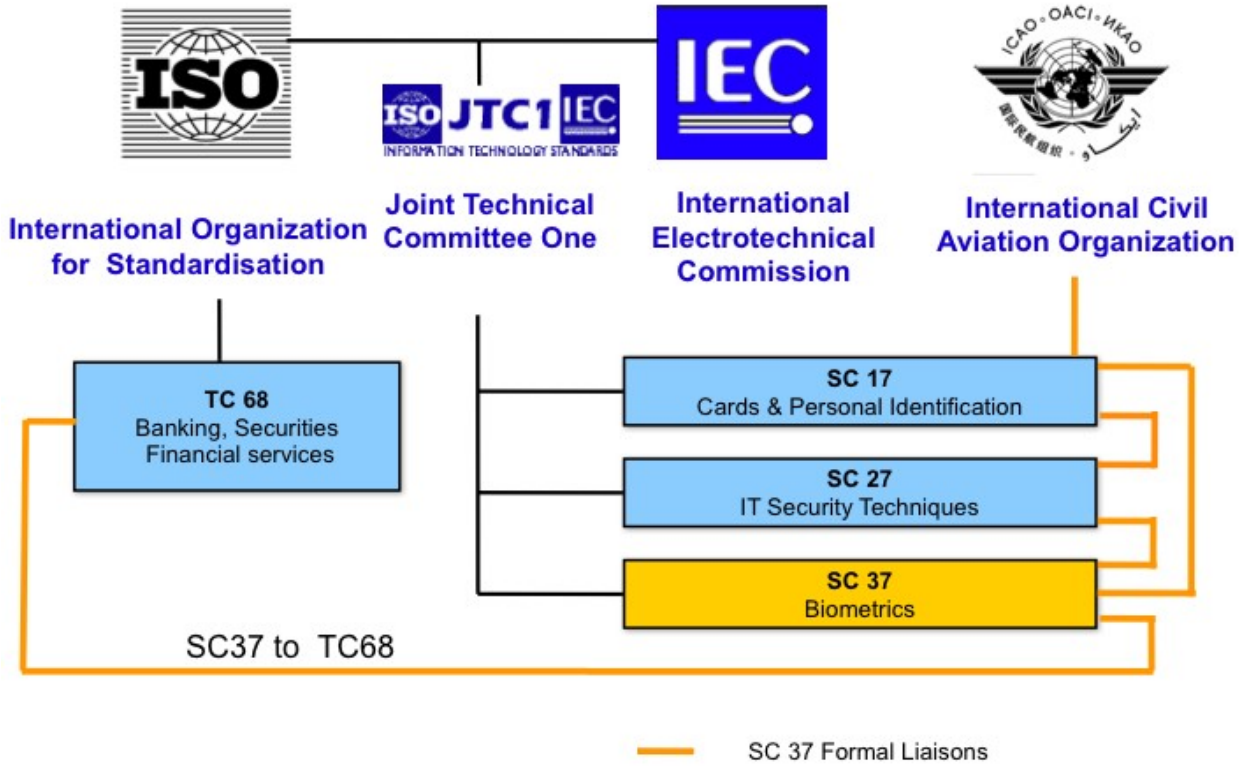

Abbildung 101: Zusammenwirken der internationalen Standardisierungs-Komittees

Bei der Entwicklung von ISO Standards gilt als Voraussetzung, dass eine ausgereifte Technologie vorliegen muss und Entscheidungen die zu einer einheitlichen Systemsicht bzw. einem einheitlichen Datenaustauschformat führen, im Konsens des Subcommittee gefunden werden müssen. Dazu werden mehrere Entwicklungsstufen durchlaufen, die sich wie folgt darstellen:

- Working Draft (WD)
- Committee Draft (CD)
- Final Committee Draft (FCD)
- Final Draft International Standard (FDIS)
- International Standard (IS)

Sofern ein Standardentwurf vollumfänglich eine Technologie beschreibt und Konsens über die technischen Festlegungen besteht, kann das Subcommittee den Entwurf zum CD-level befördern. Wesentliche technische Ergänzungen zu einem Standardentwurf sind danach nur bis zum FCD-level möglich.

Das Zusammenspiel der relevanten Biometrie-Standards wird in der folgenden Zwiebel-Schalen-Darstellung deutlich.

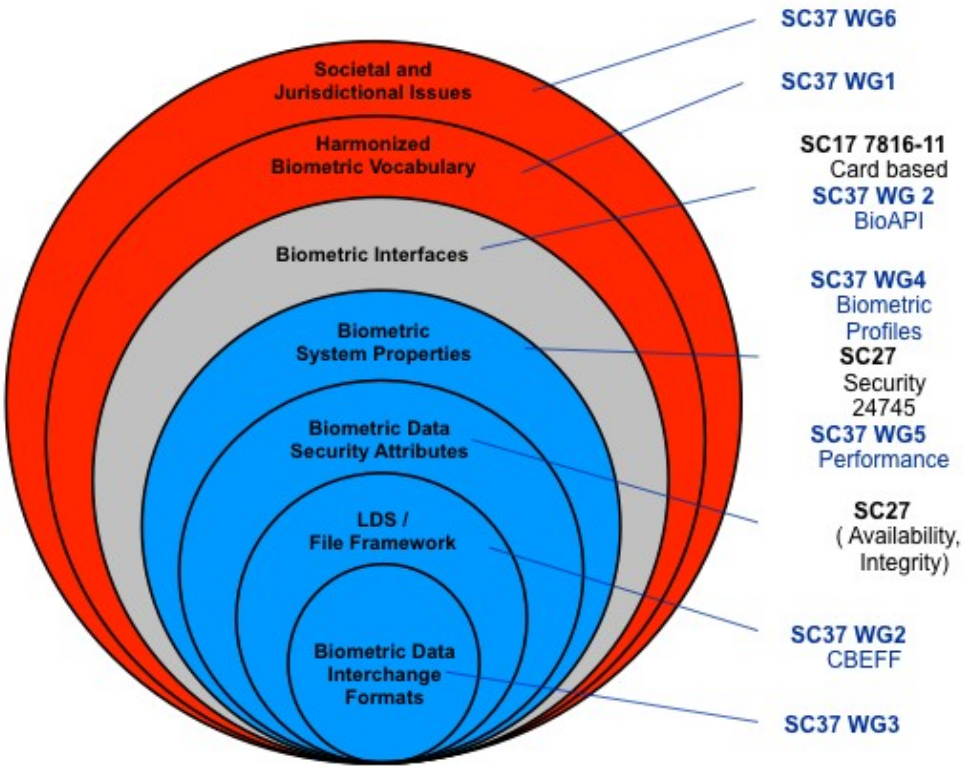

Abbildung 102: Zwiebel-Schalen-Modell der biometrischen Standardisierung

Im Kern des Zwiebel-Schalen-Modells werden die Datenaustauschformate im SC37 behandelt. In der zweiten Schale werden logische Datenstrukturen definiert, um assoziierte Metadaten zu speichern. Ein bekannter Standard ist das Common Biometric Exchange Formats Framework (CBEFF), das als ISO/IEC19785 publiziert wurde. In der dritten Schale werden Sicherheitsattribute definiert, die der Verfügbarkeit, Authentizität, Integrität und Vertraulichkeit von biometrischen Informationen dienen. In der vierten Schale werden Systemeigenschaften beschrieben, wie etwa die Erkennungsleistung eines biometrischen Systems. Die fünfte Schale behandelt Schnittstellen zu biometrischen Sensoren und Komponenten, die für die Systemintegration relevant sind.

Ein wichtiger Teil der Tätigkeiten des JTC1 Subcommittees SC37 ist es, Datenaustauschformate zu formulieren, nach denen die Repräsentation einer biometrischen Charakteristik, z.B. eines Gesichtsbilds oder des Bilds eines Fingerabdrucks, in einem spezifizierten Datensatz kodiert werden kann. Dieser Datensatz kann dann als Referenz in einer SmartCard oder in einer Datenbank abgelegt werden. Wenn es sich um ein offenes System handelt, muss diese Referenz interoperabel

sein, d.h. ein anderer Hersteller muss das Format des Datensatzes lesen und verstehen können und zudem auf derartigen Daten eine gute Erkennungsleistung herstellen können.

Besonders weite Verbreitung gefunden haben die Biometrie-Standards für elektronische Reisepässe. Bei der Spezifikation des ePasses war die Interoperabilität ein sehr hohes Ziel, so dass bei der Definition der Logischen Datenstruktur (LDS) zwei bildbasierte Standards eingeflossen sind: Das ist einerseits der Standard ISO/IEC IS 19794-5 [ISOface] zur Speicherung von Gesichtsbildern und andererseits der Standard ISO/IEC IS 19794-4 [ISOfinger]zur Speicherung von Fingerbildern.

Für den Bereich Biometrische Systeme ist die wichtigste Tätigkeit des JTC1 Subcommittees SC17 die Bearbeitung des On-Card-Comparison Standards ISO/IEC 24789, der für Token-basierte Systeme relevant ist.

Das JTC1 Subcommittee SC27 beschäftigt sich in der Working Group 5 unter anderem mit Identity Management Systemen und Privacy-Enhancing-Technologies und behandelt in diesem Kontext auch Biometrische Verfahren und den Schutz von biometrischen Referenzdaten.

7.3 Biometrische Systeme nach ISO/IEC 24745

In diesem Kapitel wird die Entwicklung des für Template Protection relevanten Standards ISO/IEC 24745 Biometric Template Protection dargestellt, der im JTC1 Subcommittee 27 in der Working Group 5 bearbeitet wird. Die in Kapitel 2 vorgestellte TURBINE Referenzarchitektur war Grundlage für die Erweiterung des ISO Standards 24745 im Oktober 2008, wodurch Mechanismen zur Erneuerbarkeit und Diversifikation von biometrischen Referenzen in den Standard integriert wurden. Diese Beiträge finden sich nun in den Kapiteln 6.1.4 und 7.2 sowie im Annex C von ISO/IEC 24745.

Der folgende Text stellt eine Zusammenfassung des ISO/IEC Committee Draft[13] 24745 „Information technology – Security techniques – Biometric template protection“ dar [ISOtp]. Die im Weiteren verwendete Terminologie entspricht den Begriffsdefinitionen aus ISO/IEC 24745 und den darin referenzierten Dokumenten (siehe Stichwortverzeichnis und auch die Erläuterung der Begriffe zur Referenzarchitektur in Kapitel 2).

Biometrische Systeme dienen der automatischen Erkennung von Individuen auf Basis physiologischer und verhaltensabhängiger Charakteristika. Diese Charakteristika sind physische Eigenschaften von Körperteilen, das Resultat physiologischer und verhaltensbedingter Prozesse des Körpers bzw. Kombinationen aus beidem. Bei den am häufigsten verwendeten physiologischen Charakteristika handelt es sich um Fingerabdrücke, das Gesicht, die Iris, Handgeometrien, Hand- und Fingervenen, die Retina sowie DNA. Die gebräuchlichsten verhaltensabhängigen Charakteristika sind Handschrift, Gang und Stimme. All diese Charakteristika erfüllen die folgenden im Kontext biometrischer Systeme wünschenswerten Eigenschaften [JBN99]:

- Allgemeingültigkeit (jedes Individuum sollte über diese Charakteristika verfügen)
- Eindeutigkeit (jedes Individuum sollte über eine unterschiedliche Ausprägung der Charakteristika verfügen)
- Beständigkeit (die Charakteristika sollten keine durch einen Alterungsprozess hervorgerufenen Varianzen aufweisen)

13 Stand Juli 2009

- Erfassbarkeit (die Charakteristika sollten quantitativ erfassbar (messbar) sein).

Biometrische Systeme werden im Wesentlichen zur Authentisierung und Identifikation eines Individuums eingesetzt. Hierzu vergleicht ein biometrisches System eine vom Individuum genommene Probe mit einer oder mehreren gespeicherten biometrischen Referenzen. Bei einer biometrischen Referenz (BR) handelt es sich um ein biometrisches Sample, ein biometrisches Template oder ein biometrisches Modell, das ein Individuum eindeutig innerhalb eines bestimmten Kontextes identifizieren kann.

Ein biometrisches System besitzt im Wesentlichen die folgenden drei Teilprozesse:

- **Enrolmentprozess**: Biometrische Charakteristika werden erfasst, verarbeitet und in Form einer biometrischen Referenz zusammen mit der Identitätsreferenz (biometrischer Datensatz, siehe Abschnitt 7.4.3) des Individuums gespeichert.
- **Identifikationsprozess**: Identifikation eines Individuums auf Basis der erfassten biometrischen Charakteristika und gespeicherten biometrischen Referenzen.
- **Verifikationsprozess**: Verifikation der Identität eines Individuums durch Vergleichen einer biometrischen Probe mit gespeicherten biometrischen Referenzen unter Verwendung der Identitätsreferenz des Individuums. Bei einer Identitätsreferenz (IR) handelt es sich um ein oder mehrere nicht-biometrische Attribute eines Individuums (z.B. Name, Personalausweisnummer, etc.) bzw. dessen Identität in einem bestimmten Kontext. Identitätsreferenzen werden in einer Identitätsreferenzdatenbank (Database for Identity References - DBIR) gespeichert.

Zur Vermeidung von Identitätskonflikten und zur Gewährleistung eines akkuraten Identifikations- und Verifikationsprozesses bedarf es der Zusammenarbeit zwischen biometrischem System und Identitätsmanagementsystem (IdMS; siehe ISO/IEC CD 24760 für weitere Details [ISOidm]).

Die Architektur eines biometrischen Systems gliedert sich nach [ISOtp] in die folgenden fünf Subsysteme:

- **Biometrisches Erfassungssubsystem**: Beinhaltet biometrische Datenerfassungsgeräte oder Sensoren und bildet die erfassten biometrischen Charakteristika auf biometrische Samples ab.
- **Signalverarbeitungssubsystem**: Extrahiert biometrische Merkmalsdaten aus biometrischen Samples.
- **Datenspeichersubsystem**: Dient zur Speicherung erfasster biometrischer Referenzen und Identitätsreferenzen, meist in separaten Datenbanken.
- **Vergleichssubsystem**: Vergleicht erfasste biometrische Samples mit gespeicherten biometrischen Referenzen und liefert ein Ähnlichkeitsmaß (Vergleichswert) als Ergebnis.
- **Entscheidungssubsystem**: Entscheidet auf Grund des Ähnlichkeitsmaßes und einer Entscheidungsrichtlinie über die Identität des zum erfassten biometrischen Sample und der gespeicherten biometrischen Referenz gehörenden Individuums.

Darüber hinaus kann ein biometrisches System über weitere funktionale Einheiten, wie zum Beispiel ein Referenz-Adaptions-Subsystem zur automatischen Anpassung einer gespeicherten biometrischen Referenz oder ein Administrationssubsystem zur Verwaltung des biometrischen Systems, verfügen.

Die konzeptionelle Struktur eines biometrischen Systems sowie die zwischen den Komponenten während eines Enrolmentprozesses bzw. eines Verifikations- oder Identifikationsprozesses ausgetauschten Daten veranschaulicht Abbildung 103.

Biometrische Referenzen, im Speziellen in Verbindung mit Identitätsreferenzen, stellen sensible personenbezogene Daten (Personal Identifiable Information - PII) [NTN02] dar, die entweder unmittelbar (z.B. Gesichtsfoto) oder indirekt (z.B. Fingerabdruck-Minutien) zur Identifizierung einer Person und auf Grund deren Eindeutigkeit potenziell als eindeutiger Identifikator (Universal Unique Identifier - UUID) zur datenbankübergreifenden Verknüpfung von Daten genutzt werden können. Dies stellt eine Gefährdung der Privatsphäre des Individuums dar und sollte vermieden werden. Insbesondere sollten biometrische Daten im Besitz und unter Kontrolle der betroffenen Person bleiben und biometrische Samples von biometrischen Systemen nur gespeichert werden, wen dies dringend erforderlich ist. Weiterhin sollte ein biometrisches System Mechanismen zum Erzeugen diversifizierbarer Referenzen zur Verfügung stellen, um das Widerrufen und Erneuern biometrischer Referenzen zu ermöglichen. Zusätzlich zu den Aspekten der persönlichen Privatsphäre sollten gesellschaftliche Aspekte wie Barrierefreiheit, Gesundheit und Sicherheit der Nutzer biometrischer Systeme, Nutzerfreundlichkeit und Nutzerakzeptanz beim Einsatz biometrischer Systeme berücksichtigt werden (siehe ISO/IEC TR 24714-1 [ISOh], [ISOi]).

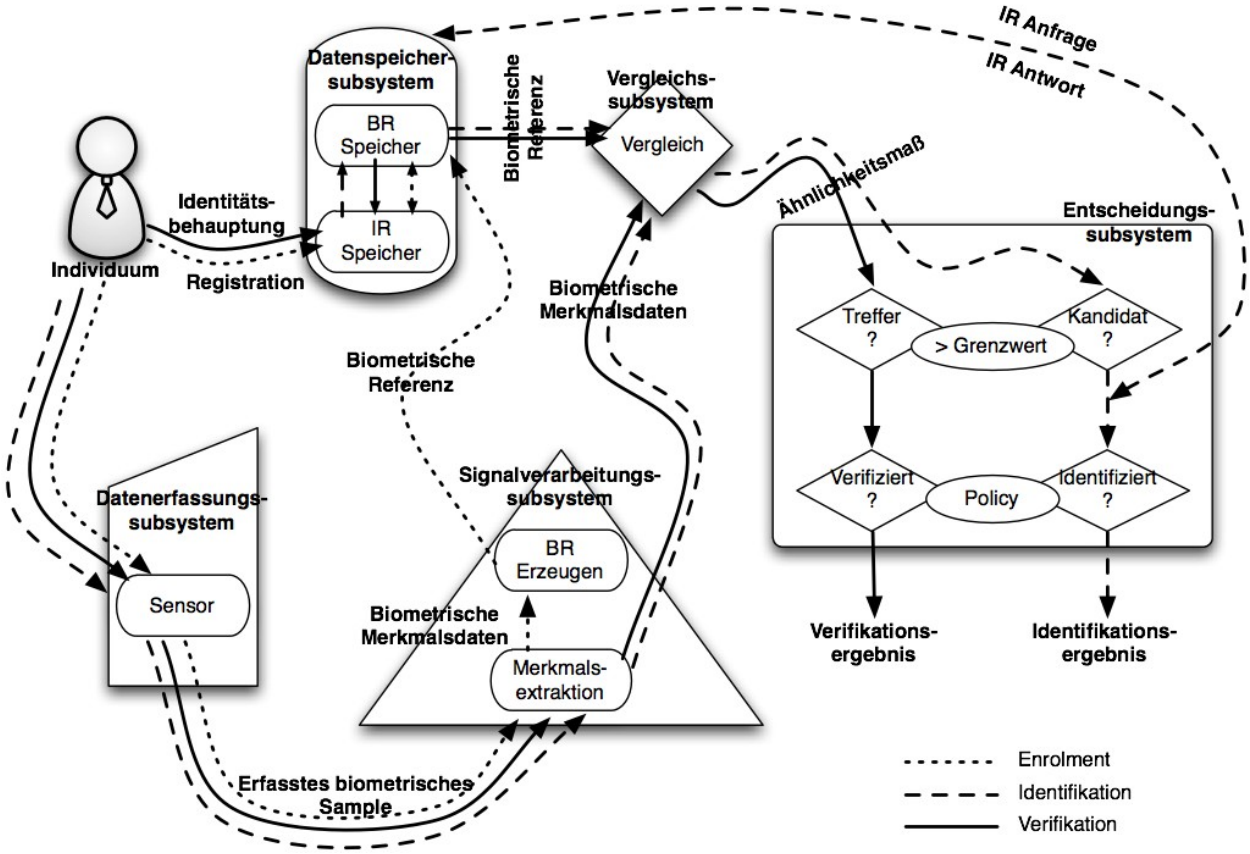

Abbildung 103: Konzeptionelle Struktur eines biometrischen Systems

Die Abbildung 103 entspricht der Referenzarchitektur eines biometrischen Systems, wie sie von ISO SC37 entwickelt wurde. Im Unterschied zur Referenzarchitektur in Kapitel 2 ist in Abbildung

103 jedoch die Komponentenbeschreibung eines generischen biometrischen System dargestellt. In einem herkömmlichen Biometrie-System wird eine biometrische Referenz (BR) nach Merkmalsextraktion als biometrisches Template oder als biometrisches Modell (z.B. ein in der Sprechererkennung übliches Personen-spezifisches Hidden-Markov-Modell) gebildet. In einem Biometrie-System, das Template-Protection Verfahren entsprechend der Referenzarchitektur in Kapitel 2 integriert hat, wird eine biometrische Referenz aus den beiden Komponenten pseudonymer Identifikator (PI) und den unterstützenden Daten (AD) gebildet, wie in Kapitel 2 beschrieben. Das in Abbildung 103 gezeigte Datenspeichersubsystem kann ein lokaler oder zentraler Speicher sein, in dem die biometrische Referenz gegebenenfalls verteilt gespeichert wird. Die verschiedenen Modelle eines möglichen Systementwurfs werden am Ende dieses Kapitels aufgezeigt.

7.4 Sicherheitsaspekte biometrischer Systeme

In diesem Abschnitt werden Sicherheitsaspekte biometrischer Systeme beschrieben. Die Struktur dieses Abschnittes und die Auswahl der Anforderungen ist aus [ISOtp] übernommen. Es werden zunächst die grundlegenden Sicherheitsanforderungen an biometrische Systeme sowie Angriffsvektoren biometrischer Systeme und mögliche Gegenmaßnahmen erläutert. Im Anschluss wird die Sicherheit biometrischer Daten in Bezug auf Art und Weise der Speicherung biometrischer Daten diskutiert. Die Formulierung der Anforderungen überschneidet sich mit der Darstellung in Abschnitt 2.3. Sie wurde jedoch bewusst hier aufgenommen, um die Zielsetzung von ISO/IEC CD 24745 darzustellen.

7.4.1 Sicherheitsanforderungen

Nachfolgend werden die Sicherheitsanforderungen an biometrische Systeme erläutert. Die Sicherheitsanforderungen lassen sich in die Teilaspekte Vertraulichkeit, Integrität, Verfügbarkeit sowie Erneuerbarkeit und Widerrufbarkeit unterteilen.

Vertraulichkeit

Vertraulichkeit ist die Eigenschaft, die den Schutz von Informationen vor nicht autorisiertem Zugriff und unerlaubter Veröffentlichung beschreibt. Im Kontext biometrischer Systeme bedeutet dies den Schutz biometrischer Daten, wie z.B. biometrischer Samples oder biometrischer Referenzen.

Während eines Identifikations- oder Verifikationsprozesses werden die während des Enrolmentprozesses in einer biometrischen Referenzdatenbank (Database for Biometric References - DBBR) gespeicherten biometrischen Referenzen an das Vergleichssubsystem übermittelt und dort mit dem erfassten biometrischen Sample (der biometrischen Probe) verglichen. Auf Grund möglicher räumlicher Distanzen zwischen der biometrischen Referenzdatenbank und des Vergleichssubsystems ist die Übermittlung von Daten über potenziell unsichere Kommunikationswege erforderlich und stellt somit eine Gefährdung der übermittelten Daten auf Grund der Möglichkeiten des Lesens, Veränderns oder Ersetzens der Daten durch unautorisierte Dritte dar.

Zur Wahrung der Vertraulichkeit biometrischer Daten ist daher der Einsatz von Kryptografie, wie zum Beispiel symmetrischer oder asymmetrischer Verschlüsselungsverfahren oder die Verwendung

pseudonymer Identifikatoren zum Schutz gespeicherter und übertragener Daten notwendig. Eine informative Übersicht derartiger Methoden wird in Abschnitt 7.8 gegeben.

Integrität

Integrität ist die Eigenschaft, die die Unversehrtheit und Korrektheit von Daten und Verfahren sicherstellt. Die Integrität einer biometrischen Referenz ist elementar, um Aussagen über die Gesamtsicherheit eines biometrischen Systems treffen zu können. Die Integrität der Authentisierung hängt unmittelbar von der Integrität der biometrischen Referenz ab. Nicht-vertrauenswürdige biometrische Referenzen oder nicht-vertrauenswürdige biometrische Samples führen zu einer nicht-vertrauenswürdigen Authentisierung. Nicht-vertrauenswürdige biometrische Referenzen oder biometrische Samples können auf Grund von fehlerhaft arbeitender Hard- oder Software, durch beabsichtigte oder unbeabsichtigte Modifikation einer biometrischen Referenz oder durch das Ersetzen einer gespeicherten biometrischen Referenz durch eine alternative Referenz durch einen Angreifer oder eine befugte Person entstehen.

Zur Wahrung der Integrität müssen biometrische Systeme geeignete Sicherheitsmaßnahmen vorsehen. Diese Maßnahmen können enthalten, sind jedoch nicht limitiert auf, hinreichende Kontrolle des Zugriffs auf biometrische Daten sowie kryptografische Maßnahmen zur Ermöglichung einer Integritätsprüfung (siehe Abschnitt 7.8), idealer Weise kombiniert mit weiteren Techniken (wie zum Beispiel Timestamping) zum Schutz vor missbräuchlicher Verwendung gestohlener biometrischer Daten oder Replay-Attacken.

Verfügbarkeit

Verfügbarkeit ist die Eigenschaft eines Systems, bei Bedarf eines autorisierten Nutzers zugänglich und funktionsfähig zu sein. Eine Beeinträchtigung der Funktionsfähigkeit eines biometrischen Systems kann aus Löschen, Verschieben oder Zerstören notwendiger biometrischer Daten in einer biometrischen Referenzdatenbank durch einen Angreifer resultieren. Zum Schutz vor derartiger Beeinträchtigungen ist die Durchführung einer Zugriffskontrolle auf die biometrische Referenzdatenbank notwendig. Integritätsprüfungen können darüber hinaus die Verfügbarkeit legitimer biometrischer Daten erhöhen.

Zusätzlich zur Sicherstellung der Verfügbarkeit korrekter biometrischer Daten ist die Verfügbarkeit der Komponenten und Ressourcen eines biometrischen Systems zu gewährleisten. Beispielsweise könnte ein biometrisches System durch überhöhten Netzwerkverkehr überlastet oder durch den Ausfall von Verbindungen zwischen Subsystemen gestoppt werden wodurch eine Beeinträchtigungen der Nutzbarkeit eines biometrischen Systems entsteht. Diese Risiken können durch vorausschauende Kapazitätsplanung sowie durch hinreichende Redundanz des Netzwerks minimiert werden.

Biometrische Systeme verfügen über ein immanentes Sicherheitsrisiko (vgl. Kapitel 2). Biometrische Messungen sind inhärent störanfällig, wodurch jede Messung in einem leicht variierenden biometrischen Sample und somit biometrischem Template resultiert und es dadurch eines hinreichend robusten Verifikationsprozesses bedarf. Diese Robustheit führt dazu, dass ein Angreifer mit geringer Wahrscheinlichkeit fälschlicher Weise erfolgreich authentifiziert wird (Falsch-Akzeptanz). Sicherheitsmaßnahmen, die hohen Falsch-Akzeptanz-Raten entgegenwirken führen in der Regel zu erhöhten Falsch-Rückweisungs-Raten und können hierdurch die Verfügbarkeit des Systems beeinträchtigen. Gegenmaßnahmen sollten daher vorsichtig eingeführt werden, sodass die Erkennungsleistung des biometrischen Systems nicht das für die jeweilige Anwendung akzeptable Maß unterschreitet.

Erneuerbarkeit und Widerrufbarkeit

Die Kompromittierung biometrischer Referenzen stellt ein wesentliches, die Sicherheit eines biometrischen Systems und die Privatsphäre eines Individuums gefährdendes, Risiko dar. Individuen verfügen über eine limitierte Anzahl von z.B. Fingern und Iriden, deren Nutzung im Kontext biometrischer Authentisierung durch Kompromittierung korrelierender biometrischer Referenzen für immer unmöglich wird. Für bestimmte Angriffsarten kann diesem Risiko durch Verwendung von Methoden zur Erneuerung und zum Widerrufen biometrischer Referenzen entgegengewirkt werden. Bei Erneuerbarkeit und Widerrufbarkeit biometrischer Referenzen handelt es sich um bedeutende Maßnahmen zur Wahrung der Privatsphäre eines Individuums durch Vermeidung unerwünschter Verknüpfungen über Datenbanken hinweg sowie zur Erhöhung der Datenbanksicherheit durch Datenseparierung.

Erneuerbare biometrische Referenzen (Renewable Biometric Reference - RBR) werden durch Diversifikation im Erstellungsprozess erzeugt und erlauben das Generieren mehrerer unterschiedlicher, zum selben biometrischen Charakteristikum gehörender biometrischer Referenzen. Erneuerbare biometrische Referenzen bestehen aus einem pseudonymen Identifikator (PI) sowie dazugehörenden unterstützenden Daten (AD), die beide – wie in Kapitel 2 beschrieben – während des Enrolmentprozesses erzeugt werden.

Erneuerbare biometrische Referenzen bieten nicht nur einen Schutz gegen Kompromittierung biometrischer Referenzen, sondern erlauben zusätzlich z.B. das zeitliche Limitieren der Gültigkeit einer biometrischen Referenz unter Verwendung ergänzender Daten (SD) während des Erstellungsprozesses.

7.4.2 Sicherheitsgefährdungen und Gegenmaßnahmen

Dieses Kapitel beschreibt potenzielle Schwachstellen biometrischer System ausgehend von der in Abbildung 103 dargestellten grundsätzlichen Architektur biometrischer Systeme sowie Gegenmaßnahmen zur Verhinderung derer Ausnutzung.

Wie in Abschnitt 7.3 beschrieben, bestehen biometrische Systeme aus mindestens fünf miteinander verbundenen Subsystemen (Datenerfassungssubsystem, Signalverarbeitungssubsystem, Datenspeichersubsystem, Vergleichssubsystem, Entscheidungssubsystem), von denen jedes Subsystem sowie die zwischen den Systemen liegenden Kommunikationskanäle eigene Angriffsvektoren besitzen. Tabelle 18 gibt eine Übersicht über Gefährdungen der Subsysteme und dazugehörige Gegenmaßnahmen. Tabelle 19 beschreibt während der Datenübertragung entstehende Gefährdungen und Gegenmaßnahmen. Zusätzlich zu den in Tabelle 18 und 19 dargestellten technischen Gegenmaßnahmen existieren weitere administrative Gegenmaßnahmen zum Schutz biometrischer Systeme und Daten. Siehe hierzu ITU-T X.tpp-1 [ISOe] und ISO 19092:2008 [ISOf].

Darüber hinaus besteht grundsätzlich die Gefahr des Austauschens gewisser biometrischer Subsysteme durch manipulierte Systeme (z.B. durch den physikalischen Austausch oder durch Umleiten des Netzwerkverkehrs), wogegen die Signierung biometrischer Komponenten hilfreich sein kann.

Subsystem	*Gefährdungen*	*Gegenmaßnahmen*
Erfassungs-subsystem	Täuschung des biometrischen Sensors durch künstliche, leblose biometrische Charakteristika (Sensor Spoofing mit künstlichen Plagiaten)	Lebenderkennung Multimodale Biometrie Challenge/Response Verfahren
Signalverarbeitungs-subsystem	Einfügen gefälschter Daten	Verwendung geprüfter und freigegebener Algorithmen
Vergleichssubsystem	Manipulation von Ähnlichkeitsmaßen (berechneten Vergleichswerten)	Sicherung des Servers und/oder Clients Implementierung von Vergleichs- und Entscheidungssubsystem in einer geschützten Komponente (z.B. On-Card-Comparison, „tamper resistant module“, ...)
Datenspeicher-subsystem	Kompromittierung der Datenbank	Verwendung erneuerbarer biometrischer Referenzen Datenseparation Zugriffskontrolle
	Unautorisierte Veröffentlichung personenbezogener Daten (Biometric Reference - BR, Identity Reference IR) Unautorisiertes Austauschen von gespeicherten Daten (BR, IR) Unautorisierte Modifikation von BR, IR Unautorisiertes Löschen von BR, IR	Zugriffskontrollen Sicherung von BR, IR durch elektronische Signaturen Sicherung von BR, IR durch Datenverschlüsselung
Entscheidungs-subsystem	Leichte und kontinuierliche Modifikation eines biometrischen Samples zur Erreichung der notwendigen Entscheidungsgrenzwerte (Hill-Climbing Attacke)	Verwendung grob dargestellter Vergleichswerte Sichere Kommunikationskanäle
	Erhöhung der Falsch-Akzeptanz-Rate durch Manipulation von Entscheidungsgrenzwerten (z.B. Heruntersetzen des Grenzwertes)	Zugriffskontrolle Schutz der verwendeten Daten (z.B. durch Kryptografie)

Tabelle 18: Gefährdungen biometrischer Subsysteme und dazugehörige Gegenmaßnahmen.

Kommunikationsverbindung(en)	*Übertragene Daten*	*Gefährdungen*	*Gegenmaßnahmen*
Datenerfassungs-subsystem ↔ Signalverarbeitungs-subsystem Signalverarbeitung ↔ Vergleichssubsystem	Biometrisches Sample und Merkmalsdaten	Abhören der Daten	Verwendung einer gesicherten Kommunikationsverbindung (z.B. durch Verschlüsselung)
		Wiederholte Datenübermittlung (Replay-Attacke)	Einsatz von Challenge-Response-Verfahren
		Brute Force	Ablehnen von Datenverbindungen / Fehlbedienungszähler
Datenspeichersubsystem ↔ Vergleichssubsystem	Biometrische Referenz	Abhören der Daten	Verwendung einer gesicherten Kommunikationsverbindung (z.B. durch Verschlüsselung)
		Wiederholte Datenübermittlung (Replay-Attacke)	Einsatz von Challenge-Response-Verfahren
		Man-in-the-middle Angriff	Verwendung einer gesicherten Kommunikationsverbindung und Authentikation der Teilnehmer (z.B. durch Ende-zu-Ende-Verschlüsselung) Verwendung von PKI / Zertifikaten zur Sicherung der biometrischen Daten
		Hill-Climbing Attacke	Verwendung grob dargestellter Vergleichswerte Sichere Kommunikationskanäle
Vergleichssubsystem ↔ Entscheidungssubsystem	Ähnlichkeitsmaß	Manipulation des Ähnlichkeitsmaßes	Verwendung einer gesicherten Kommunikationsverbindung (z.B. durch Verschlüsselung)

Tabelle 19: Durch Datenübertragung auftretende Gefährdungen biometrischer Systeme und dazugehörige Gegenmaßnahmen.

7.4.3 Sicherheit biometrischer Datensätze

Bei einem zur biometrischen Authentisierung notwendigen biometrischen Datensatz handelt es sich um ein Tupel bestehend aus biometrischer Referenz oder erneuerbarer biometrischer Referenz und Identitätsreferenz. Biometrische Datensätze werden während des Enrolment-Prozesses gespeichert. Eine Struktur zur Speicherung biometrischer Datensätze standardisiert ISO/IEC 19785 (Common Biometric Exchange Format Framework, CBEFF) [ISOg].

Um die Privatsphäre einer betroffenen Person zu schützen, ist es möglich, (erneuerbare) biometrische Referenzen und Identitätsreferenzen in unterschiedlichen Datenbanken zu speichern (Datenseparation). Zusätzlich erhöhen Erneuerbarkeit und Widerrufbarkeit biometrischer Referenzen darüber hinaus die Sicherheit von Datenbanken und Applikationen sowie den Schutz der Privatsphäre einer betroffenen Person. Die Möglichkeiten des Erneuerns bzw. Widerrufens biometrischer Referenzen stehen nur bei Verwendung erneuerbarer biometrischer Referenzen (RBR) zur Verfügung.

Die Komponenten eines biometrischen Datensatzes können grundsätzlich auf vier unterschiedliche Arten gespeichert werden:

- **Klartext**: Hierbei werden (erneuerbare) biometrische Referenz und/oder Identitätsreferenz unmittelbar, ohne Anwendung weiterer kryptografischer Methoden gespeichert. Das Speichern eines Datums im Klartext bietet keine Möglichkeiten der Wahrung der Vertraulichkeit bzw. der Integrität des Datums an sich.
- **Signiert**: Hierbei werden (erneuerbare) biometrische Referenz und/oder Identitätsreferenz zusammen mit einer daraus abgeleiteten Signatur bzw. einem Message Authentication Code (MAC) gespeichert. Das Speichern signierter Daten ermöglicht die Sicherstellung der Integrität der Daten.
- **Verschlüsselt**: In diesem Fall werden (erneuerbare) biometrische Referenz und/oder Identitätsreferenz verschlüsselt gespeichert. Die Verschlüsselung der Daten ermöglicht die Wahrung der Vertraulichkeit der Daten. Eine schwache Integritätskontrolle ist darüber hinaus ggf. möglich.
- **Signiert und Verschlüsselt:** (Erneuerbare) Biometrische Referenz und/oder Identitätsreferenz werden signiert und zusammen mit der Signatur verschlüsselt gespeichert. Dieser Modus ermöglicht gleichzeitig die Wahrung der Vertraulichkeit der Daten, sowie eine Integritätskontrolle.

Zur Speicherung biometrischer Datensätze wird aus Sicht des Schutzes der Privatsphäre betroffener Personen die Separierung von IR und BR empfohlen. Idealer Weise werden die IR bzw. BR enthaltenden Datenbanken von unterschiedlichen Organisationen betrieben und deren Datensätze mit unterschiedlichen Schlüsseln kryptografisch gesichert. Bei Verwendung erneuerbarer biometrischer Referenzen sollten PI und AD ebenfalls getrennt von einander gespeichert werden.

Zur Herstellung der Verbindung zwischen biometrischer Referenz und Identitätsreferenz im Falle der Datenseparierung bedarf es eines beider Seiten bekannten Identifikators (Common Identifier - CI). Abschnitt 7.7 gibt Beispiele für die Implementierung eines gemeinsamen Identifikators CI.

7.5 Anwendungsmodelle biometrischer Systeme

Dieser Abschnitt beschreibt Anwendungsmodelle biometrischer Systeme. Eine Unterscheidung der Modelle kann auf Basis des Speicherortes biometrischer Datensätze sowie des Implementierungsorts des Vergleichssubsystems getroffen werden.

Im Folgenden werden hierzu acht Modelle beschrieben, die nach aktuellem Stand der Technik in realen Anwendungen zum Einsatz kommen, sowie deren Auswirkungen in Bezug auf die Sicherheit des Systems und der verarbeiteten Daten. Die hierbei verwendeten Standorte des Datenspeichers bzw. des Vergleichssubsystems lassen sich wie folgt beschreiben:

- **Server:** Ein Server ist ein System, das über ein Netzwerk mit einem Client kommuniziert und Daten zur Verfügung stellt bzw. Operationen ausführt.
- **Client**: Bei einem Client handelt es sich um ein Computersystem oder ein äquivalentes Endgerät (z.B. PDA, Smartphone), das in Form eines Kiosksystems zur Verfügung stehen kann und an ihn angeschlossene oder integrierte biometrische Sensoren besitzt.
- **Token:** Ein Token (z.B. Smartcard, elektronischer Personalausweis) ist ein tragbares Gerät, das biometrische Daten speichern und in manchen Fällen auch vergleichen kann (z.B. On-Card-Comparison Lösung auf Basis von Smartcards).

Die im Weiteren beschriebenen Modelle A bis F sind auf den Einsatz von sowohl biometrischen Referenzen, als auch erneuerbaren biometrischen Referenzen unter der Annahme anwendbar, dass PI und AD am gleichen Ort gespeichert werden. Modelle G und H hingegen sind ausschließlich für den Einsatz mit erneuerbaren biometrischen Referenzen anwendbar und beschreiben Modelle der Separation von PI und AD. Durch die verteilte Speicherung von pseudonymen Identifikatoren (PI) und unterstützenden Daten (AD) sind die Modelle G und H ausschließlich zur Verifikation der Identität eines Individuums in einem bestimmten Kontext einsetzbar.

7.5.1 Modell A – Speichern und Vergleichen auf Server

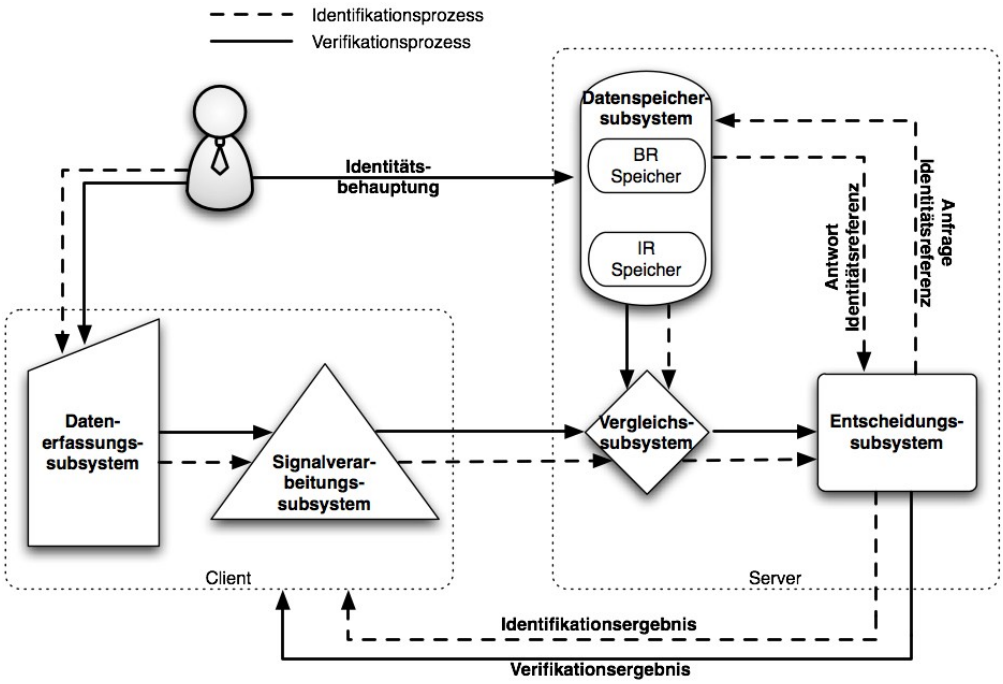

Abbildung 104: Modell A - Speichern und Vergleich auf Server unter Verwendung biometrischer Referenzen (BR).

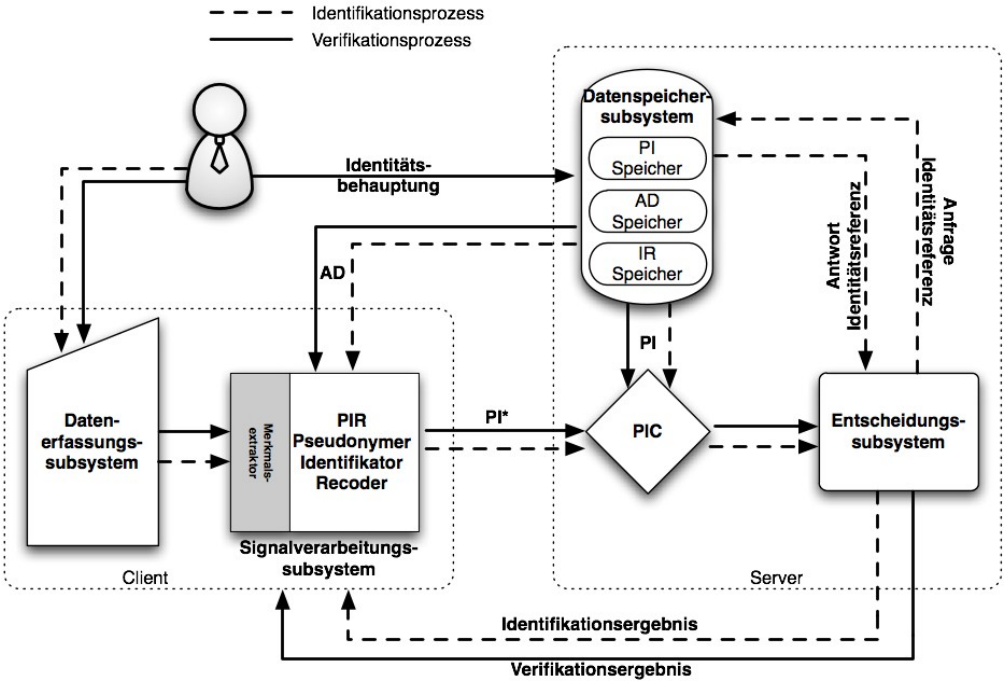

Abbildung 105: Modell A - Speichern und Vergleichen auf Server unter Verwendung erneuerbarer biometrischer Referenzen (BR).

In diesem Modell (siehe Abbildung 104) werden biometrische Referenzen während des Enrolmentprozesses auf dem Server gespeichert. Im Zuge eines Verifikations- bzw.

Identifikationsprozesses werden biometrische Samples am Client erfasst und daraus extrahierte Merkmale zum Vergleichen an den Server übermittelt.

Dieses Modell eignet sich besonders beim Einsatz von Endgeräten mit moderaten Systemressourcen und kann sowohl zur Identifikation, als auch Verifikation eingesetzt werden. Aus Sicht des Schutzes der Privatsphäre der enrolten Personen ist der Einsatz eines diesem Modell folgenden biometrischen Systems nur in Verbindung mit erneuerbaren biometrischen Referenzen empfohlen. Sensible personenbezogene Daten (biometrische Referenz, Identitätsreferenz) werden gemeinsam in einem zentralen Speicher auf einem Server vorgehalten und erfordern hierdurch hinreichenden Schutz der Datenbank. Zusätzlich bedarf es eines sicheren Kommunikationskanals sowie eines Vertrauensverhältnisses zwischen Client und Server.

Automatische Fingerabdruck-Identifikationssysteme (AFIS) werden häufig diesem Modell folgend implementiert.

7.5.2 Modell B – Speichern auf Token, Vergleich auf Server

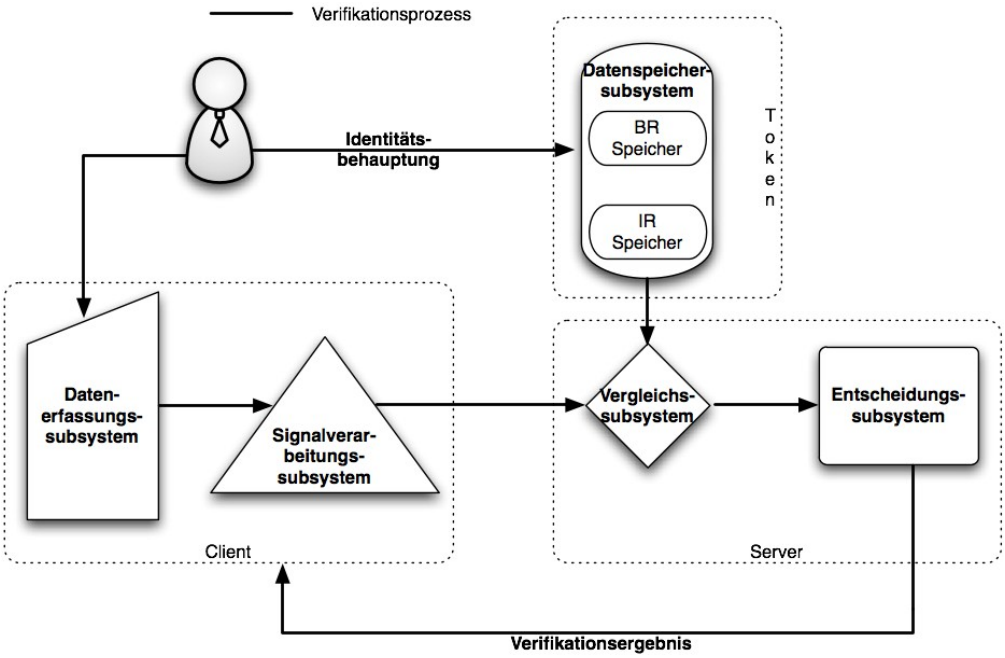

Abbildung 106: Modell B - Speichern auf Token, Vergleich auf Server unter Verwendung biometrischer Referenzen (BR).

Dieses Modell (siehe Abbildung 106) sieht das Speichern biometrischer Datensätze auf einem benutzerspezifischen Token während des Enrolmentprozesses vor. Analog zu Modell A findet die Datenerfassung während einer biometrischen Authentisierung am Client statt, sodass gespeicherte biometrische Referenzen zusammen mit von einem erfassten biometrischen Sample extrahierten Merkmalen zum Vergleichen über das Netzwerk an einen Server übertragen werden müssen. Im Falle der Verwendung erneuerbarer biometrischer Referenzen übernimmt der Client das Erzeugen eines neuen pseudonymen Identifikators PI* (PIR Ansatz, siehe Kapitel 2.4.3) und übermittelt diesen gemeinsam mit dem auf dem Token gespeicherten pseudonymen Identifikator PI an den

Server. Die zur Erzeugung verwendeten unterstützenden Daten AD werden nicht über das Netzwerk übertragen.

Dieses Modell setzt serverseitiges Vertrauen der vom Client erhaltenen Daten voraus und wird im Wesentlichen zur Verifikation der Identität eines Individuums eingesetzt. Es wird ein hinreichendes Maß an Netzwerksicherheit benötigt, um den Datentransfer zwischen Client und Server zu schützen und gewährleisten zu können, dass die übermittelten Daten originär vom zu verifizierenden Individuum stammen und nicht im Rahmen des Kommunikationsvorgangs modifiziert oder ausgetauscht wurden.

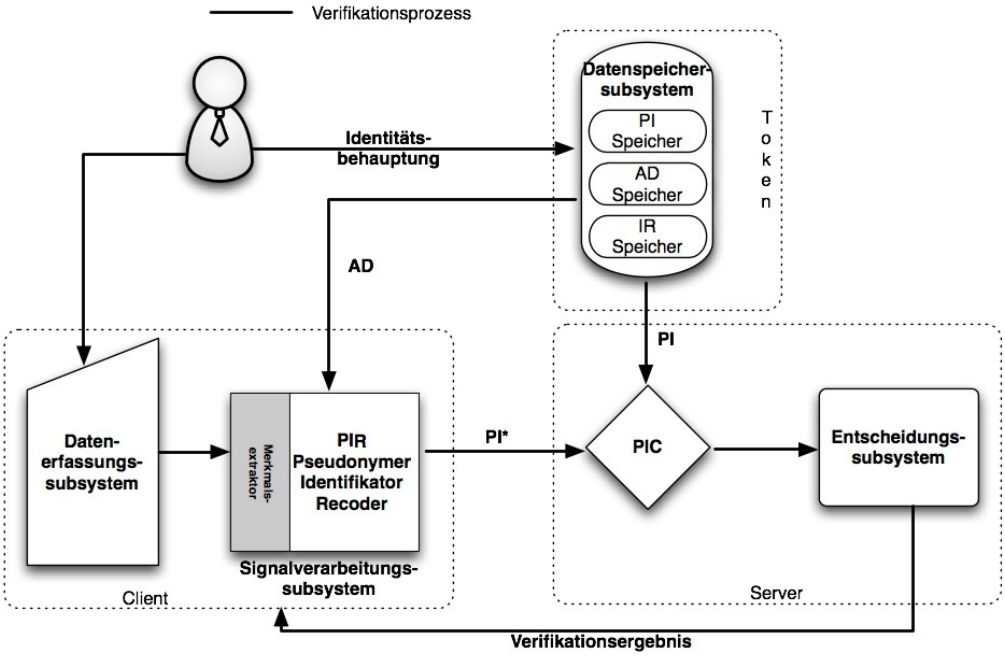

Abbildung 107: Modell B - Speichern auf Token, Vergleich auf Server unter Verwendung erneuerbarer biometrischer Referenzen (RBR).

Da Identitätsreferenz und biometrische Referenz auf einem benutzerspezifischen Token gespeichert werden, bedarf es keiner speziellen Maßnahmen zur Sicherung von Datenbanken. Darüber hinaus gibt es weder Verknüpfungen zwischen biometrischer Referenz und Identitätsreferenz, die für eine datenbankübergreifende Verknüpfung genutzt werden könnten, noch findet eine Übermittlung der Identitätsreferenz an den Server statt. Aus Sicht der Privatsphäre der erfassten Personen handelt es sich daher um ein günstiges Modell.

7.5.3 Modell C – Speichern auf Server, Vergleich auf Client

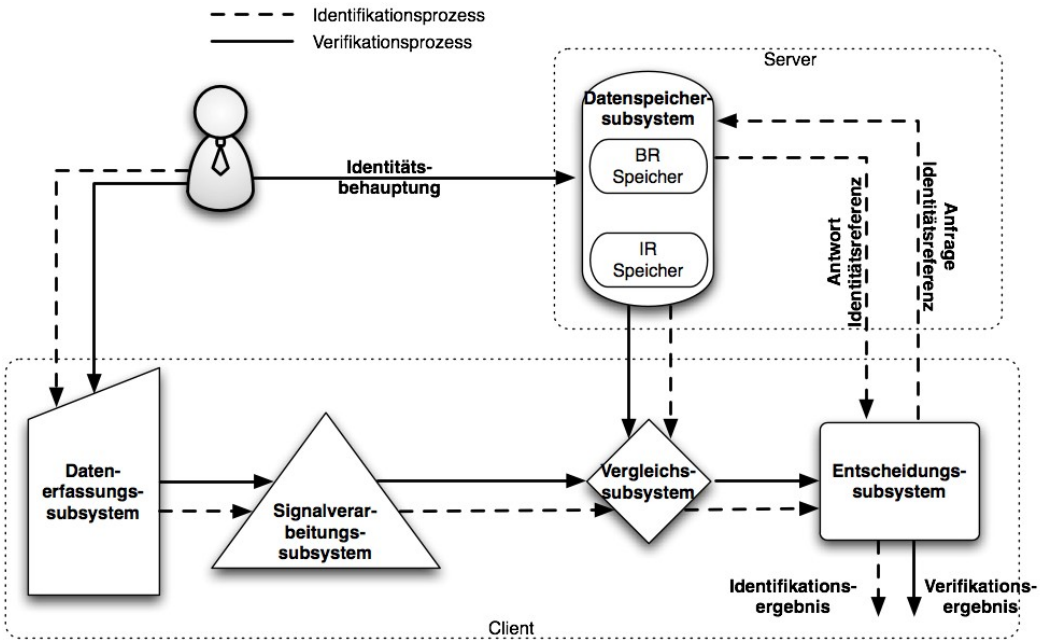

Abbildung 108: Modell C – Speichern auf Server, Vergleich auf Client unter Verwendung biometrischer Referenzen (BR).

Modell C (siehe Abbildung 108) sieht das Speichern biometrischer Datensätze während des Enrolmentprozesses auf einem Server und das Vergleichen während eines Authentisierungsvorgangs auf dem Client vor. Hierzu erfasst der Client durch an ihn angeschlossene biometrische Sensoren ein biometrisches Erkennungssample der betroffenen Person und extrahiert benötigte Merkmalsdaten.

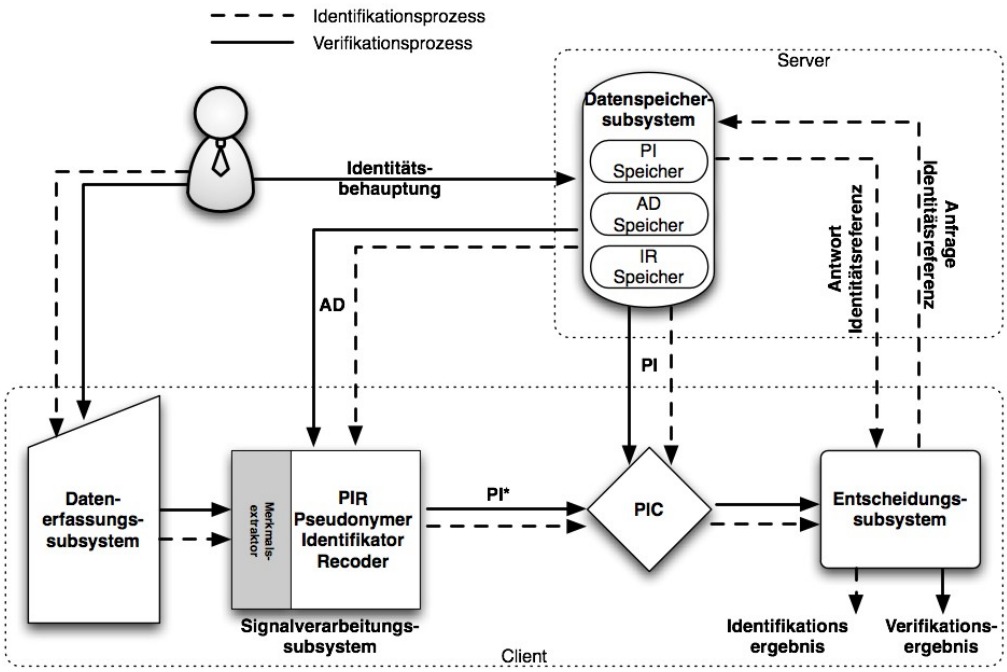

Abbildung 109: Modell C - Speichern auf Server, Vergleich auf Client unter Verwendung erneuerbarer biometrischer Referenzen (RBR).

Im Falle eines Verifikationsprozesses fordert der Client einen zum Individuum gehörenden biometrische Datensatz an und vergleicht die biometrische Referenz mit den auf dem Client extrahierten Merkmalsdaten. Im Falle eines Identifikationsprozesses bedarf es der Übermittlung mehrere biometrischer Referenzen und Identitätsreferenzen von Server zu Client.

Auf Grund der Datenübertragung bedarf dieses Modell eines hinreichenden Maßes an Netzwerksicherheit. Darüber hinaus ist eine entsprechende Sicherung des Datenspeichersubsystems auf Grund des zentralen Speicherns sensibler personenbezogener Daten notwendig, um die Privatsphäre der erfassten Personen zu schützen. Die Verwendung erneuerbarer biometrischer Referenzen in diesem Modell wird daher empfohlen.

7.5.4 Modell D – Speichern und Vergleich auf Client

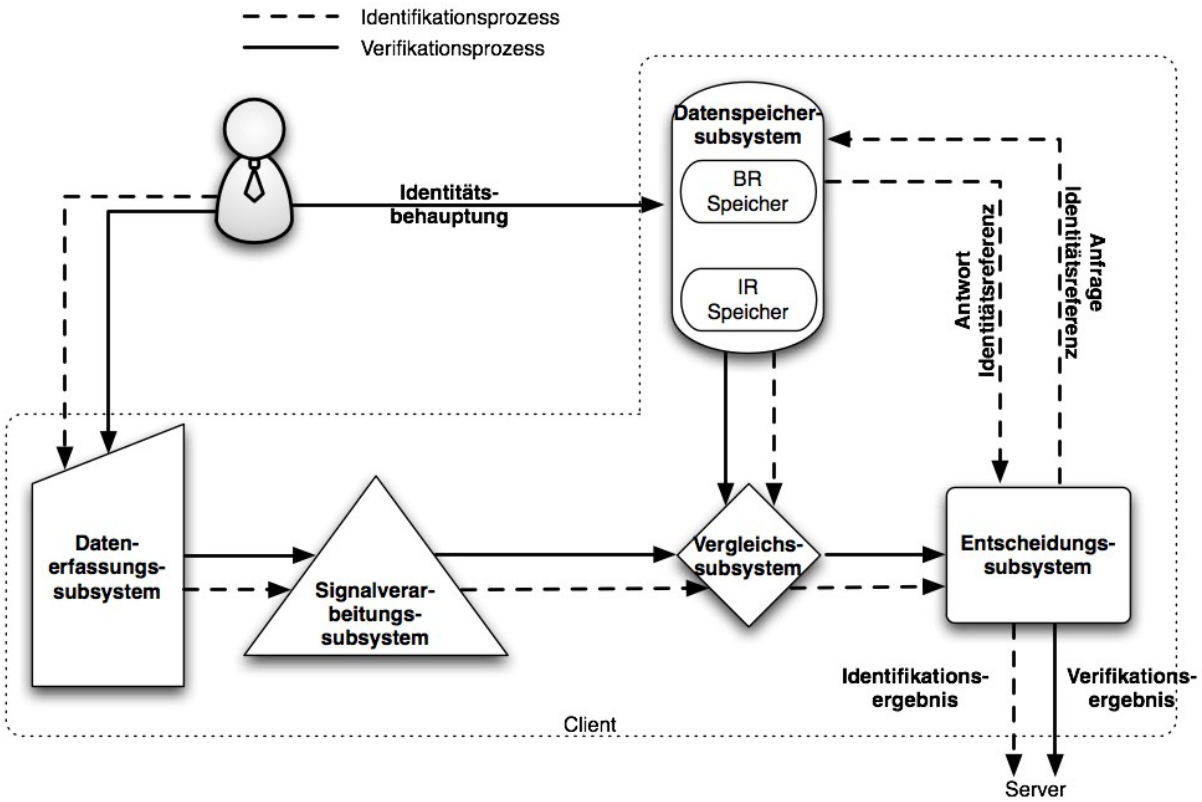

Abbildung 110: Modell D - Speichern und Vergleich auf Client unter Verwendung biometrischer Referenzen (BR).

In diesem Modell (siehe Abbildung 110) werden biometrische Datensätze während des Enrolments auf dem Client gespeichert. Während eines Verifikations- bzw. Identifikationsprozesses erfasst der Client über angeschlossene Sensoren ein biometrisches Erkennungssample der betroffenen Person, das zum Vergleich mit der biometrischen Referenz dient.

Dieses Modell kann vollständig eigenständig und isoliert betrieben werden – hierbei übernimmt der Client alle Teilaufgaben. Alternativ kann die schlussendliche Entscheidung an einen Server delegiert werden. Hierzu muss das Entscheidungssubsystem ausgelagert und das errechnete Ähnlichkeitsmaß an den Server übermittelt werden.

Modell D kann sowohl zur Identifikation eines Individuums, als auch zur Verifikation der vorgegebenen Identität eines Individuums genutzt werden und wird häufig zur lokalen Authentisierung von Nutzern von mobilen Endgeräten eingesetzt.

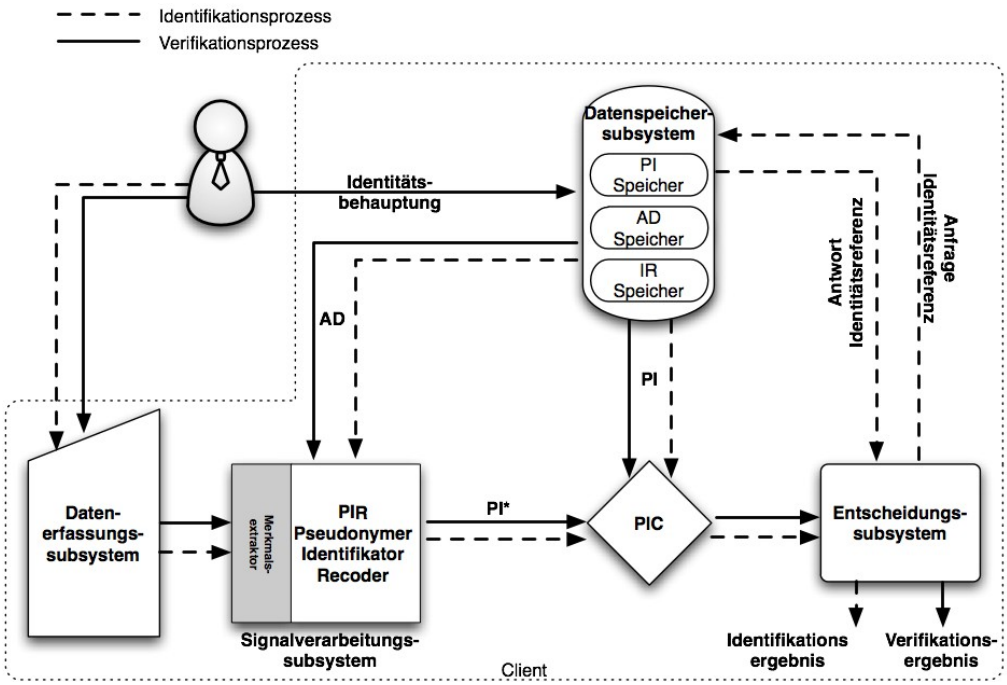

Abbildung 111: Modell D - Speichern und Vergleich auf Client unter Verwendung erneuerbarer biometrischer Referenzen (RBR).

Da in diesem Modell die Übermittlung sensitiver personenbezogener Daten nicht vorgesehen ist, spielt Netzwerksicherheit keine bedeutende Rolle für die Sicherheit des biometrischen Systems. Dennoch stellt dieses Modell Anforderungen an die Sicherheit des clientseitig implementierten Datenspeichersubsystems und der Einsatz erneuerbarer biometrischer Referenzen wird empfohlen.

7.5.5 Modell E – Speichern auf Token, Vergleich auf Client

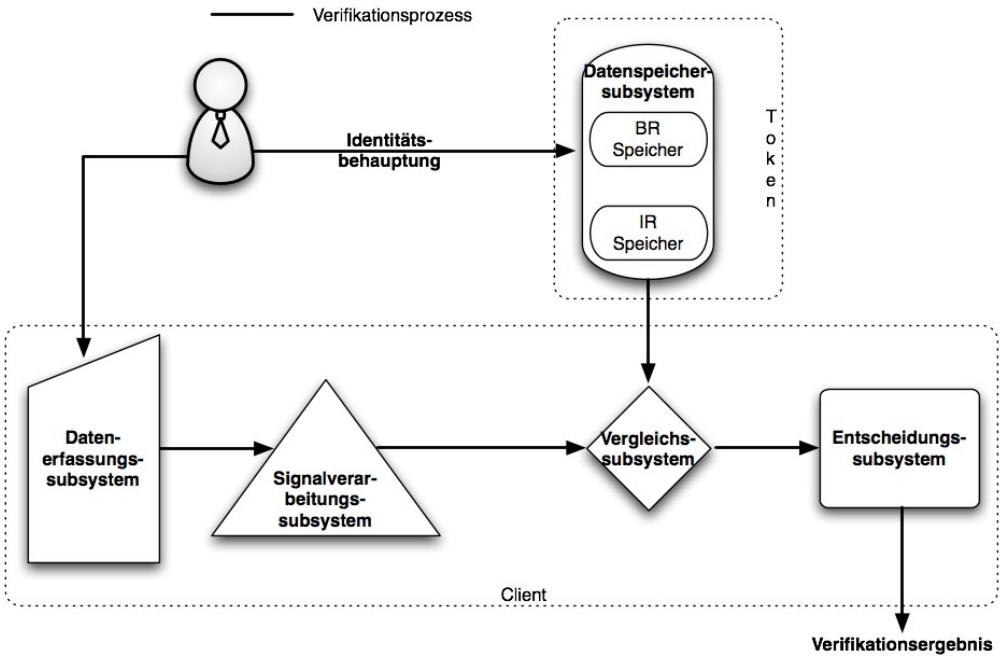

Abbildung 112: Modell E - Speichern auf Token, Vergleich auf Client unter Verwendung biometrischer Referenzen (BR).

Modell E (siehe Abbildung 112) sieht das Speichern biometrischer Datensätze während des Enrolmentprozesses auf einem Token vor. Der Vergleich der gespeicherten biometrischen Referenz mit einem erfassten biometrischen Erkennungssample einer betroffenen Person findet auf dem Client statt.

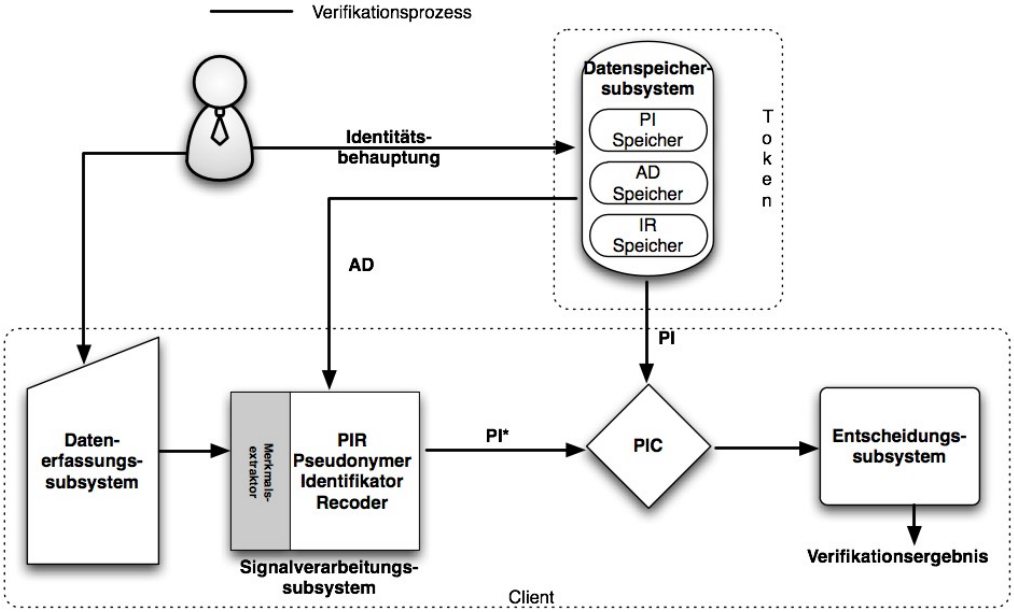

Abbildung 113: Modell E - Speichern auf Token, Vergleich auf Client unter Verwendung erneuerbarer biometrischer Referenzen (RBR).

In dem hier beschriebenen Modell werden keine sensiblen personenbezogenen Daten an einen Server oder über Netzwerkverbindungen ausgetauscht, sodass keine speziellen Anforderungen an die Netzwerksicherheit gestellt werden. Die Sicherheit der auf dem Token gespeicherten biometrischen Datensätze muss jedoch durch entsprechende Maßnahmen gewährleistet sein. Darüber hinaus sollte die Kommunikation zwischen Client und Token unter Verwendung des Secure Messaging Mechanismus (ISO/IEC 7816-4) gesichert werden.

Zum Einsatz kommt dieses Modell zum Beispiel bei der Grenzkontrolle. Hierbei wird ein elektronischer Personalausweis oder Reisepass als Token verwendet während sich der Client in Form eines Kiosks in einer Grenzkontrollstation befindet. Biometrische Referenz und Identitätsreferenz werden auf dem integrierten Chip des elektronischen Personaldokuments gespeichert.

7.5.6 Modell F – Speichern und Vergleich auf Token

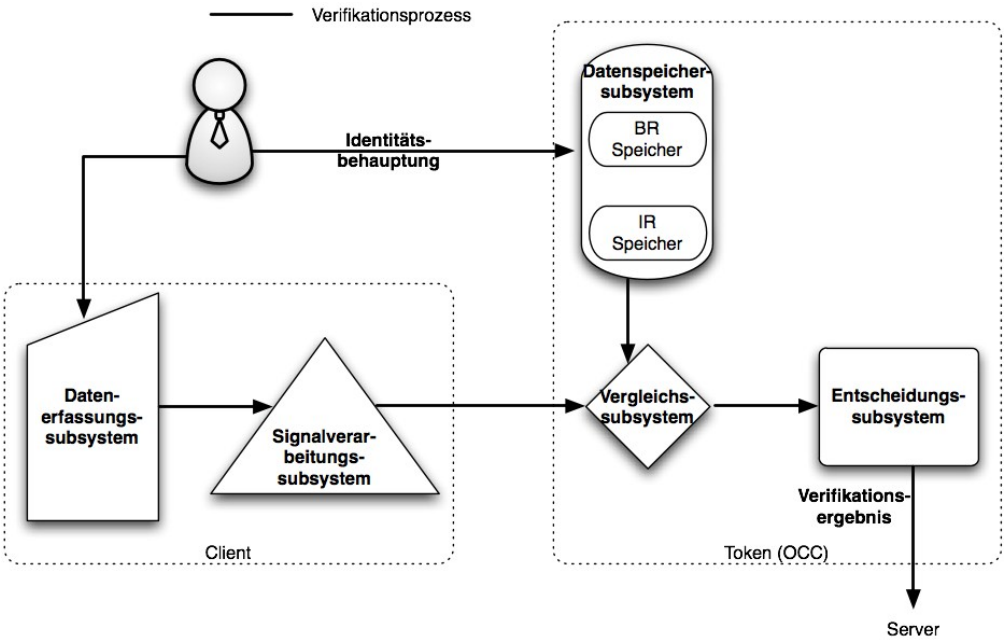

Abbildung 114: Modell F - Speichern und Vergleich auf Token (OCC) unter Verwendung biometrischer Referenzen (BR).

Während der Enrolmentphase werden in diesem Modell (siehe Abbildung 114) biometrische Datensätze auf einem benutzerspezifischen Token gespeichert. Darüber hinaus verfügt das Token über das Vergleichs- und Entscheidungssubsystem und ist somit in der Lage, die biometrische Authentisierung eigenständig durchzuführen. Das hierfür notwendige biometrische Erkennungssample wird vom Client erfasst und an das Token übermittelt. Im Anschluss wird das Ergebnis des Entscheidungsprozesses an den Server übermittelt.

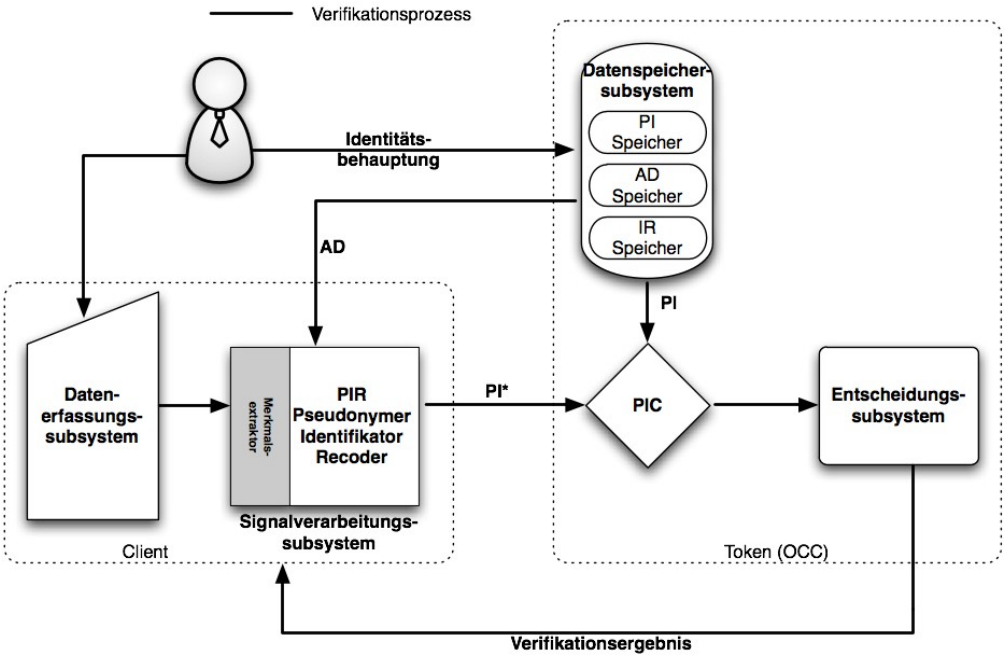

Abbildung 115: Modell F - Speichern und Vergleich auf Token (OCC) unter Verwendung erneuerbarer biometrischer Referenzen (RBR).

Diese Art eines On-Card-Comparison Modells stellt den stärksten Schutzmechanismus personenbezogener Daten dar, da keinerlei sensible personenbezogene Daten herausgegeben werden. Sollte dieses Modell zusammen mit erneuerbaren biometrischen Referenzen verwendet werden, ist lediglich die Übermittlung unterstützender Daten an den Client zur Generierung eines PI* (PIR Ansatz) notwendig. Zum Schutz der Kommunikation zwischen Client und Token sollte der in ISO/IEC 7816-4 beschriebene Secure Messaging Mechanismus verwendet werden.

Implementationsaspekte derartiger, Modell F folgender, kartenbasierter Lösungen werden durch ISO/IEC 24789 (On-Card-Comparison) standardisiert.

7.5.7 Modell G – Verteiltes Speichern auf Token und Server, Vergleich auf Server

Modell G (siehe Abbildung 116) bezieht sich ausschließlich auf den Einsatz mit erneuerbaren biometrischen Referenzen und verfolgt das Konzept von Datenseparierung durch verteiltes Speichern der Komponenten erneuerbarer biometrischer Referenzen (IR, PI, AD, ggf. CI).

Ein pseudonymer Identifikator (PI) wird zusammen mit einem gemeinsamen Identifikator (CI) während der Enrolmentphase in einem sich auf dem Server befindenden Datenspeicher gespeichert. Die hierzu gehörenden unterstützenden Daten (AD) werden jedoch zusammen mit der Identitätsreferenz (IR) des Individuums und dem gemeinsamen Identifikator (CI) auf einem benutzerspezifischen Token gespeichert.

Während eines Verifikationsprozesses werden CI und AD dem Client in Form des Tokens präsentiert. Auf Basis der AD und eines gewonnenen biometrischen Erkennungssamples bestimmt der Client einen pseudonymen Identifikator PI*, der zusammen mit CI an den Server übertragen

und zum Vergleichen mit dem auf dem Server gespeicherten pseudonymen Identifikator PI genutzt wird.

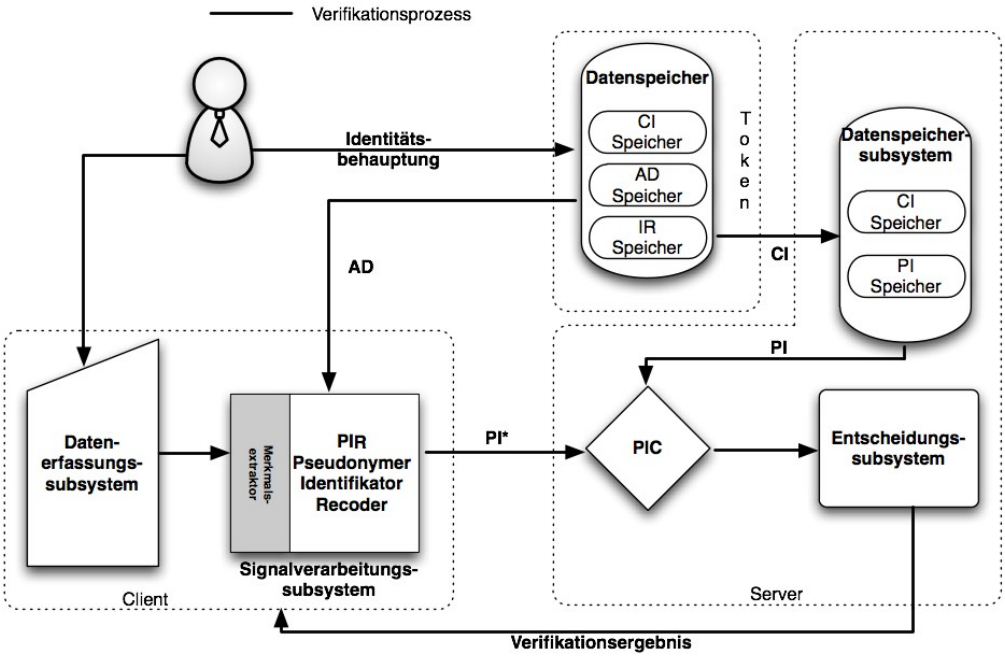

Abbildung 116: Modell G – Verteiltes Speichern auf Token und Server, Vergleich auf Server unter Verwendung erneuerbarer biometrischer Referenzen (RBR).

Ein bedeutender Vorteil dieses Modells ist die Verteilung der erneuerbaren biometrischen Referenz auf unterschiedliche Systeme (Token, Server), wodurch die Durchführung eines Verifikationsprozesses lediglich durch Vorliegen korrekter Daten auf beiden Systemen sowie ausschließlich durch Zustimmen der betroffenen Person ermöglicht wird. Darüber hinaus ermöglicht dieses Modell ein serverseitiges Widerrufen biometrischer Referenzdaten (PI), ohne hierzu Zugriff auf das Token zu benötigen.

Abwandlungen dieses Modells sind möglich in der Form

- des Speicherns der Identitätsreferenz auf dem Server anstatt auf dem Token
- des Speicherns von CI, IR und AD auf dem Client und PI und CI auf dem Server – ohne die Verwendung eines Tokens
- des Speicherns von PI auf sowohl dem Server, als auch dem Token um hierdurch die Möglichkeit einer 3-Faktor Authentisierung zu erreichen.

Modell G eignet sich besonders für die Tätigung sicherer online Transaktionen, wie zum Beispiel online Banking, online Kreditkartentransaktionen und zur Verwendung als PIN-Ersatz an Geldautomaten.

7.5.8 Modell H – Verteiltes Speichern auf Token und Client, Vergleich auf Client

Vergleichbar zu Modell G basiert Modell H (siehe Abbildung 117) auf dem Ansatz des verteilten Speicherns von IR, AD und CI auf einem Token und PI und CI auf dem Client. Bei Bedarf ist das Speichern von IR auf dem Client anstelle des Tokens ebenfalls möglich.

Zur Verifikation der vorgegebenen Identität eines Individuums übermittelt das Token CI und AD an den Client. Der Client ermittelt den zu CI gehörenden pseudonymen Identifikator (PI) und führt PI, AD sowie ein erfasstes biometrisches Erkennungssample dem auf dem Client implementierten PI Verifier (PIV) zu (vgl. Abschnitt 2.4), der die erhaltenen Daten vergleicht und ein Verifikationsergebnis ausgibt.

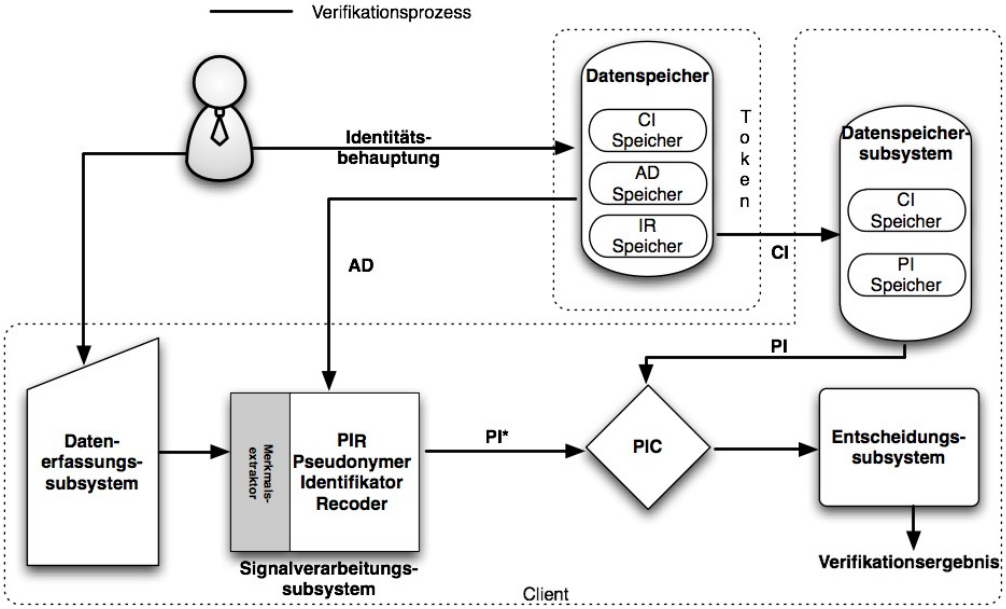

Abbildung 117: Modell H - Verteiltes Speichern auf Token und Client, Vergleich auf Client unter Verwendung erneuerbarer biometrischer Referenzen (RBR).

In diesem Modell kann der Client eine eigenständige funktionale Einheit darstellen, die zur biometrischen Authentisierung an Flughäfen, in öffentlichen Gebäuden oder in der Grenzkontrolle (z.B. unter Verwendung eines elektronischen Personaldokuments als Token) genutzt werden kann.

7.5.9 Auswirkungen der Bedrohungen auf die Modelle A bis H

In diesem Abschnitte wird die Tabelle 18 in einer veränderten Form gezeigt und die Auswirkungen von Gefährdungen auf die primär betroffenen Modelle A bis H dargestellt.

Subsystem	***Gefährdungen***	***Primär betroffene Modelle***
Erfassungs-subsystem	Täuschung des biometrischen Sensors durch künstliche, leblose biometrische Charakteristika (Sensor Spoofing mit künstlichen Plagiaten)	Modell A,B,C,D,E,F,G,H
Signalverarbeitungs-subsystem	Einfügen gefälschter Daten	Modell A,B,C,D,E,F,G,H

Subsystem	*Gefährdungen*	*Primär betroffene Modelle*
Vergleichssubsystem	Manipulation von Ähnlichkeitsmaßen	Modell C,D,E,H
Datenspeicher-subsystem	Kompromittierung der Datenbank	Modell A,C
	Unautorisierte Veröffentlichung personenbezogener Daten (Biometric Reference - BR, Identity Reference IR)	Modell A,C
	Unautorisiertes Austauschen von gespeicherten Daten (BR, IR)	Modell D
	Unautorisierte Modifikation von BR, IR	Modell D
	Unautorisiertes Löschen von BR, IR	Modell A,C
Entscheidungs-subsystem	Leichte und kontinuierliche Modifikation eines biometrischen Samples / Manipulation von Entscheidungsgrenzwerten	Modell C,D,E,H

Tabelle 20: Auswirkung von Gefährdungen auf einzelne Modelle biometrischer Systeme.

7.6 Privatsphärenmanagement biometrischer Informationen

Auf Grund der zunehmenden Verbreitung biometrischer Systeme auf lokaler, nationaler und internationaler Ebene wird dem Schutz der Privatsphäre von biometrischen Systemen betroffener Personen eine steigende Bedeutung zugemessen, die sich nicht zuletzt in dementsprechender Gesetzgebung und Regulierung manifestiert. Nationale Gesetzgebung sowie regulatorische Auflagen sind daher grundsätzlich beim Betrieb biometrischer Systeme und dem Umgang mit biometrischen Daten zu beachten.

Der Begriff der Privatsphäre im Kontext von Biometrie lässt sich sowohl unter dem Aspekt der persönlichen Privatsphäre, als auch unter dem einer informationellen Selbstbestimmung betrachten. Die informationelle Selbstbestimmung bezieht sich auf den Umgang mit biometrischen Daten während derer Verarbeitung und Lebenszyklus, während gesellschaftliche, kulturelle und ethische Aspekte (z.B. Angst und Vorbehalte gegenüber der Nutzung biometrischer Systeme) im Fokus der persönlichen Privatsphäre einer von biometrischer Verarbeitung betroffenen Person stehen. Beide Typen sind im Kontext der Privatsphäre sind eng miteinander verbunden und voneinander abhängig. Unter dem Aspekt des Schutzes der Privatsphäre von betroffenen Personen stellen sich somit die folgenden Anforderungen an den Umgang mit biometrischen Informationen:

- **Erneuerbarkeit**: Auf Grund der engen Verbindung zwischen mit biometrischen Daten verknüpften Identitätsreferenzen und biometrischen Referenzen und der daraus resultierenden Kompromittierung einer Identitätsreferenz bei Kompromittierung einer biometrischen Referenz sollten biometrische Referenzen grundsätzlich erneuerbar sein.

- **Widerrufbarkeit**: Auf Grund der Möglichkeit der Kompromittierung einer biometrischen Referenz und den daraus resultierenden Sicherheitsrisiken oder der Möglichkeit des Ablaufens der Gültigkeit einer biometrischen Referenz in einem bestimmten Kontext, sollten biometrische Referenzen widerrufbar sein.
- **Nichtverknüpfbarkeit**: Die grundsätzliche Eignung biometrischer Referenzen als UUID und die daraus entstehende Möglichkeit des eindeutigen Verknüpfens sensibler Informationen über Anwendungsgrenzen hinweg bedroht die Privatsphäre von Biometrie betroffener Personen. Aus diesem Grund sollten biometrische Referenzen durch Verwendung diversifizierbarer biometrischer Referenzen, Verschlüsselung und/oder Datenseparierung nicht verknüpfbar gemacht werden.
- **Unumkehrbarkeit**: Biometrische Daten dürfen nur in im Voraus definierten Anwendungsszenariosn verwendet werden. Eine Erweiterung des Anwendungsrahmens setzt die Einwilligung des Besitzers der biometrischen Informationen voraus. Um unerlaubte Verwendung in anderen Anwendungsszenarios, zu anderen Zwecken als Identifikation und Verifikation und insbesondere die Inferenz von zum Besitzer biometrischer Referenzen gehörender Daten (medizinische Informationen, ethnische Eigenschaften, etc.) zu vermeiden, sollten biometrische Referenzen unumkehrbar sein.
- **Datenminimierung**: Da es sich bei biometrischen Daten um (sensible) personenbezogene Daten handelt, sollte ISO/IEC 29100 als Richtlinie zum Umgang mit diesen Daten angewandt werden. Darüber hinaus sollte Datenseparierung zur Minimierung der an einem zentralen Ort vorgehaltenen Daten implementiert werden.

Darüber hinaus sollten die folgenden Aspekte in den jeweiligen Abschnitten des Verarbeitungs- bzw. Lebenszyklus biometrischer Daten beachtet werden:

- **Datenerfassung**: Vor der Erfassung biometrischer Daten sollte das Einverständnis der betroffenen Person eingeholt werden. Hierzu sollten dem Individuum mindestens Art und Umfang, Verwendungs- und Verarbeitungszweck der biometrischen Daten sowie Informationen über für die Datenerfassung verantwortliche Personen mitgeteilt werden. Ferner sollte der betroffenen Person mitgeteilt werden, ob die Nutzung eines biometrischen Systems freiwillig oder verpflichtend ist und ob etwaige alternative Identifikations- und/oder Verifikationsmethoden existieren.
- **Übermittlung biometrischer Daten**: Im Falle der Übermittlung biometrischer Daten an Dritte sollte die Zustimmung der betroffenen Person eingeholt werden, sofern eine Übermittlung nicht gesetzlich notwendig ist. Hierzu sollten dem Individuum Informationen über Empfänger der biometrischen Daten, Art der übermittelten Daten sowie deren Verwendungszweck mitgeteilt werden. Besondere Sorgfalt sollte auf Grund möglicherweise unterschiedlicher nationaler Gesetzgebung in einer grenzüberschreitenden Datenübermittlung gelten.
- **Datennutzung**: Eine Nutzung biometrischer Daten sollte nur im vom Individuum genehmigten Rahmen geschehen, sofern keine abweichende gesetzliche Notwendigkeit und Legitimation besteht. Eine Erweiterung des Nutzungsrahmens bedarf der Zustimmung der betroffenen Person.
- **Datenspeicherung**: Auf Grund des Speicherns (sensibler) personenbezogener Daten sollten die abgelegten biometrischen Daten nicht mit weiteren vom Individuum erfassten personenbezogenen Daten in Verbindung zu bringen sein. Ferner sollten die im Kapitel 3

beschriebenen Sicherheitsmaßnahmen implementiert werden. Das Speichern von unverarbeiteten biometrischen Samples (Roh-Bilddaten) sollte vermieden werden.

- **Archivierung**: Archivierung biometrischer Daten sollte falls nicht zwingend erforderlich vermieden werden. Falls erforderlich sollte eine Archivierung biometrischer Daten unter Verwendung starker kryptografischer Methoden stattfinden.
- **Löschen der Daten**: Nach Ablauf der Gültigkeit biometrischer Daten, nach Ende deren Verwendungszwecks oder im Falle des Widerrufens der Zustimmung der betroffenen Person sollten alle erfassten biometrischen Daten bzw. daraus abgeleiteten Informationen sicher gelöscht werden. Hierzu sollte der Datenspeicher formatiert bzw. ein biometrischer Datensatz überschrieben werden.

Die Verantwortung zur Wahrung des Schutzes der Privatsphäre einer betroffenen Person während des Umgangs mit biometrischen Daten obliegt dem Betreiber des biometrischen Systems. Hierzu sollte der Betreiber eine betroffene Person in den Prozess der biometrischen Datenerfassung und Verarbeitung einbeziehen, mindestens jedoch seine Zustimmung einholen, Möglichkeit zur Einsichtnahme der gespeicherten Daten bieten sowie einen Mechanismus zum Widerrufen einer Zustimmung bereitstellen und die oben beschriebenen Anforderungen und Aspekte umsetzen. Ferner sollte der Betreiber die Aktualität der biometrischen Daten sicherstellen und auf Änderungswünsche der betroffenen Person reagieren. Bei einer möglichen Korrumpierung der von einem Individuum gespeicherten Daten sollte die betroffene Person umgehend informiert und entsprechende Gegenmaßnahmen eingeleitet werden.

7.7 Sichere Verknüpfung von Datensätzen in separierten IR-/BR-Datenbanken

Zur Erhöhung des Schutzes sensibler personenbezogener Daten sowie zur Sicherstellung der Privatsphäre der von Biometrie betroffenen Personen sollte eine separierte Speicherung von Identitätsreferenzen und biometrischen Referenzen vorgesehen werden. Zur Verknüpfung der IR-/BR-Datensätze über unterschiedliche Datenbanken hinweg wird ein gemeinsamer Identifikator (CI) genutzt, der keine Rückschlüsse über die miteinander verknüpften Daten zulassen sollte und darüber hinaus Aufschluss über erlaubte oder unerlaubte Modifikation von Datensätzen einer Datenbank geben sollte.

Zur Erreichung dieser Ziele und Anforderungen empfiehlt sich die Generierung eines gemeinsamen Identifikators CI unter Verwendung kryptografischer Algorithmen. Hierzu wird während der Enrolmentphase eine übermittelte Identitätsreferenz (möglicher Weise zusammen mit weiteren Attributen) unter Verwendung eines lokalen geheimen Schlüssels sIR verschlüsselt. Analog hierzu wird die einer BR-Datenbank übermittelte biometrische Referenz im Rahmen des Enrolments mit einem zweiten lokalen geheimen Schlüssel sBR chiffriert. Zusätzlich zur Verschlüsselung wird auf der IR-Datenbank ein zur Identitätsreferenz (sowie möglicher Weise weiteren Attributen) gehörender Hashwert h(IR) errechnet. Dieser Hashwert wird in Abhängigkeit von der Sicherheit (Vertraulichkeit, Authentizität) der Kommunikationsverbindung mit einem beiden Datenbanken bekannten geheimen Schlüssel sÜ verschlüsselt und an DBBR übertragen. Nach Erhalt der Daten entschlüsselt DBBR die Nachricht mit dem geheimen Schlüssel sÜ und erhält somit den Hashwert h(IR) der Identitätsreferenz IR. Zur Erzeugung eines gemeinsamen Identifikators CI bestimmt DBBR den Hashwert der biometrischen Referenz h(BR). Die daraufhin durch Verschlüsselung des Paares (h(IR), h(BR)) mit dem DBIR und DBBR bekannten Schlüssel sCI entstandene Bitfolge

stellt den gemeinsamen Identifikator CI dar. DBBR sendet im Anschluss den mit sÜ verschlüsselten Hashwert h(BR) an DBIR, woraufhin DBIR nach Entschlüsselung mit sÜ ebenfalls unter Verwendung von sCI den gemeinsamen Identifikator aus (h(IR), h(BR)) errechnen kann. Nach Errechnung von CI auf beiden Seiten wird CI zusammen mit den mit sIR bzw. sBR verschlüsselten Referenzen IR bzw. BR in DBIR bzw. DBBR gespeichert.

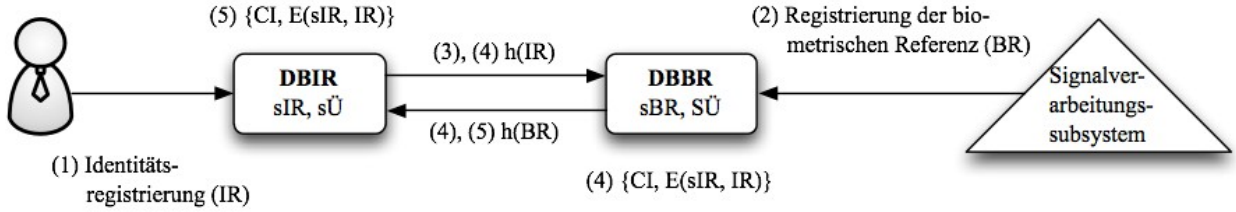

Abbildung 118: Separate IR- und BR-Datenbanken

Im Rahmen eines Verifikations- bzw. Identifikationsprozesses kann nun der so gewonnene gemeinsame Identifikator CI zusammen mit den Hashwerten der Referenzen, d.h. h(IR) bzw. h(BR), zwischen den Systemen ausgetauscht werden, um eine sichere Verknüpfung zwischen IR und BR herstellen zu können.

7.8 Kryptografische Algorithmen zur Sicherung biometrischer Systeme

Zur Erlangung von Vertraulichkeit können kryptografische Methoden zum Verschlüsseln der gespeicherten und übermittelten Daten verwendet werden. Die Arten der zum Einsatz kommenden Algorithmen lassen sich im Wesentlichen in symmetrische und asymmetrische Verschlüsselungsverfahren unterscheiden. Symmetrische Verschlüsselungsalgorithmen basieren auf einem den miteinander kommunizierenden Parteien bekannten geheimen Schlüssel (shared secret, secret key), während bei asymmetrischen Verschlüsselungsverfahren jeder Kommunikationsteilnehmer über ein aus öffentlichem Schlüssel (public key) und privatem Schlüssel (private key) bestehendes Schlüsselpaar verfügt. Symmetrische Verfahren werden unter anderem aus Performanzgründen häufiger im Kontext biometrischer Systeme eingesetzt als asymmetrische Verfahren. Weitere Informationen zu symmetrischen Verfahren können ISO/IEC 18033-3 [ISOk] und ISO/IEC 18033-4 [ISOl] entnommen werden. Informationen zu asymmetrischen Verfahren finden sich in ISO/IEC 18033-2 [ISOj].

Zur Wahrung der Datenintegrität in biometrischen Systemen eignet sich der Einsatz von Message Authentication Code (MAC) Algorithmen oder digitaler Signaturen. ISO/IEC 9797-1 und ISO/IEC 9797-2 [ISOo] beschreiben Arbeitsweisen von MAC Algorithmen. Digitale Signaturen können in zwei Modi eingesetzt werden. ISO/IEC 9796 [ISOn] beschreibt ein Schema, das die Extraktion einer ganzen Nachricht oder ein Teil derer aus der digitalen Signatur ermöglicht, wohingegen ISO/IEC 14888 [ISOm] ein Schema beschreibt, das keine Datenextraktion ermöglicht. Für die Verwendung digitaler Signaturen empfiehlt sich der Einsatz einer Public-Key Infrasktruktur (PKI).

Zur Wahrung von Integrität und Erlangung von Vertraulichkeit kann eine Kombination von Datenverschlüsselung und MAC Algorithmen bzw. digitaler Signatur eingesetzt werden. Siehe hierzu ISO/IEC 19772 [ISOp].

8 Literaturverzeichnis

17	ISO/IEC JTC 1 SC 17. Application of biometrics to cards and personal identification.
37	ISO/IEC JTC 1 SC 37. Biometrics.
AC09	A. Adler and R. Cappell:Template SecurityEncyclopedia of Biometrics, 2009
AC93	Ahlswede, R. and Csiszar, I.:Common randomness in information theory and cryptography in secret sharing
Adler05	Andy Adler:Vulnerabilities in biometric encryption systemsAudio- and Video-based Biometric Person Auth., 2005
AI	ANSI/NIST-ITL. American national standards for biometrics. http://fingerprint.nist.gov/standard/.
BCI+07	Julien Bringer, Hervé Chabanne, Malika Izabachène, David Pointcheval, Qiang Tang, Sébastien Zimmer: An application of the goldwasser-micali cryptosystem to biometric authentication. In ACISP, pages 96–106, 2007.
BCPT07	Julien Bringer, Hervé Chabanne, David Pointcheval, Qiang Tang: Extended private information retrieval and its application in biometrics authentications. In CANS, pages 175–193, 2007.
BCR04	R. Bolle and J. H. Connell and N.K. RathaSystem and method for distorting a biometric for transactions with enhanced security and privacy
BD08	Ileana Buhan and Jeroen Doumen and Pieter Hartel and Qiang Tang and Raymond Veldhuis:Embedding renewable cryptographic keys into continuous noisy dataInformation and communications security, 10th international conference ICICS, 2008
Ber08	Christen Bergman: Match-on-card for secure and scalable biometric authentication. In Nalini K. Ratha and Venu Govindaraju, editors, Advances in Biometrics. Springer, 2008.
BK06	J. Merkle, U. Martini, U. Korte, M. Krawczak, M. Niesing, R. Plaga, C. Tiemann, H. Vinck: Basic Concept for BioKey. Version 1.5.3. Report of Project BioKeyS. 2006
Bre08	J. Breebaart, C. Busch, J. Grave, E. Kindt: A Reference Architecture for Biometric Template Protection based on Pseudo Identities, in Proceedings BIOSIG 2008
bus09	C. Busch, D. Lodrova, E. Tabassi, W. Krodel: Semantic Conformance Testing for Finger Minutiae Data, Proceedings of IEEE IWSCN 2009, Trondheim, pp. 17-23, ISBN 978-82-997105-1-0, May 2009
CKL03	T. Charles Clancy and Negar Kiyavash and Dennis J. Lin:Secure Smartcard-Based Fingerprint AuthenticationACM Workshop on Biometrics: Methods and Applications, 2003

CLM07 R. Cappelli, A. Lumini, D. Maio and D. Maltoni:Fingerprint Image Reconstruction from Standard TemplatesIEEE Transactions on Pattern Analysis Machine Intelligence, 2007

CS07 A. Cavoukian und A. Stoianov: Biometric encryption: a positive-sum technology that achieves strong authentication, security and privacy. Whitepaper information and Privacy Commissioner/Ontario, 2007. available from www.ipc.on.ca.

CTG05 Tee Connie and Andrew Teoh and Michael Goh and David Ngo:PalmHashing: a novel approach for cancelable biometricsInformation Processing Letters, 2005

DCB+08 N. Delvaux, H. Chabanne, J. Bringer, B. Kindarji, P. Lindeberg, J. Midgren, J. Breebaart, T. Akkermans, M. van der Veen, R. N. J. Veldhuis, E. Kindt, K. Simoens, C. Busch, P. Bours, D. Gafurov, B. Yang, J. Stern, C. Rust, B. Cucinelli, D. Skepastianos: Pseudo identities based on fingerprint characteristics. In Intelligent Information Hiding and Multimedia Signal Processing, 2008, IIHMSP '08 International Conference, Harbin, China, pages 1063–1068, Los Alamitos, August 2008. IEEE Computer Society Press.

DKM+07 S.C. Draper, A. Khisti, E. Martinian, A. Vetro, J.S. Yedidia: Using distributed source coding to secure fingerprint biometrics. In Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference on, volume 2, pages II–129–II–132, April 2007.

DORS07 Y. Dodis, R. Ostrovsky, L. Reyzin, and A. Smith: Fuzzy extractors: How to generate strong keys from biometrics and other noisy data. Technical Report 2003/235, Cryptology ePrint archive, http://eprint.iacr.org. 2007

DRS04 Yevgeniy Dodis, Leonid Reyzin, Adam Smith: Fuzzy extractors: How to generate strong keys from biometrics and other noisy data. In EUROCRYPT, pages 523–540, 2004.

ecc07 CEN TC 224 WG15 Identification card systems: European Citizen Card, 2007.

EE95 European Parliament and European Council. Directive 1995/46/ec of the european parliament and of the council of 24 october 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data. http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31995L0046:EN:HTML, 1995.

FbFV Karthik Nandakumar and Anil K. Jain and Sharath PankantiFingerprint-based Fuzzy Vault: Implementation and Performance

Gutt84 Antonin Guttman:R-Trees: A Dynamic Index Structure for Spatial Searching

HB09 Daniel Hartung and Christoph Busch:Why Vein Recognition Needs Privacy Protectioniih-msp, pp.1090-1095, 2009 Fifth International Conference on Intelligent Information Hiding and Multimedia Signal Processing, 2009

HD05 Feng Hao and Ross Anderson and John DaugmanCombining cryptography with biometrics effectively

INN Innovatrics: IDKit PC SDK, version September 2009

ISOa ISO/IEC. 19092:2008 - financial services - biometrics - security framework.

ISOb	ISO/IEC. 19794 - information technology - biometric data interchange formats.
ISOc	ISO/IEC IS 19794-2:2005 - information technology - biometric data interchange formats part 2: Finger minutiae data.
ISOd	ISO/IEC. CD 24787 – indentification cards - On-card-biometric-Comparison
ISOe	ISO/IEC 9594-2, ITU-T X.tpp-1 (Telebiometric Protection Procedure-Part1): A guideline of technical and managerial countermeasures for biometric data security
ISOf	ISO 19092:2008, Financial Services – Biometrics – Security framework
ISOface	ISO/IEC 19794-5:2005, Biometric data interchange formats - Part5: Face image data
ISOfinger	ISO/IEC 19794:2005, Biometric data interchange formats - Part4: Finger image data
ISOg	ISO/IEC JTC1/SC37 19785-4 Information technology – Common Biometric Exchange Formats Framework – Part 4: Security block format specifications
ISOh	ISO/IEC 24714-1: Biometrics – jurisdictional and societal considerations for commercial applications – Part 1: General guidance
ISOi	ISO/IEC 24714-2: Biometrics – jurisdictional and societal considerations for commercial applications – Part 2: Practical Application to Specific Contexts
ISOidm	ISO/IEC CD 24760, Information technology - Security techniques - A framework for identity management
ISOj	ISO/IEC 18033-2: 2006, Information technology - Security techniques - Encrypion algorithms - Part 2: Asymmetric ciphers
ISOk	ISO/IEC 18033-3: 2005, Information technology - Security techniques - Encryption algorithms - Part 3: Block ciphers
ISOl	ISO/IEC 18033-4: 2005, Information technology - Security techniques - Encryption algorithms - Part 4: Stream ciphers
ISOm	ISO/IEC 14888 (All parts): Information technology - Security techniques - Digital signaures with appendix
ISOn	ISO/IEC 9796 (All parts): Information technology - Security techniques - Digital signature schemes giving message recovery
ISOo	SO/IEC 9797 (All parts): Information technology - Security techniques - Message authentication codes (MACs)
ISOp	ISO/IEC 19772: Information technology - Security techniques - Authenticated encryption
ISOq	International Standards ISO/IEC FDIS 29109-1 Information Technology - Conformance Testing Methodology for Biometric Interchange Formats defined in ISO/IEC 19794, Februar 2009
ISOtp	ISO/IEC CD 24745 Information technology - Security techniques - Biometric template protection

Jain08	A. Jain, K. Nandakumar, A. Nagar: Biometric Template Security, EURASIP Journal on Advances in Signal Processing, Volume 2008
JBN99	A.K. Jain, R. Bolle, S. Pankanti (Eds): Personal Identification In a Networked Society, Kluwer (1999)
JPH+99	Anil K. Jain, Salil Prabhaka, Lin Hong, Sharath PankantiFingerCode: A Filterbank for Fingerprint Representation and Matching. IEEE Computer Society Conference on Computer Vision and Pattern Recognition: 2187, 1999.
JS02	A. Juels, M. Sudan:A fuzzy vault scheme, In Proceedings IEEE Information Theory 2002, pages 408, 2002
JW99	A. Juels und M. Wattenberg: A fuzzy commitment scheme. In Proc. 6th ACMCCCS, pages 28–36, 1999.
KCZ06	Adams Kong and King-Hong Cheung and David Zhang and Mohamed Kamel and Jane You:An analysis of BioHashing and its variants
KKM08	U. Korte, M. Krawczak, J. Merkle, et al., A cryptographic biometric authentication system based on genetic fingerprints. Tagungsband Sicherheit 2008. Lecture Notes of Informatics P-128. Springer-Verlag. 2008.
KMN09	U. Korte, J. Merkle, M. Niesing: Datenschutzfreundliche Authentisierung mit Fingerabdrücken, in DuD - Datenschutz und Datensicherheit, 1/2009
Kor08	U. Korte, M.. Krawczak, J. Merke, R. Plage, M. Niesing, C. Tiemann, H. Vinck, U. Martini: A cryptographic biometric authentication system based on genetic fingerprints, in Proceedings Sicherheit 2008
KYE+04	Alper Kanak and Gebze Yüksek and Teknoloji EnstitüsüBIOMETRICS FOR COMPUTER SECURITY AND CRYPTOGRAPHY
LB08	Lucas Kevin Ballard: Robust Techiques For Evaluation Biometric Cryptographic Key Generators,The Johns Hopkins University, 2008
Lod09	D. Lodrova, C. Busch, E. Tabassi, W. Krodel, M. DrahanskySemantic Conformance Testing Methodology for Finger Minutiae Data, in Proceedings BIOSIG 2009, pages 31-42, 2009
LT03	J.-P. M. G. Linnartz, P. Tuyls: New shielding functions to enhance privacy and prevent misuse of biometric templates. In In AVBPA, pages 393–402, 2003.
Mal05	D. Maltoni, D Maio, A. Jain, S.Prabhakar: rHandbook of Fingerprint Recognition, Springer, 2005
MMT09	Preda Mihalescu and Axel Munk and Benjamin Tams:The Fuzzy Vault For Fingerpints is Vulnerable To Brute Force Attack BIOSIG 2009, 2009
MRW99	F. Monrose, M.K. Reiter, R. Wetzel: Password hardening based on keystroke dynamics, in Proceedings 6th ACM CCSS, pages 73-82, 1999
MS09	Ao, Meng and Li, Stan Z.:Near Infrared Face Based Biometric Key BindingICB '09: Proceedings of the Third International Conference on Advances in Biometrics, 2009
nbis	NIST Biometric Image Software http://fingerprint.nist.gov/NBIS/index.html

NIST-SD14 National Institute of Standards and TechnologySpecial Database 14, http://www.nist.gov/ts/msd/srd/nistsd14.cfm

NIST-SD29 National Institute of Standards and TechnologySpecial Database 29, http://www.nist.gov/ts/msd/srd/nistsd29.cfm

nist07 National Institute of Standards and Technology: Biometric Data Specification for Per-sonal Identity Verification, NIST Special Publication 800-76-1, 2007, http://csrc.nist.gov/publications/nistpubs/800-76-1/SP800-76-1_012407.pdf

NJ09 A. Nagar and A. K. Jain:On the Security of Non-Invertible Fingerprint Template TransformsIEEE Workshop on Information Forensics and Security (WIFS), 2009

NJP07 K. Nandakumar, A.K. Jain, S. Pankanti: Fingerprint-based fuzzy vault: Implementation and performance. Information Forensics and Security, IEEE Transactions on, 2(4):744–757, Dec. 2007.

NNJ07 Karthik Nandakumar and Abhishek Nagar and Anil K. Jain: Hardening Fingerprint Fuzzy Vault Using Password, International conference on biometrics 2007, 2007

NT Neurotechnology: Verifinger 6.0 Extended SDK, version September 2009

NTN02 S. Nanavati, M. Thieme, R. Nanavati: Biometrics Identity Verification in a Networked World, Wiley (2002)

Par03 ARTICLE 29 Data Protection Working Party. Working document on biometrics working document on biometrics. http://ec.europa.eu/justice_home/fsj/privacy/docs/wpdocs/2003/wp80_en.pdf, 2003. Last visited: November 26, 2009

RCB01 N.K. Ratha and J.H. Connell and R. Bolle:Enhancing security and privacy of biometric-based authentication systems

RCCB07 N. K. Ratha, S. Chikkerur, J. H. Connell, R. M. Bolle: Generating cancelable fingerprint templates. IEEE Trans. pattern analysis and machine intelligence, 29(4):561–572, 2007.

S79 Shamir, AdiHow to share a secret, Commun. ACM, 1979

SB07 W. J. Schreier, T.E. Boult: Cracking Fuzzy Vault and Biometric Encryption, in Proceedings BCC 2007

SK09 Hirata, Shinji and Takahashi, Kenta:Cancelable Biometrics with Perfect Secrecy for Correlation-Based MatchingICB '09: Proceedings of the Third International Conference on Advances in Biometrics, 2009

SKK04 Marios Savvides and B.V.K. Vijaya Kumar and P.K. Khosla:CANCELABLE BIOMETRIC FILTERS FOR FACE RECOGNITIONProceedings of the 17th International Conference on Pattern Recognition ICPR 2004, 2004

Smith04 Adam Davison Smith: Maintaining Secrecy when Information Leakage is Unavoidable, MASSACHUSETTS INSTITUTE OF TECHNOLOGY, 2004

SRS98 C. Soutar, D. Roberge, A. Stoianov, R. Gilroy, B. V. K. Vijaya Kumar: Biometric Enryption, in R. L. van Renesse, editor, Society of Photo-Optical Instrumentation

	Engineers (SPIE) Conference Series, volume 3314 of Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, pages 178–188, April 1998.
ST06	Berry Schoenmakers, Pim Tuyls: Efficient binary conversion for paillier encrypted values. In EUROCRYPT, pages 522–537, 2006.
STP09	K. Simoens and P. Tuyls and B. Preneel:Privacy Weaknesses in Biometric Sketchesthe 2009 IEEE Symposium on Security and Privacy, IEEE Computer Society, 2009
Tab04	E. Tabassi, C. Wilson, C. WatsonNational Institute of Standards and Technology NISTIR 7151, Fingerprint Image Quality, August 2004
TAK05	P. Tuyls, A. H. M. Akkermans, T. A. M. Kevenaar, G. J. Schrijen, A. M. Bazen, R. N. J. Veldhuis: Practical biometric authentication with template protection. In Audio and video-based biometric person authentication, pages 436–446.
tcM	ANSI/INCITS technical committee M1. Biometrics. http://m1.incits.org/.
TG04	P. Tuyls and J. Goseling:Capacity and examples of template protecting biometric authentication systems Biometric authentication workshop (BioAW 2004), 2004
TGG04	Andrew Teoh Beng Jin and David Ngo Chek Ling and Alwyn Goh:Biohashing: two factor authentication featuring fingerprint data and tokenised random number
TGN06	Teoh, A.B.J. and Goh, A. and Ngo, D.C.L.:Random Multispace Quantization as an Analytic Mechanism for BioHashing of Biometric and Random Identity InputsPattern Analysis and Machine Intelligence, IEEE Transactions on, 2006
TI09	Tanya Ignatenko: Secret-Key Rates and Privacy Leakage in Biometric Systems, 2009
TUR08	European Integrated Project TURBINE: http://www.turbine-project.eu
TVI04	P. Tuyls and E. Verbitskiy and T. Ignatenko and D. Schobben and T. H. Akkermans:Privacy protected biometric templates: ear identificationProceeding of SPIE, 2004
tw	Identity Theft Website: http://www.idtheft.com
UPJ05	Umut Uludag and Sharath Pankanti and Anil K. Jain:Fuzzy vault for fingerprintsin Proc. AVBPA, Lecture Notes in Computer Science 3546, 2005
VDR09	Anthony Vetro and Stark C. Draper and Shantanu Rane and Jonathan S. YedidiaSecuring Biometric Data
VK06	M. van der Veen and T. Kevenaar and G.-J. Schrijen and T. H. Akkermans and Z. Fei:Face biometrics with renewable templatesSecurity, Steganography, and Watermarking of Multimedia Contents VIII, 2006
VT03	E. Verbitskiy and P. Tuyls and D. Denteneer and J.-P. LinnartzReliable Biometric Authentication with Privacy Protection
wk09	http://en.wikipedia.org/wiki/Apriori_algorithm. Zuletzt besucht: Juli 2009
XR09	Xu, Haiyun and Veldhuis, Raymond N.J. and Bazen, Asker M. and Kevenaar, Tom A.M. and Akkermans, Ton A.H.M. and Gokberk, Berk: Fingerprint

Verification Using Spectral Minutiae RepresentationsIEEE Transactions on Information Forensics and Security, 4, 2009

XV08 Xu, Haiyun and Veldhuis, Raymond N.J. and Kevenaar, Tom A.M. and Akkermans, Anton H.M. and Bazen, Asker M: Spectral minutiae: A fixed-length representation of a minutiae set2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Workshop on Biometrics, 2008

YB09 B. Yang, C. Busch: Parameterized Geometric Alignment for Minutiae-Base Fingerprint Template Protection, in Proceedings IEEE BTAS 2009

YV04 S. Yang and I.M. Verbauwhede:Secure fuzzy vault based fingerprint verification systemConference Record of the Thirty-Eighth Asilomar Signals, Systems and Computers, 2004

Zhou07 Xuebing ZhouTemplate Protection and its Implementation in 3D Face Recognition Systems

ZK10 Xuebing Zhou and Ton Kalker:On the Security of BiohashingSPIE of Electronic Imaging, 2010

ZWBK09 Xuebing Zhou and Stephen D. Wolthusen and Christoph Busch and Arjan Kuijper:A Security Analysis of Biometric Template Protection SchemesInternational Conference on Image Analysis and Recognition ICIAR 2009, 2009

9 Stichwortverzeichnis

10 Glossar

Approximation. Näherung der Anfrage um sehr effizient unnötige Vergleiche auszuschließen.

Bifurkation. Gabelung einer Papillarlinie.

Biometrisches Kryptoverfahren oder -system. Biometrisches Authentisierungsverfahren, bei dem nicht das Template selbst abgespeichert wird, sondern nur ein Verifikationsstring, der mit Hilfe von fehlerkorrigierenden Codes aus dem Template und einem benutzerindividuellen geheimen Schlüssel berechnet wird, und nur wenig Information über das Template preisgibt.

BioNN. Identifikationsverfahren, das auf dem Abgleich geometrischer Beziehungen der Minutienpositionen basiert. Es ist rotations- und translationsinvariant.

BioSimJoin. Identifikationsverfahren, das durch Einsatz einer Indexstruktur sehr effizient den Abgleich unterschiedlicher Fingerabdrücke bearbeitet.

Chaff Points/Streupunkte. Im Verfahren von Juels und Sudan zufällige Punkte aus dem Raum der potenziellen biometrischen Merkmale (zum Beispiel Minutien), die zusammen mit den echten Merkmalen abgespeichert werden und diese damit verschleiern.

Comparison-Algorithmus. Im Kontext dieses Feinkonzepts ein Algorithmus, der zu zwei Mengen von Minutien ein Mapping und die Anzahl der in beiden Mengen vorkommenden Minutien ermittelt.

Core Point. Entspricht dem Zentrum der nördlichsten Singularität (Schleife, Delta oder Wirbel) eines Fingerabdruckbilds.

Dimensionalität. Anzahl der Features, die ein beliebiges Objekt beschreiben.

Enrolment. Phase in einem biometrischen Authentisierungsverfahren, in dem die biometrischen Templates der Benutzer erfasst und für die Verifikation hinterlegt werden.

Entropie. Informationsgehalt eines Bitstrings. Im Kontext dieses Feinkonzeptes wird nur die Min-Entropie betrachtet.

Falsche Matches. Durch den Comparison-Algorithmus ausgegebene Übereinstimmungen (Mappings) der zur Authentisierung präsentierten Minutien mit hinterlegten Chaff Points.

Falschakzeptanzrate (False Accept Rate FAR). Die Wahrscheinlichkeit bzw. relative Häufigkeit, dass unberechtigte Benutzer als genuine Benutzer authentifiziert werden.

Falschrückweisungsrate (False Reject Rate FRR). Die Wahrscheinlichkeit bzw. relative Häufigkeit, dass berechtigte Benutzer nicht erfolgreich authentifiziert werden.

Falschrübereinstimmungsrate (False Match Rate FMR). Die Wahrscheinlichkeit bzw. relative Häufigkeit, dass Templates berechtigter Benutzer nicht erfolgreich authentifiziert werden.

Falschrnichtübereinstimmungsrate (False Non-Match Rate FNMR). Die Wahrscheinlichkeit bzw. relative Häufigkeit, dass Templates unberechtigter Benutzer erfolgreich authentifiziert werden.

Feature. Biometrisches Merkmal. Im Kontext dieses Feinkonzeptes werden als Features Minutien von Fingerabdrücken verwendet.

Feature-Extraktions-Algorithmus. Im Kontext dieses Feinkonzeptes ein Algorithmus, der aus einem digitalen Fingerabdruck die Minutien ermittelt.

FingerQS-Datenbank. Vom BSI zur Analyse von Fingerabdruckerkennungsalgorithmen aufgebaute Datenbasis von Fingerabdruckbildern.

Fusion. Die Zusammenführung der biometrischen Informationen.

Genuine. authentisch

GeoMatch. Identifikationsverfahren, welches zwei Fingerabdrücke basierend auf geometrischen Beziehungen von Minutientupeln vergleicht. Es ist rotations- und translationsinvariant.

Gleichfehlerrate (Equal Error Rate EER). Die Falschakzeptanz-Rate, wenn die gleich Falschrückweisungs-Rate bei gegebener Einstellung in biometrischem System ist

Hash-Funktion. Im Kontext dieses Feinkonzeptes eine kollisionsresistente Hash-Funktion, das heißt eine Funktion, die zu Eingaben beliebiger Länge Hashwerte fester Länge berechnet, so dass es praktisch nicht möglich ist, zwei Eingaben zu berechnen, die den gleichen Hashwert besitzen.

Hamming-Distanz. Die Anzahl der unterschiedliche Stelle von zwei Blöcke

Imposter. nicht-authentisch

Korrekte Matches. Durch den Comparison-Algorithmus ausgegebene Übereinstimmungen (Mappings) der zur Authentisierung präsentierten Minutien mit hinterlegten Minutien (dem Template).

Korrelation. Statistischer Zusammenhang zwischen der Häufigkeitsverteilung zweier Messgrößen.

Künstliche Minutie. Eine Minutie, die aus der Zusatzinformation, wie PIN oder Passwort, generiert wird.

Mapping von Minutien. Die ein-eindeutige Zuordnung von Minutien aus einer Menge von Minutien mit einer anderen Minutienmenge.

Matrix-Comparator. Identifikationsverfahren, welches zwei Fingerabdrücke basierend auf den Distanzbeziehungen ihrer Minutien vergleicht. Es ist rotations- und translationsinvariant.

Min-Entropie. Logarithmus (zur Basis 2) der maximalen Wahrscheinlichkeit einer Zufallsvariable.

Minutie. Verzweigung oder Endpunkt der Papillarlinien eines Fingerabdrucks.

NIST SD14: Datenbank mit Fingerabdrücken bereitgestellt vom NIST (siehe [NIST-SD14])

Normalisierung. Ausrichtung eines Fingerabdruckes vor dem Matching, so dass die längste Strecke zwischen zwei Minutien in vertikaler Lage liegt.

Packungsdichte. Verhältnis zwischen der Anzahl der in einem Raumbereich platzierten, sich nicht überschneidenden Sphären und dem Volumen des Raumbereiches.

Papillarlinie. Charakteristische Linie in der Haut der Handinnenseite und der Fußsohle. Die Papillarlinien des Fingers definieren den Fingerabdruck.

Platzkomplexität. Maß für das Anwachsen des Speicherbedarfs eines Algorithmus.

Polynom-Interpolation. Verfahren zur Berechnung des eindeutig definierten Polynoms mit Grad k aus $k + 1$ Stützstellen. Das bekannteste Verfahren ist die Lagrange-Interpolation.

Pseudo-Zufallszahlengenerator. Algorithmus, der aus einem Bitstring konstanter Länge (dem Seed) eine unendliche Bitfolge berechnet, so dass die Vorhersage eines Bits ohne Kenntnis des Seeds – auch bei Kenntnis vorheriger Ausgabebits – nicht effizient möglich ist.

Quantisierung. Transformation auf ganzen Zahlen durch Division mit einer Konstante und anschließender Rundung.

Rayleighverteilung. Statistische Verteilung, die sich für die (euklidische) Länge eines 2-dimensionalen Vektors ergibt, dessen Koordinaten normalverteilt und statistisch unabhängig sind.

R-Baum. Eine in Datenbanksystemen verwendete räumliche Indexstruktur.

Reed-Solomon-Code. Fehlerkorrigierender Code, bei dem die Daten durch Evaluation eines Polynoms kodiert und durch Reed-Solomon-Decodierung decodiert werden.

Reed-Solomon-Decodierung. Verfahren zu Berechnung eines Polynoms mit Grad k aus einer Menge von n Punkten (x, y), von denen mehr als $k + 1$ viele auf dem Polynom liegen. Das im Kontext dieses Feinkonzeptes betrachtete Verfahren erfordert, dass mindestens $(n + k) / 2$ viele Punkte auf dem Polynom liegen.

ROC-Kurve. ROC-Kurve ist ein Diagramm, bei der die Detektionswahrscheinlichkeit (1-Falschrückweisungsrate) als Funktion der Falschakzeptanzrate aufgetragen ist. Sie ist ein Hilfsmittel, um die gesamte Erkennungsleistung eines biometrischen Systems zu beobachten.

Seitenkapazität. Maximale Anzahl von Einträgen in einer Datenseite des R-Baums.

Streupunkte. Siehe Chaff Points

Template. Datensatz, der die für eine Authentisierung genutzten Informationen eines biometrischen Merkmals enthält.

Template Fusion. Bezeichnet hier die Fusion der Fingerbildminutien mit den künstlich generierten Minutien zu einem Minutientemplate.

Toleranzparameter. Parameter d und ϵ des Comparison-Algorithmus, die seine Toleranz bei der Zuordnung der Minutien zweier Fingerabdrücke festlegen.

Zusatzinformation. Informationen, wie PIN oder Passwort, die zur Erzeugung künstlicher Merkmale, wie z.B. künstlicher Minutien, genutzt werden.