

Modellbasierte Charakterisierung von Herdbefunden in Mammogrammen

Diplomarbeit

zur Erlangung des Grades eines Diplom-Informatikers im Studiengang Computervisualistik

vorgelegt von

Sven-René von der Heide

Betreuer: Prof. Dr.-Ing. Dietrich Paulus, Institut für Computervisualistik,
Fachbereich Informatik
Erstgutachter: Prof. Dr.-Ing. Dietrich Paulus, Institut für Computervisualistik,
Fachbereich Informatik
Zweitgutachter: M.Sc. Matthias Elter, Fraunhofer-Institut für Integrierte Schaltungen

Koblenz, im Mai 2008

Erklärung

Ich versichere, dass ich die vorliegende Arbeit selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe und dass die Arbeit in gleicher oder ähnlicher Form noch keiner anderen Prüfungsbehörde vorgelegen hat und von dieser als Teil einer Prüfungsleistung angenommen wurde. Alle Ausführungen, die wörtlich oder sinngemäß übernommen wurden, sind als solche gekennzeichnet.

Die Richtlinien der Arbeitsgruppe für Studien- und Diplomarbeiten habe ich gelesen und anerkannt, insbesondere die Regelung des Nutzungsrechts.

Mit der Einstellung dieser Arbeit in die Bibliothek bin ich einverstanden. ja nein

Der Veröffentlichung dieser Arbeit im Internet stimme ich zu. ja nein

Koblenz, den

Unterschrift

Inhaltsverzeichnis

1	Einleitung	7
1.1	Medizinischer Hintergrund	9
1.2	Ziel der Arbeit	11
2	Stand der Technik	13
2.1	Computer-assistierte Detektion/Diagnose	13
2.2	Computer-assistierte Charakterisierung	17
3	Material und Methoden	19
3.1	Verwendete Datenbanken	19
3.2	Normierungsmethoden	20
3.3	Eigenfaces	22
3.3.1	Die Hauptkomponentenanalyse (PCA)	22
3.3.2	Die Eigenface-Methode	23
3.4	Die Fisherface-Methode	28
3.5	Die Eigenfeature-Regularization-Extraction-Methode	30
3.6	Testdurchführung	35
3.6.1	Kreuzvalidierung	35

3.6.2	Klassifikator	36
3.6.3	Receiver-Operating-Characteristics(ROC)-Analyse	37
4	Experimente und Ergebnisse	41
4.1	Versuchsaufbau und -durchführung	41
4.2	Ergebnisse	43
5	Diskussion	49
5.1	Fisherface-Methode	49
5.2	ERE-Methode	50
5.3	Eigenface-Methode	50
5.4	Vergleich der Methoden	51
6	Zusammenfassung	55
7	Ausblick	57
	Literaturverzeichnis	59
A	Verzeichnisse	63
A.1	Notationstabelle	64
A.2	Literaturrecherche	65
A.3	Patentrecherche	66
A.4	Demo-Applikation	68
A.5	Verwendete Bibliotheken	69

Kapitel 1

Einleitung

Nach offiziellen Statistiken des Statistischen Bundesamtes ist Brustkrebs die am häufigsten zum Tode führende Krebsart. Das Statistische Bundesamt veröffentlichte 2006 Zahlen zum Thema Brustkrebs bei Frauen¹. Darin wird die Zahl der Todesfälle an Brustkrebs im Jahr 2004 mit 17.590 beziffert. Diese Zahl entspricht einem Anteil an durch Krebs verursachten Todesfälle von 17%. Im Jahr 2005 war Brustkrebs in Deutschland mit 153.000 Fällen der zweithäufigste Behandlungsgrund. Die *Dachdokumentation Krebs* geht von 55.100 Neuerkrankungen im Jahr aus [RK206]. Die *American Cancer Society* beziffert die Zahl der Brustkrebsdiagnosen in den USA für das Jahr 2004 auf 215.990 Frauen. Die Zahl der Todesfälle durch Brustkrebs wird mit 40.110 Frauen angegeben [GCF07].

Diese Zahlen verdeutlichen die Dringlichkeit, dass die Behandlungsmethoden und die Vorbeuge von Brustkrebs verbessert werden muss. Die am erfolgversprechenste Methode ist die Früherkennung von Brustkrebs. Durch eine frühzeitige Erkennung von Brustkrebs wird die Überlebensrate erhöht und die Behandlungsmöglichkeiten verbessert.

Für eine mögliche Früherkennung werden Frauen und Radiologen sensibilisiert. Dazu werden in Deutschland Programme zur Reihenuntersuchung der Brust angeboten². Diese

¹Die veröffentlichten Zahlen stammen aus einer Pressemitteilung des Statistischen Bundesamtes. Diese Pressemitteilung (Nr.090 vom 02.03.2006) ist unter dem Titel *Brustkrebs ist gefährlichste Krebsart für Frauen* auf der Internetpräsenz <http://www.destatis.de> erreichbar

²Die Koordination der Einführung bundesweiten und flächendeckenden Mammographie-Screenings in Deutschland wird von der *Kooperationsgemeinschaft Mammographie* durchgeführt. Weitere Informa-

Screening-Programme beinhalten kontinuierliche, periodische Untersuchungen der Brust im Abstand von 2 Jahren. Durch eine Reihenuntersuchung der Brust können Mammographieaufnahmen miteinander verglichen werden. Dies erleichtert die Findung kleinster Veränderung im Brustgewebe. Die Teilnahme an diesen Programmen des mammographie-Screenings ist freiwillig und für Frauen im Alter von 50-69 Jahren kostenlos.

Ein Herd in der Brust wird mittels der Mammographie untersucht. Dabei werden unter Verwendung von Röntgenstrahlen Aufnahmen des Weichgewebes der Brust gemacht. Diese werden vom Radiologen auf Abnormalitäten untersucht. Anhand der rein subjektiven Einschätzung eines Radiologen werden die Abnormalitäten klassifiziert und entsprechend in die Kategorien benigne und maligne (gutartige/bösartige) Herdbefunde eingeordnet. Aufgrund dieser Einschätzung kommt es zu einer Frühbehandlung oder zu keiner weiteren Untersuchung des Befundes. Studien aus [OMBF06] zufolge, werden 10% bis 30% von Brustkrebskrankungen im Frühstadium nicht erkannt. Eine weitere Fehlrate beziffert die fälschlicherweise als möglicher Brustkrebs eingestuft Herde; diese Fehlrate wird mit 35% angegeben [OMBF06]. Ein solche Fehleinschätzung hat für den Patienten weitreichende, negative Folgen. Der Patient wird einer unnötigen psychischen und einer körperlichen Belastung ausgesetzt, die bei korrektem Herdbefund zu vermeiden ist. Um solche Fehleinschätzungen zu minimieren, wird zunehmend die *Computer-aided Detection/Diagnosis* (CAD) eingesetzt.

Das Ziel dieser Arbeit ist die Evaluation von Methoden multivariater Datenanalyse, eingesetzt zur Diagnose von Herdbefunden. Die aus der Gesichtserkennung bekannten Methoden Eigenfaces [TP91] und Fisherfaces [BHK97] werden auf Mammographieaufnahmen angewendet, um eine Einordnung von Herdbefunden nach benign oder malign zu tätigen. Eine weitere implementierte Methode wird als *Eigenfeature Regularization and Extraction* [JMK08] bezeichnet. In diesem Kapitel folgt eine Einführung zum Thema *Medizinischer Hintergrund*. Die Arbeit gliedert sich wie folgt in fünf weitere Kapitel. In Kapitel 2 wird ein Überblick zum aktuellen Stand der computer-assistierte Detektion/Diagnose gegeben. Weiterführende Literaturempfehlungen und Nennung vorhandener Patente werden im Anhang aufgeführt. Im Kapitel 3 *Material und Methoden* wird die verwendete Bilddatenbank

tionen findet man auf der Internetpräsenz der Gemeinschaft unter <http://www.kooperationsgemeinschaft-mammographie.de>

vorgestellt, Normierungsschritte aufgeführt und die implementierten Methoden beschrieben. Um die in Kapitel 4 aufgezeigten Ergebnisse nachzuvollziehen, werden vorangehend Versuchsaufbau und -durchführung beschrieben. Es folgen die Kapitel 5 *Diskussion*, Kapitel 6 *Zusammenfassung* und abschließend ein Ausblick zum Thema der Arbeit in Kapitel 7.

1.1 Medizinischer Hintergrund

Die Mammographie³ bezeichnet die Röntgenuntersuchung der Brust. Dabei werden weiche Röntgenstrahlen durch die Brust geschickt, um das Weiche Brustgewebe abbilden zu können [SFR02]. Die Röntgenaufnahmen zur Untersuchung der Brust, werden aus zwei verschiedenen Positionen getätigt. Die gebräuchlichsten Arten von Mammographieaufnahmen sind CC- und MLO-Aufnahmen. CC steht für *craniocaudal view* und bezeichnet Aufnahmen einer Brust entlang der vertikalen Körperachse. Als *mediolateral oblique* (MLO) werden Mammographieaufnahmen entlang der horizontalen Körperachse bezeichnet. Bild 1.1 zeigt eine schematische Darstellung der Aufnahmetechnik, Bild 1.2 zeigt eine CC- und eine MLO-Aufnahme.

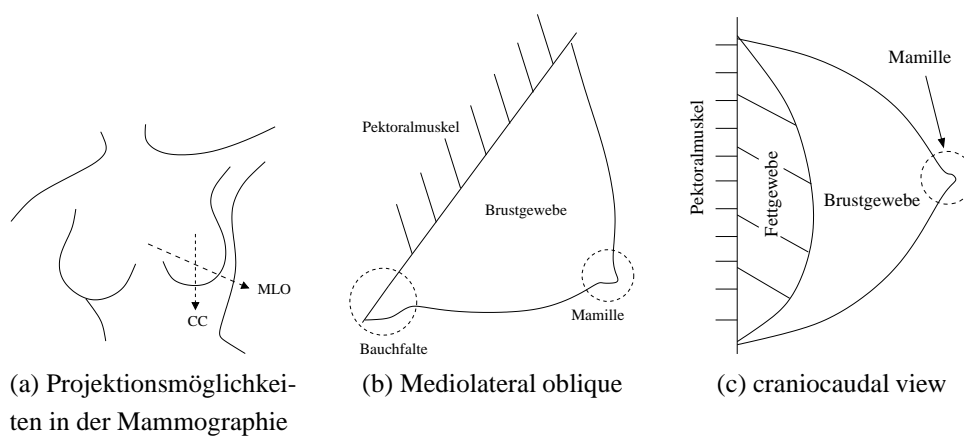
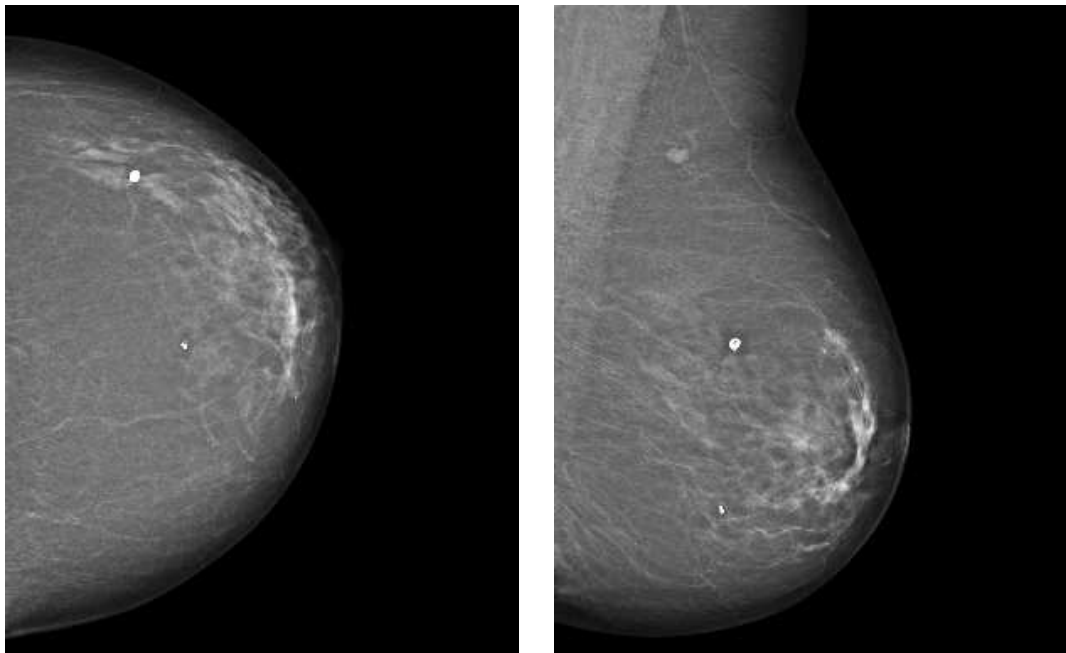


Bild 1.1: Verwendete Aufnahmetypen in der Mammographie [NOV04].

³*Mammo* ist die lateinische Bezeichnung für die weibliche Brust, *Graphie* bedeutet so viel wie Darstellungsverfahren und kommt aus dem Griechischen



(a) CC-Aufnahme einer linken Brust

(b) MLO-Aufnahme einer linken Brust

Bild 1.2: Mammographieaufnahmen von verschiedenen Positionen einer linken Brust.

Mammogramme werden hinsichtlich auffälliger Regionen untersucht. Die wichtigsten Arten von Läsionen sind Mikroverkalkungen und Knoten im Brustgewebe. Mikroverkalkungen können Hinweise auf Brustkrebs geben. Gewebeverdichtungen, bzw. Knoten in der Brust können weitere Hinweise für Krebs sein. Knoten können unter anderem durch die Kriterien Form, Ausrichtung, Randbegrenzung und Dichte des Gewebes als gut- oder bösartig eingestuft werden. Einstufungen nach dem Kriterium Form des Knoten, sind in Bild 1.3 dargestellt. Erschwerend für die Diagnose ist das Problem der Gewebeüberlagerung. Bild 1.4 veranschaulicht anhand von zwei Beispiele von Herden im Brustgewebe, wie die Einstufungen nach dem Kriterium Form vorgenommen werden. Durch die herkömmliche Aufnahmetechnik in der Mammographie, wird ein 3D-Körper im zweidimensionalen Raum abgebildet. Dadurch kann Gewebe die zu betrachtende Regionen überlagern. Da Mikroverkalkungen oft einen höheren Kontrast als Gewebeverdichtungen aufweisen, ist die Gewebeüberlagerung von Regionen für Herdbefunde problematischer. Die Schwierigkeit liegt in der Findung der Begrenzung von Gewebeverdichtungen. Gewebeüberlagerung

kann nicht immer eindeutig von der Region getrennt werden, die untersucht werden soll. Diese Problematik ist in Bild 1.4 ersichtlich. Die dortigen Diagnosevorschläge der Herdbefunde sind für Bild 1.4a benign und für Bild 1.4b malign.

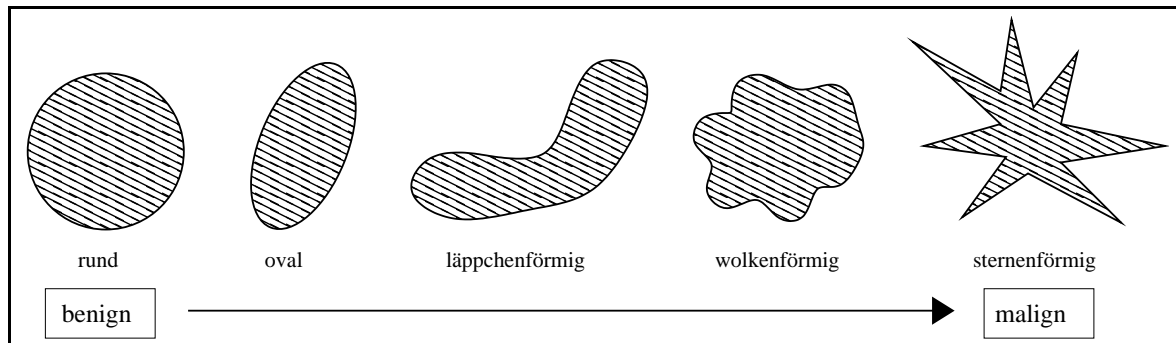


Bild 1.3: Einordnung der Formen von Knoten nach benign bis malign.

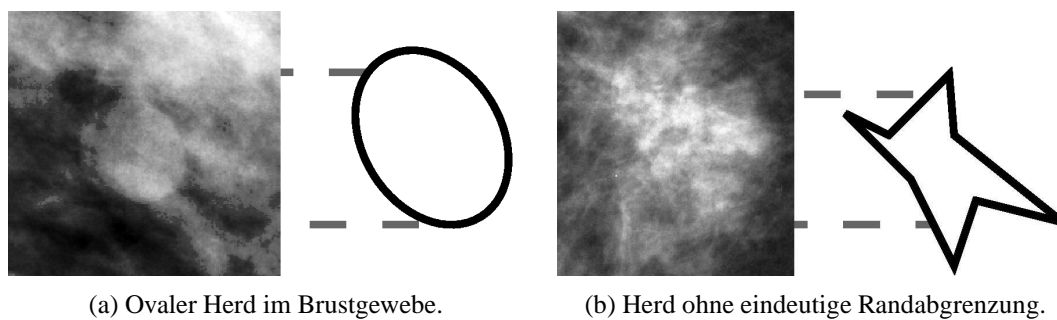


Bild 1.4: Beispiele für Herdbefunde und mögliche Einordnung nach Form.

1.2 Ziel der Arbeit

Ziel der Arbeit ist die Implementierung und Evaluation von Methoden zur Diagnose von Herdbefunden in Mammogrammen. Die implementierten Verfahren bilden diagnostizierte Herdbefunde in Modellen ab. Anhand dieser Modelle werden Herdbefunde charakterisiert um eine Diagnose zu stellen. Zu diesen Verfahren gehören die Eigenface- und Fisherface-Methode. Zusätzlich wird in der Arbeit eine Methode namens *Eigenfeature Extraction and*

Regularization implementiert und evaluiert. Beschrieben werden die jeweiligen Methoden in Kapitel 3, die erzielten Ergebnisse werden in Kapitel 4 dargestellt und diskutiert.

Die Diagnose eines Herdbefundes, erstellt mit einer der implementierten Methoden, kann in einem System zur computer-assisierten Diagnose zur Entscheidungsfindung eingesetzt werden. Kapitel 2 gibt im Abschnitt 2.1 einen Überblick zum Thema computer-assistierte Diagnose und Detektion im Bereich der Mammographie. Um die Ergebnisse zu visualisieren, wird eine Applikation erstellt. Diese Demo-Applikation ist im Anhang beschrieben und liegt der Arbeit bei.

Kapitel 2

Stand der Technik

2.1 Computer-assistierte Detektion/Diagnose

Im Bereich der Mammographie werden Computersysteme zur Detektion und Diagnose von Herdbefunden allgemein als CAD-Systeme¹ bezeichnet. Um zwischen Computersystemen zu unterscheiden, die ausschließlich für die Detektion oder Diagnose eingesetzt werden, wurden die Bezeichnungen CADe für computer-assistierte Detektion und CADx für computer-assistierte Diagnose eingeführt. Diese Systeme sollen Radiologen bei der Findung von auffälligen Herden im Brustgewebe und deren Einordnung nach benign und malign helfen. Das erste CADx-System wurde 1967 von Winsberg et al. in [WEM⁺67] publiziert. Dieses System konnte sich in diesem frühen Stadium der technischen Möglichkeiten nicht durchsetzen. Ein weiterer entscheidender Faktor für den Misserfolg eines solchen Computersystems für die Diagnose von Herdbefunden, war dessen Konzept. Das CADx-System wurde so konzipiert, dass ein Radiologe auf die Entscheidungsfindung der Diagnose keinen Einfluss nehmen kann. Mitte der 80er Jahre wurde an der Universität Chicago das erste CADx-System entwickelt, das nicht ausschließlich eine Diagnose von Herdbefunden stellen soll, sondern den Radiologen in seiner Entscheidungsfindung unterstützt. Daraus resultiert die Bezeichnung *computer-assistierte* Diagnose, bzw. Detektion. Das Computersysteme eine assistierende Rolle in der Radiologie einnehmen, hat sich bis

¹CAD steht für die englische Bezeichnung *computer-aided diagnosis/detection*

heute bewährt² [DHR07].

CADe-Systeme haben bei der Detektion eine hohe Sensivität, jedoch eine geringere Spezifität gegenüber Radiologen. In klinischen Studien hat sich gezeigt, dass durch die Verwendung von CADe-Systemen 10% mehr Krebsherde detektiert werden [Nis07]. Diese prozentuale Steigerung ist vergleichbar mit einer Untersuchung von Mammographieaufnahmen von zwei Radiologen. Um die Gegenprüfung eines Mammographie-Screenings durch einen zweiten Radiologen zu vermeiden, lassen sich solche Systeme einsetzen. Dadurch werden Kosten reduziert und Radiologen in ihrer Entscheidungsfindung unterstützt. Bild 2.1a zeigt den allgemeinen Ablauf eines computer-assistierten Detektionssystems.

Computer-assistierte Diagnose wird ebenso zur Unterstützung von Radiologen eingesetzt wie CADe-Systeme. Durch den Einsatz von CADx-Systeme wird die Zahl der Fehldiagnosen reduziert. Mammogramme sind im Kontext der Früherkennung von Brustkrebs schwer zu interpretieren. Zum einen können die potenziellen Herde sehr klein, zum anderen können Regionen von Brustgewebe überlagert sein. In diesem Zusammenhang ist die Qualität der Mammogramme entscheidend, da sich eine schlechte Bildqualität in Form von Rauschen und mangelnder Kontrast auf die Diagnose auswirken kann. Für Radiologen ist es wichtig auf ihren Erfahrungswerten zurückgreifen zu können, um eine möglichst genaue Diagnose stellen zu können. Wie auch bei der Detektion, wird ohne computer-assistierte Diagnose in der Regel ein zweiter Radiologe hinzugezogen. CADx-Systeme sollen als Kontrollfunktion für Radiologen eingesetzt werden. Für diese Aufgabe werden detektierte Regionen mit möglichen Tumorbildungen an das CADx-System übergeben. Dieses extrahiert Merkmale, um eine Diagnose stellen zu können. Ein möglicher Ablauf einer Untersuchung von Mammogrammen lässt sich in fünf Schritte gliedern [RAD07].

1. Vorverarbeitung des Radiologen:

Der Radiologe untersucht das Mammogramm auf Regionen, die einen potentiellen Herdbefund beinhalten. Diese werden markiert und an das CADx-System übergeben.

²Ein Beispiel für ein solches CAD-System ist das *R2 CAD*. Dies wurde 1998 eingeführt und seit dem für klinische Studien genutzt. Weitere Informationen sind auf der Internetpräsenz des Herstellers *Hologic* unter <http://www.r2tech.com> zu finden.

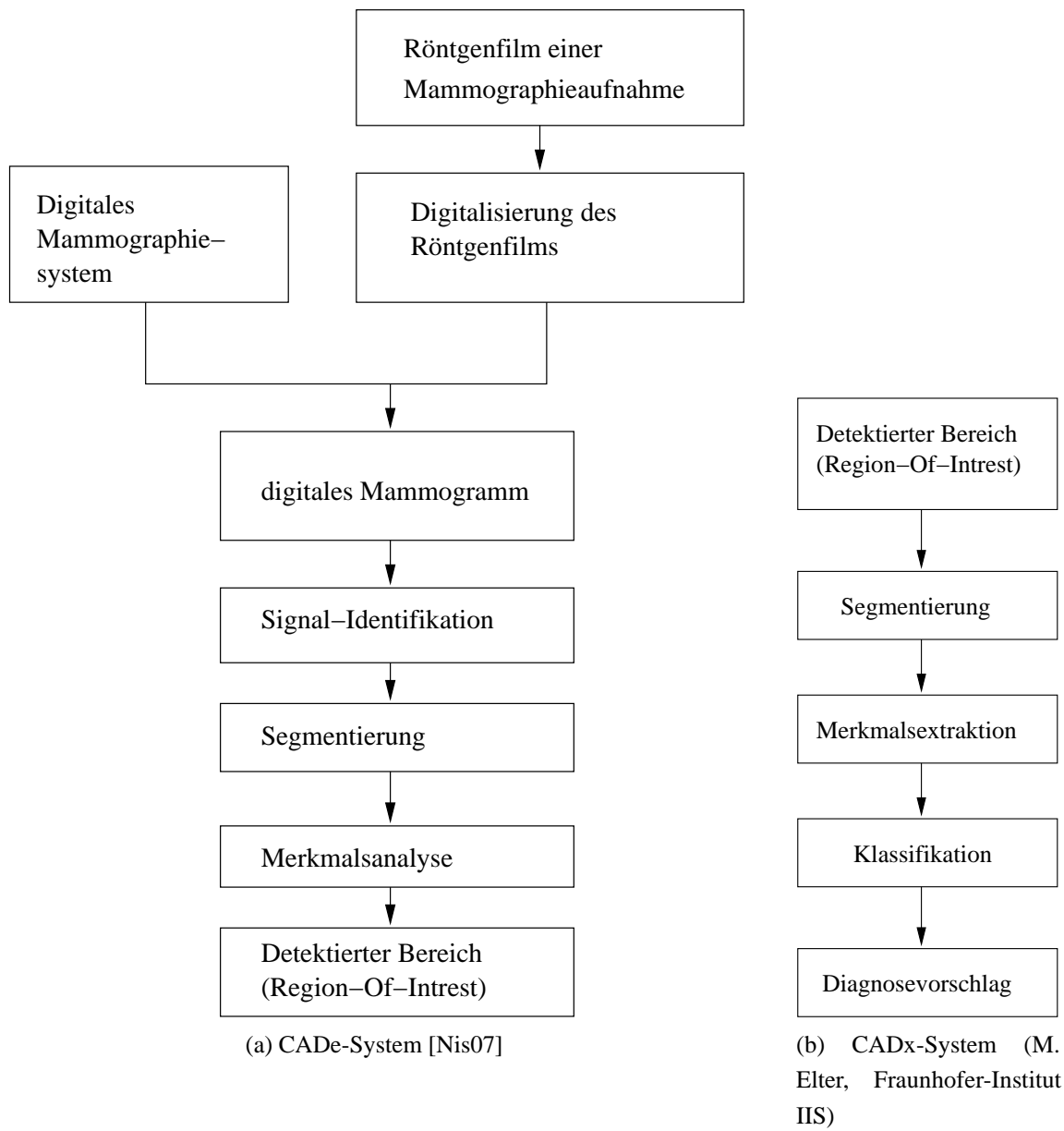


Bild 2.1: Ablauf von computer-assistierter Detektion/Diagnose in der Mammographie

2. Vorverarbeitung des CADx-Systems:
Das CADx-System schneidet die markierten Regionen aus, um daraus Merkmale zur Diagnose zu extrahieren.
3. Kontrolle des Radiologen:
Die ausgeschnittenen Regionen der Mammographieaufnahme werden nochmals auf ihre Relevanz hin untersucht. Unnötig markierte Regionen kann der Radiologe wieder entfernen, so dass keine Diagnose für eine solche Region erstellt werden muss. Desweiteren überprüft der Radiologe ein zweites Mal die Mammographieaufnahme, um weitere mögliche Regionen zu markieren.
4. Diagnose des CADx-Systems:
Der im System implementierte Algorithmus stellt anhand extrahierter Merkmale den Diagnosevorschlag, ob der Herdbefund benign oder malign ist.
5. Bestätigung einer Diagnose:
Die vom CADx-System gestellte Diagnose wird vom Radiologen geprüft. Bestätigt dieser die Diagnose, ist sie das Ergebnis der Untersuchung.

Freer und Ulissey veröffentlichten 2001 in [FU01] eine Studie zum Thema CAD-Systeme in der Mammographie. Darin wurden 12.860 Mammogramme unter Verwendung von CAD-Systemen untersucht. Die CAD-Systeme konnten 19,5% mehr Herde detektieren, als es Radiologen im Vorfeld konnten. Die Steigerung der detektierten Bereiche erhöhte die Detektion von malignen Herdbefunden in der Früherkennung von 73% auf 78%. Herdbefunde die falsch als malign eingestuft wurden, konnten mittels des CAD-Systems um 38% verringert werden. Diese Studie zeigt das enorme Potential von CAD-Systemen, die zur Unterstützung der Radiologen eingesetzt werden. Die Fehlerrate übersehener maligner Herde, sowie die Belastung von Patienten durch unnötig durchgeführte Biopsien wird verringert. Eine Biopsie wird durchgeführt, wenn sich mit Mammographie und Sonographie nicht zweifelsfrei klären lässt, ob ein Herdbefund benign oder malign ist. Dabei wird aus dem zu untersuchenden Bereich eine Gewebeprobe entnommen. Diese Gewebeprobe wird histologisch, zytologisch, immunhistologisch oder gentechnologisch untersucht um eine Diagnose zu stellen.

2.2 Computer-assistierte Charakterisierung

Die Charakterisierung von Herdbefunden kann anhand verschiedener Kriterien vorgenommen werden. Zu diesen Kriterien gehören Form und Ausrichtung, sowie Randbegrenzung und Randabgrenzbarkeit. Systeme zur computer-assistierten Diagnose, die Röntgenaufnahmen zur Entscheidungsfindung verwenden, stellen oft keine eigenen Diagnosen. Das Patent *US 2002/0181797 A1*³ beschreibt eine Methode, die den Radiologen eine verbesserte Grundlage für seine Diagnose bereitstellt. In dieser Methode werden die Herdbefunde nicht vom Computersystem nach Kriterien wie Form oder Ausrichtung diagnostiziert, sondern es wird dem Radiologen eine verbesserte grafische Darstellung bereitgestellt.

Ein Patent zur Charakterisierung von Herdbefunden ist in *P-Nr.: 6453058* beschrieben. Diese Methode extrahiert Merkmale der Herdbefunde und überführt diese in einen relationalen Attributengraph. Dieser Graph ist die Grundlage für die Diagnose. Eine andere Methode der computer-assistierten Charakterisierung ist die Verwendung des Kriteriums Form eines Herdes. Dabei wird die Form des zu untersuchenden Herdbefundes segmentiert und als Entscheidungsgrundlage verwendet.

Eine andere Art computer-assistierter Diagnose in der Mammographie ist die Verwendung von Ultraschallaufnahmen. Sind Mammographieaufnahmen mittels Röntgentechnik in 2D vergleichbar mit Ultraschallaufnahmen in 2D, gibt es im Bereich der Sonographie Verfahren, die auf 3D-Aufnahmen von Herdbefunden in der Brust angewendet werden. Bezeichnet wird dies als dreidimensionale Mammosonographie [Tho08]. Bei dieser Art der Diagnose werden die Kriterien zur Entscheidungsfindung dreidimensional betrachtet.

³Diese Zeichenfolge beschreibt eine Patentnummer. Im folgenden werden anderen Formaten von Patentnummer die Abkürzung *P-Nr.* vorangestellt.

Kapitel 3

Material und Methoden

3.1 Verwendete Datenbanken

Die Grundlage für die Testdurchführung der Methoden aus den Abschnitten 3.3, 3.4 und 3.5 ist die Bilddatenbank DDSM¹ [HBK⁺01] [HBK⁺98]. Die Bilddatenbank wird von vielen Forschungsgruppen² für die Analyse von Mammographieaufnahmen genutzt. Oliver et al. erzielten ihre Testergebnisse in [OMBF06] unter Verwendung von Testdaten aus der DDSM. Die Hauptunterstützung im Aufbau dieser verifizierten Datenbank erfolgte durch das Brustkrebsforschungsprogramm *U.S. Army Medical Research and Materiel Command* und der *University of South Florida*. Aufgebaut und zur freien Verfügung gestellt wurde die DDSM unter anderem, um die Entwicklung computergestützter Diagnose im Bereich der Mammographie voran zu treiben.

Die Datenbank enthält 2500 Brustkrebsstudien. Jeder Studie sind CC-, MLO-Aufnahmen der linken und rechten Brust zugeordnet, deren Bilddaten in 12bit vorliegen. Zusätzlich enthält jede Studie Daten über Regionen der Mammographieaufnahmen in Form

¹DDSM ist die Abkürzung für die englische Bezeichnung *Digital Database for Screening Mammography*

²Die Forschungsgruppe *Image Analysis Research Laboratory* der *University of South Florida* verwendet die DDSM als Referenzdatenbank für die Testdurchführung entwickelter Software für den Bereich Mammographie. Am *Navy Surface Warfare Center* wird die DDSM-Datenbank ebenfalls als Referenz verwendet[BB07].

von Plain-Text Annotationen. Die Studien sind nach benignen und malignen Befunden geordnet. Ebenso finden sich Informationen zu den Aufnahmeparametern der DDSM-Bilddatenbank, die eine Trennung nach Scannertyp der Studien zulassen.

Ein Problem der DDSM-Bilddatenbank ist die Qualität der Mammographieaufnahmen. Die nachträglich von Röntgenfilm digitalisierten Aufnahmen entsprechen nicht der heute im Mammographie-Screening verwendeten Technik. Die nachträgliche Digitalisierung der Bilddaten verstärkt das Problem vorhandenen Rauschens und geringem Kontrast der Aufnahmen. Neuere Aufnahmetechniken in der Mammographie erstellen direkt digitale Daten, ohne einen Röntgenfilm verwenden zu müssen.

3.2 Normierungsmethoden

Die in implementierten Methoden zur Datenanalyse, angewendet auf Mammographieaufnahmen, stellen besondere Anforderungen an die zu verwendenden Bilder. Das Ergebnis einer Detektion von Herden in Mammographieaufnahmen sind Parameter, die den zu untersuchenden Bereich definieren. Diese Bereiche werden als Region-Of-Interest (ROI) bezeichnet. Solche Regionen unterscheiden sich in ihrer Größe und in den Seitenverhältnissen. Um die Methoden auf Bilder anzuwenden, müssen diese auf eine einheitliche Größe gebracht werden. Dabei reicht es nicht aus, dass die Bilder, transformiert in einen Spaltenvektor, die gleiche Länge haben. Zusätzlich werden gleichbleibende Seitenverhältnisse gefordert. Bild 3.1 zeigt ein Beispiel, in dem zwei ROIs auf eine einheitliche Größe gebracht werden. In diesem Bild wird deutlich, dass aufgrund der angepassten Seitenverhältnisse der ROI das Bild mehr Informationen enthält als die ursprünglich markierte Region. Diese Art der Normierung ist notwendig, da bei der Bildnormierung von Mammogrammen vermieden werden muss, unkontrolliert mögliche Herde abzuschneiden. Dies hätte negative Auswirkungen auf die Ergebnisse der Methoden. Die in Bild 1.3 dargestellten Formen und deren Einstufungen von rund bis sternenförmig in benign und malign verdeutlicht, dass ein ovaler Herd durch eine Normierung mit abschneiden der linken und rechten Seiten einer runden Form angenähert wird. Dies bedeutet, dass eine solche Normierung das Ergebnis der Diagnose negativ beeinflussen würde. Da dieser Sachverhalt in Extremfällen einen eindeutig bösartig einzustufenden Herd als gutartig erscheinen lassen würde, wird

die Normierung wie beschrieben auf die Mammographieaufnahmen angewendet.

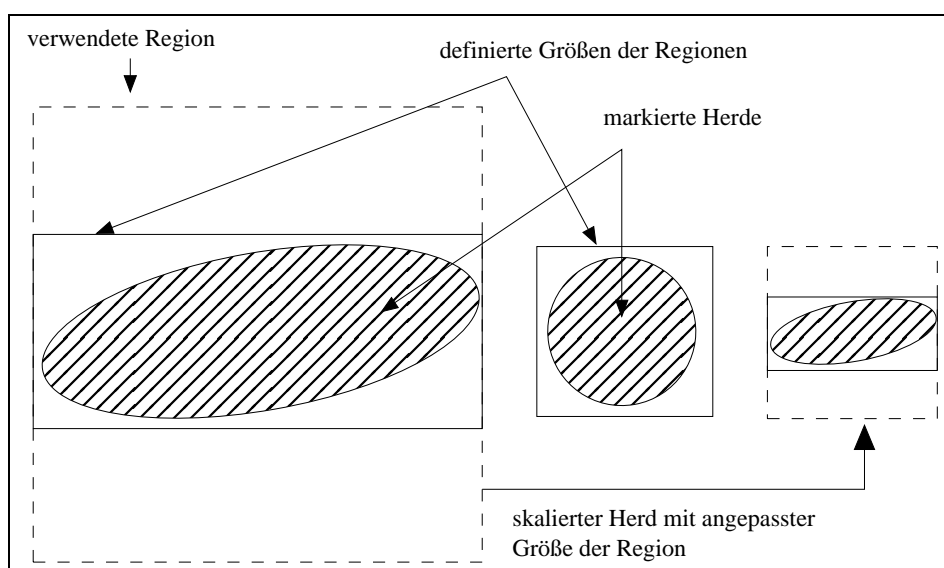


Bild 3.1: Beispiel einer Bildgrößennormierung

Ein weiteres Problem ist in Bild 3.1 dargestellt: das Verhältnis der nötigen Skalierung. Weichen die verwendeten Skalierungsfaktoren stark von einander ab, passiert es in der Anwendung, dass kleine Bildausschnitte sehr hoch skaliert werden müssen. Dadurch werden bei Anwendung der Bildvergleichsmethoden Informationen verwendet, die unkontrolliert hinzugefügt wurden. Dies kann das Ergebnis negativ beeinflussen. Um die Skalierungsfaktoren einheitlich wählen zu können, wird bei den in Kapitel 4 beschriebenen Ergebnisse eine weitere Normierungsmethode angewendet. Diese Normierungsmethode wird von Oliver et al. in [OMBF06] auf die Eigenface-Methode angewendet. Dabei werden die Bilder nach Größen der Ausgangsdaten in Gruppen aufgeteilt. In dieser Arbeit werden drei Gruppen erstellt. Dabei sind die Grenzgrößen der Bilder möglichst geschickt zu wählen, so dass die Skalierungsfaktoren auch innerhalb einer Gruppe nicht zu stark voneinander abweichen. Die Bilder der Testmenge werden auf ihre Gruppenzugehörigkeit hin untersucht und der entsprechenden Trainingsmenge zum Vergleich zugeordnet.

Da ein 3D-Körper in 2D abgebildet wird, kommt es in der Bildaufnahme zu Gewebeüberlagerungen. Werden Regionen markiert, kann nicht eindeutig bestimmt werden, was

das Gewebe des Herdes ist und welches genau das überlagernde Gewebe. Dadurch sind die Mammographieaufnahmen in den markierten Regionen in ihrem Kontrast nie optimal. In dieser Arbeit werden 8bit zur Berechnung verwendet. Dies entspricht dem Grauwertbereich $[0 \dots 255]$. Auf diesen Grauwertbereich werden die Bilder normiert. Eine weitere Normierungsmethode wird im folgenden als 8bit-Stretching bezeichnet. Bei dieser Methode werden die Bildausschnitte so normiert, dass der Grauwertbereich von $[0 \dots 255]$ voll ausgenutzt wird.

3.3 Eigenfaces

3.3.1 Die Hauptkomponentenanalyse (PCA)

Die Hauptkomponentenanalyse (PCA³) wurde zu Beginn des 20.Jahrhunderts vom Mathematiker Karl Pearson formuliert und findet heute in den unterschiedlichsten Einsatzgebieten Anwendung. So wird die Hauptkomponentenanalyse zur Datenanalyse in den Bereichen der Wirtschafts-, Sozial-, Geo- und Biowissenschaften, Psychologie, Chemie, Medizin sowie in der Informatik angewendet. Wie es unterschiedliche Einsatzbereiche für die PCA gibt, haben sich auch verschiedene Bezeichnungen für die Hauptkomponentenanalyse etabliert. In der Chemie wird die PCA als Faktorenanalyse, in der Signalverarbeitung und im Bereich des Datenretrieval wird sie als Karhunen-Loeve-Transformation (KLT) bezeichnet. In der Mathematik wird die PCA unter der Rubrik Eigenwertprobleme eingeordnet und auch als Hauptachsentransformation bezeichnet[Kes07].

Das Ziel der Hauptkomponentenanalyse besteht darin, eine Menge von beobachtbaren Variablen auf eine kleinere Menge, sogenannter latenter Variablen, zu reduzieren. Latente Variablen werden auch als Faktoren oder Hauptkomponenten bezeichnet. Eine solche Hauptkomponente ist eine Linearkombination der ursprünglichen Variablen, wodurch eine Reduzierung der Variablenmenge zu Stande kommt. Entsprechend wird die Hauptkomponentenanalyse häufig zur Datenanalyse und Datenkompression verwendet. Die Linearkombination setzt sich aus der linearen Summe von beobachtbaren Variablen zusammen, die

³PCA ist die Abkürzung für die englische Bezeichnung *principal component analysis*

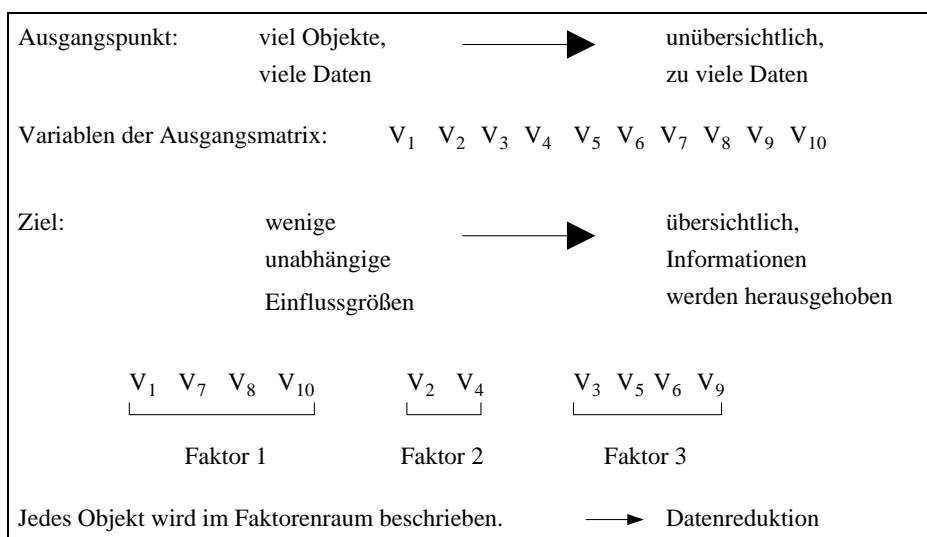


Bild 3.2: Prinzip der Hauptkomponentenanalyse [Kes07].

unterschiedlich gewichtet werden. Bild 3.2 stellt das Prinzip der Hauptkomponentenanalyse grafisch dar: Eine Menge von Merkmalsvariablen eines Objektes werden zu wenigen Hauptkomponenten zusammengefasst, um das Objekt zu beschreiben. Die PCA soll Informationen aus einer großen Menge von Daten hervorheben und diese gleichzeitig in einem niederdimensionalen Raum erhalten. In Bild 3.3 wird die Auswirkung der Wahl einer Hauptkomponente dargestellt. Bild 3.3a zeigt, dass die Projektionsgerade entlang der maximalen Variation in den Daten gewählt werden muss, damit beide Gruppen deutlich voneinander getrennt werden. Dies bedeutet, dass die Hauptkomponente in Richtung der maximalen Varianz verlaufen muss. Im Gegensatz zu Bild 3.3b zeigt Bild 3.3a, wie sich die beiden dargestellten Gruppen vermischen, wenn die Projektionsgerade in Richtung der minimalen Varianz verläuft.

3.3.2 Die Eigenface-Methode

Die Eigenface-Methode wurde von Turk und Pentland in [TP91] vorgestellt. Diese Methode verwendet zur automatisierten Gesichtserkennung die Hauptkomponentenanalyse. Die Idee dieser Methode besteht darin, mittels Informationen über Unterschiede zwischen Bil-

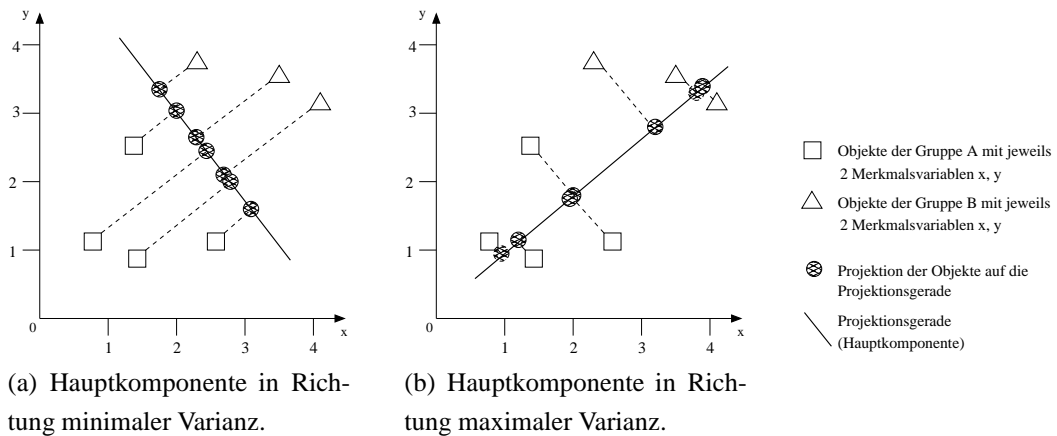


Bild 3.3: Wahl der Hauptkomponente und deren Auswirkungen [Kes07].

dern aus einer vorhandenen Trainingsmenge, Bilder zu kodieren und vergleichbar zu machen. Dabei werden die Unterschiede unabhängig von den üblichen Gesichtsmerkmalen betrachtet. Die Eigenface-Methode bedient sich nicht den Gesichtsmerkmalen wie Nase, Mund oder Augenpartie, sondern betrachtet das Gesicht als Ganzes. Diese Informationen werden so effizient wie möglich kodiert, damit sie mit anderen, in gleicher Weise kodierten Informationen, vergleichbar werden. Angewendet auf die Gesichtserkennung bedeutet dies, dass jeder Pixel eines Bildes ein Merkmal darstellt und die gesamte Menge der Merkmale das Gesicht beschreibt. Durch die Hauptkomponentenanalyse werden die Bilder auf ihre signifikanten Merkmale reduziert, welche zum Vergleich ausreichen.

Um diese Methode erfolgreich auf Bilder anwenden zu können, müssen diese bestimmte Eigenschaften aufweisen. Um diese Eigenschaften zu erhalten, sind folgende Normierungsschritte erforderlich: Die Bilder müssen

1. gleiche Breite und Höhe haben,
2. Grauwertbilder sein und
3. sollten sich über den Kontrast gegeneinander abgrenzen.

Diese drei Eigenschaften lassen sich unterschiedlich begründen. Damit Bilder über ihre Merkmale vergleichbar sind, müssen diese die gleiche Anzahl von Pixel aufweisen. Auf-

grund der Betrachtung der Bilder als Ganzes, müssen die Bilder einer Trainingsmenge die gleiche Breite und Höhe aufweisen, da sonst mit schlechteren Ergebnissen bei der Gesichtserkennung zu rechnen ist. Da sich die Beschreibung und die Implementierung der Methode auf Grauwertbilder beschränkt, wird die Eigenschaft von Grauwerten aufgeführt. Die Abgrenzung der Bilder über den Kontrast verstärkt die Unterscheidbarkeit der Bilder, so dass bessere Ergebnisse beim Vergleich solcher Bilder zu erwarten sind.

Formal lässt sich die Eigenface-Methode wie folgt beschreiben:

1. Erstellen der Datenmatrix D

Um die Datenmatrix D zu erhalten, werden alle N Bilder $I \in \mathbb{R}^{m \times n}$ der Trainingsmenge in die Spaltenvektoren Γ_i transformiert.

$$I \in \mathbb{R}^{m \times n} \Rightarrow \Gamma \in \mathbb{R}^{(m \cdot n) \times 1} \quad (3.1)$$

$$D = [\Gamma_1 \Gamma_2 \dots \Gamma_N] \in \mathbb{R}^{(m \cdot n) \times N} \quad (3.2)$$

2. Erstellen der Kovarianzmatrix K

D soll nun der Hauptkomponentenanalyse unterzogen werden. Dazu wird die Kovarianzmatrix K erstellt, um Korrelationen zwischen Merkmalen zu verdeutlichen. Dafür wird als erstes der Mittelwertvektor μ nach (3.3) gebildet. Der Vektor μ wird dazu verwendet, der Bilderanzahl N entsprechende Spaltenvektoren ϕ zu berechnen. ϕ_i beschreibt alle Bilder N über die Abweichung zum Mittelwertvektor, aus denen sich Matrix K erstellen lässt (3.5). K wird über die Streuung der gesamten Bildinformation erstellt. In der Literatur wird diese Form der Matrix K auch als *Total-Scatter-Matrix* bezeichnet und als S_T abgekürzt.

$$\mu = \frac{1}{N} \sum_{i=1}^N \Gamma_i \quad (3.3)$$

$$\phi_i = \Gamma_i - \mu, i = 1 \dots N \quad (3.4)$$

$$K = \frac{1}{N} \sum_{i=1}^N \phi_i \phi_i^T \in \mathbb{R}^{(m \cdot n) \times N} \quad (3.5)$$

Gesucht wird die Projektionsmatrix \mathbf{W}_{PCA} für die gilt $\mathbf{W}_{PCA} = \arg \max_{\mathbf{W}} |\mathbf{W}^T \mathbf{K}_T \mathbf{W}|$. Diese Matrix lässt sich aus den Eigenvektoren der Matrix \mathbf{K} aufbauen. Aus diesem Grund werden im Folgenden die Eigenwerte und die zugehörigen Eigenvektoren der Kovarianzmatrix gesucht. Dies wird als *einfaches Eigenwertproblem* bezeichnet und ist in (3.6) formuliert. Dabei ist zu beachten, dass die \mathbf{K} eine quadratische Matrix mit den Dimensionen $(n * m) \times (n * m)$ ist, λ der Eigenwert des Eigenvektors \mathbf{w} mit der Länge $(n * m)$.

$$\mathbf{K} \mathbf{w} = \lambda \mathbf{w} \quad (3.6)$$

$$\mathbf{K} = \mathbf{A} \mathbf{A}^T, \text{ mit } \mathbf{A} = [\phi_1 \phi_2 \dots \phi_N] \quad (3.7)$$

Das Eigenwertproblem aus (3.6) kann mittels der Singulärwertzerlegung gelöst werden. Wie in (3.7) dargestellt, lässt sich die Kovarianzmatrix als Matrixmultiplikation schreiben. Zur Berechnung der Eigenwerte und -vektoren wird die Eigenschaft der Singulärwertzerlegung genutzt, so dass die Matrizen der Zerlegung von \mathbf{A} die gewünschten Eigenwerte und Eigenvektoren der Matrix $\mathbf{A} \mathbf{A}^T$ beinhalten.

3. Berechnung der Eigenwerte und der zugehörigen Eigenvektoren

Die gesuchten Eigenvektoren und Eigenwerte von \mathbf{K} erhält man mittels Anwendung der Singulärwertzerlegung auf die Matrix \mathbf{A} . Dadurch erhält man die Form $\mathbf{A} = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^T$ (vgl. 3.8). Dabei gilt für $\mathbf{U} \in \mathbb{R}^{(m*n) \times (m*n)}$ und $\mathbf{V} \in \mathbb{R}^{M \times N}$. $\mathbf{\Sigma} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_M)$ hält die Singulärwerte der Matrix \mathbf{A} in absteigender Reihenfolge, die gleichzeitig die Eigenwerte von \mathbf{K} sind. Die Spaltenvektoren aus \mathbf{U} sind die korrespondierenden Eigenvektoren \mathbf{w} zu den Eigenwerten aus $\mathbf{\Sigma}$ von \mathbf{K} . Die Eigenvektoren \mathbf{w} deren Eigenwerte λ ungleich Null sind, werden als Spaltenvektoren der Projektionsmatrix \mathbf{W} verwendet. \mathbf{W} wird im folgenden als \mathbf{W}_{PCA} bezeichnet. c ist die Anzahl von Eigenwerten die den Wert Null haben. Folglich gilt $\mathbf{W}_{PCA} \in \mathbb{R}^{(m*n) \times (N-c)}$.

$$\mathbf{A} = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^T = \mathbf{U} \begin{bmatrix} \lambda_1 & & & & & & \\ & \ddots & & & & & \\ & & \lambda_{(N-c)} & & & & \\ & & & 0 & & & \\ & & & & \ddots & & \\ & & & & & & \lambda_{(m*n)} \end{bmatrix} \mathbf{V}^T, \quad (3.8)$$

mit $\lambda_1 \geq \lambda_2 \geq \dots \lambda_{(N-c)} > \dots \lambda_{(m*n)}$

Die Projektionsmatrix \mathbf{W}_{PCA} wird nun dazu verwendet, ein Bild in Form eines Spaltenvektors $\mathbf{\Gamma}$ nach (3.9) in den neuen Unterraum zu projizieren. In diesem neuen Unterraum wird das Bild durch einen Vektor $\mathbf{\Gamma}^{PCA}$ der Länge $N - c$ repräsentiert. Die Elemente des neuen Vektors werden als Eigenmerkmale⁴ bezeichnet. Es werden alle N Bilder der Trainingsmenge in den neuen Unterraum projiziert. Um ein Bild mit der Trainingsmenge vergleichen zu können, muss dieses Bild wie die Trainingsmenge in die Form $\mathbf{\Gamma} \in \mathbb{R}^{(n*m) \times 1}$ gebracht und nach (3.9) in seinen Dimensionen reduziert werden. Der Spaltenvektor des Bildes, welches mit der Trainingsmenge verglichen werden soll, wird als $\mathbf{\Gamma}_{image}^{PCA} \in \mathbb{R}^{(n*m) \times 1}$ bezeichnet. Die Eigenmerkmale der Bilder aus der Trainingsmenge werden durch $x_1, x_2, \dots, x_{(N-c)}$ des jeweils transformierten Bildes $\mathbf{\Gamma}_i^{PCA}$ beschrieben. Nun wird mittels Euklidischer Distanz 3.10 die Nähe des Bildes zu dem jeweiligen Bild der Trainingsmenge bestimmt. $\mathbf{\Gamma}_{image}^{PCA}$ wird durch $y_1, y_2, \dots, y_{(N-c)}$ beschrieben. Für den euklid'schen Abstand d von $\mathbf{\Gamma}_{image}^{PCA}$ zu $\mathbf{\Gamma}_i^{PCA}$ mit $i = 1, 2, \dots, N$ ergibt sich die Form (3.10).

$$\mathbf{\Gamma}_i^{PCA} = \mathbf{W}_{PCA}^T \mathbf{\Gamma}_i, \text{ mit } i = 1, 2, \dots, N \quad (3.9)$$

$$d(\mathbf{\Gamma}_i^{PCA}, \mathbf{\Gamma}_{image}^{PCA}) = \sum_{j=1}^{(N-c)} \sqrt{(x_j - y_j)^2} \quad (3.10)$$

⁴Für Eigenmerkmale wird auch die englische Bezeichnung *Eigenfeatures* verwendet

3.4 Die Fisherface-Methode

In der Gesichtserkennung gibt es neben der in (3.3) beschriebenen Eigenface-Methode die Fisherface-Methode [BHK97]. Diese Methode verwendet die Fisher-Lineare-Diskriminanzanalyse (LDA) nach angewandeter Hauptkomponentenanalyse. Die Fisherface-Methode projiziert ein Gesicht aus dem hochdimensionalen Bildraum in einen niederdimensionalen Eigenschaftsraum. Dieser Eigenschaftsraum ist unempfindlich gegen Veränderungen in der Beleuchtungsrichtung und gegen Veränderungen im Gesichtsausdruck. Dies ermöglicht die Einteilung der Bilder für die Trainingsmenge nach Personen. Jede Person kann in ihrer Klasse eine beliebige Menge von eigenen Gesichtsbildern haben. Die Methode ist klassenspezifisch, da die Fisherface-Methode versucht die Streuung zu formen. Die Eigenface-Methode erstellt dagegen über die gesamte Streuung der Bildinformationen die Kovarianzmatrix, auf die die Hauptkomponentenanalyse angewendet wird.

Die N Bilder $\mathbf{I} \in \mathbb{R}^{m \times n}$ werden in k Klassen unterteilt, so dass $\mathbf{I}_{ij} \in \mathbb{R}^{m \times n}$ gilt. N^k ist die Anzahl von Bildern der Klasse k . Daraus ergibt sich die Notation für den Spaltenvektor des i -ten Bildes der Klasse k ; \mathbf{I}_i^k mit $i = 1, 2, \dots, N^k$. Zur Anwendung der linearen Diskriminanzanalyse nach Fisher werden in (3.11) und (3.12) zwei Kovarianzmatrizen erstellt. In der Literatur werden diese beiden Matrizen als *Between-Class-Scatter-Matrix* (S_B) und als *Within-Class-Scatter-Matrix* (S_W) bezeichnet. Da es sich in beiden Fällen um Kovarianzmatrizen handelt, werden in dieser Arbeit die beiden Matrizen als \mathbf{K}_B und \mathbf{K}_W bezeichnet.

$$\mathbf{K}_B = \sum_{i=1}^k (\boldsymbol{\mu}^i - \boldsymbol{\mu})(\boldsymbol{\mu}^i - \boldsymbol{\mu})^T \quad (3.11)$$

$$\mathbf{K}_W = \sum_{i=1}^k \sum_{j=1}^{N^k} (\mathbf{I}_j^i - \boldsymbol{\mu}^i)(\mathbf{I}_j^i - \boldsymbol{\mu}^i)^T \quad (3.12)$$

Gesucht wird die optimale Projektionsmatrix \mathbf{W}_{OPT} mit orthonormalen Spaltenvektoren, die das Verhältnis der Determinante von \mathbf{K}_B und \mathbf{K}_W maximiert (3.13). Diese Projektionsmatrix erhält man, indem das allgemeine Eigenwertproblem nach (3.14) gelöst wird.

$$\mathbf{W}_{OPT} = \arg \max_{\mathbf{W}} \frac{|\mathbf{W}^T \mathbf{K}_B \mathbf{W}|}{|\mathbf{W}^T \mathbf{K}_W \mathbf{W}|} \quad (3.13)$$

$$= [\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_k], \text{ mit } \{w_k | k = 1, 2, \dots, p\}$$

\mathbf{W}_{OPT} besteht aus den Eigenvektoren \mathbf{w} von \mathbf{K}_B und \mathbf{K}_W die mit den absteigend sortierten Eigenwerten λ korrespondieren.

$$\mathbf{K}_B \mathbf{w}_k = \lambda_k \mathbf{K}_W \mathbf{w}_k \quad (3.14)$$

In der Bildverarbeitung besteht das Problem in der Anwendung multivarianter Datenanalyse darin, dass jeder Pixel eines Bildes als Information interpretiert wird, und diese Anzahl von Pixeln sehr groß sein kann. Dies hat zur Folge, dass bei erstellten Trainingsmengen die Anzahl der Pixel wesentlich größer ist als die Anzahl der verwendeten Bilder in der Trainingsmenge. Ein Beispiel aus Kapitel 4 verdeutlicht dies: Bei 400 Bildern in der Trainingsmenge mit Bildgrößen von 80×80 Pixeln, ist das Verhältnis der Pixelanzahl zur Anzahl verwendeter Bilder 16:1. Aufgrund dieses Verhältnisses wird \mathbf{K}_W singulär. Dadurch kann \mathbf{W} so gewählt werden, dass \mathbf{K}_W als Nullmatrix geschrieben werden kann und die Streuung innerhalb der Klassen in \mathbf{W}_{opt} nicht berücksichtigt wird.

Daher wendet die Fisherface-Methode die Hauptkomponentenanalyse an bevor die Fisher-Linear-Diskriminanz-Analyse durchgeführt wird. Nach der Hauptkomponentenanalyse werden die Bilder nicht mehr durch ihre Pixelwerte repräsentiert, sondern durch Skalare im Facespace. Diese Skalare werden auch als Eigenmerkmale bezeichnet und bedeuten eine Reduzierung der zu verwendenden Dimensionen. Die Repräsentation der Bilder durch ihre Eigenmerkmale hat den Vorteil, dass aus dem ungleichen Verhältnis von Pixelanzahl zu Anzahl von Bildern einer Trainingsmenge das maximale Verhältnis 1:1 wird. Dadurch wird die \mathbf{K}_W nicht singulär, und \mathbf{W} kann nicht mehr so gewählt werden, dass \mathbf{K}_W als Nullmatrix angesehen werden kann und der Informationsgehalt in \mathbf{W}_{OPT} minimiert wird. Daraus ergibt sich für die Fisherface-Methode in (3.15) eine neue Formulierung für \mathbf{W}_{OPT} . Als \mathbf{W}_{PCA} wird die Projektionsmatrix der Eigenface-Methode bezeichnet. \mathbf{W}_{FLD} bezeichnet die berechnete Projektionsmatrix unter Verwendung von \mathbf{W}_{PCA} . Aus \mathbf{W}_{FLD} und \mathbf{W}_{PCA} wird \mathbf{W}_{OPT} berechnet.

$$\begin{aligned}
\mathbf{W}_{opt}^T &= \mathbf{W}_{fld}^T \mathbf{W}_{pca}^T & (3.15) \\
\mathbf{W}_{pca} &= \arg \max_{\mathbf{W}} |\mathbf{W}^T \mathbf{K}_T \mathbf{W}| \\
\mathbf{W}_{fld} &= \arg \max_{\mathbf{W}} \frac{|\mathbf{W}^T \mathbf{W}_{pca}^T \mathbf{K}_B \mathbf{W}_{pca} \mathbf{W}|}{|\mathbf{W}^T \mathbf{W}_{pca}^T \mathbf{K}_W \mathbf{W}_{pca} \mathbf{W}|}
\end{aligned}$$

3.5 Die Eigenfeature-Regularization-Extraction-Methode

Die Schwierigkeit in der Gesichtserkennung besteht in der Extraktion von unterscheidbaren und stabilen Merkmalen. Die PCA maximiert die Varianz von Merkmalen, minimiert den Fehler bei der Rekonstruktion von Bildern und beseitigt das Rauschen durch Ausgrenzung entsprechender Dimensionen. Die Eigenface-Methode hat sich im Bereich der Gesichtserkennung unter Anwendung multivarianter Datenanalyse etabliert, ist aber keine optimale Lösung. Die PCA verwendet als Kovarianzmatrix die Streuung der gesamten Bildinformationen, ohne diese zu ordnen. Hingegen extrahiert die LDA unterscheidbare Merkmale unter Verwendung von Streuungsinformation zwischen und innerhalb von Bildklassen. In der Gesichtserkennung ist dieser Ansatz problematisch, da aufgrund der hohen Bilddimensionen und der weitaus geringeren Anzahl verwendeter Bilder die Kovarianzmatrix, die die Streuung innerhalb der Bildklassen beschreibt, singular wird. Um diesem Problem zu begegnen, führt die Fisherface-Methode erst eine Hauptkomponentenanalyse durch, um die Bilddimensionen auf eine geringere Menge von Eigenmerkmalen zu reduzieren. Dieses Problem ist in Kapitel 3.4 im genauen Zusammenhang erklärt. Durch diese Dimensionsreduktion wird die Unterscheidbarkeit minimiert, so dass die Vorteile der LDA nach Fisher in ihrer Wirkung abgeschwächt werden. Beide Methoden weisen die Eigenschaft auf, dass für die Berechnung des Projektionsraumes lediglich Eigenvektoren verwendet werden, deren Eigenwerte ungleich Null sind. Der so aufgespannte Raum wird als Gesichtsraum⁵ bezeichnet.

Die Hauptkomponentenanalyse bewirkt nach Anwendung auf Bilder, dass die durch ihre Pixelwerte beschriebenen Bilder im Projektionsraum durch sogenannte Gewichtung-

⁵Als Gesichtsraum wird auch die englische Bezeichnung *Facespace* verwendet.

faktoren⁶ beschrieben werden. Diese Gewichtungsfaktoren werden nicht unterschiedlich gewichtet, obwohl der Informationsgehalt mit Abnahme der zugehörigen Eigenwerte abnimmt. Die Eigenfeature-Regularization-Extraction-Methode (ERE-Methode) [JMK08] reguliert die berechneten Eigenfeatures für die weitere Verwendung, um daraus einen neuen Unterraum aufzubauen (Extraktion der regulierten Eigenmerkmale).

Diese Methode beschäftigt sich mit den Fragen, welche Informationen in den Eigenvektoren vorhanden sind, deren Eigenwerte gleich Null sind und wie lässt sich der Facespace gewichten. Bei der Rekonstruktion von Bildern aus dem Gesichtsraum in den ursprünglichen Bildraum ist es möglich, bei unterschiedlicher Gewichtung der Eigenvektoren zu erkennen, dass die Hauptinformationen der Rekonstruktion in den vorderen Eigenvektoren zu finden ist. Aus diesem Grund wird bei der ERE-Methode der Gesichtsraum vergrößert und in drei Unterräume aufgeteilt. Zu berechnen sind die Grenzen vom Facespace zum Noisespace, und vom Noisespace zum Nullspace. Nach dem die einzelnen Räume definiert wurden, lassen sich Gewichtungsfaktoren für die Eigenvektoren der Projektionsmatrix berechnen. Die Trainingsmenge wird mittels der Projektionsmatrix in den unterschiedlich gewichteten Unterraum projizieren. Diese Darstellung der Bilder wird verwendet, um eine neue Kovarianzmatrix zu erstellen, die die komplette Streuung der Bildermenge im gewichteten Unterraum repräsentiert.

Die Methode lässt sich in fünf Schritte gliedern. Im folgenden werden diese Schritte erläutert und mathematisch skizziert:

1. Erstellen der Kovarianzmatrix \mathbf{K}_W aus der Menge von Ausgangsbildern in Berechnung der dazugehörigen Eigenvektoren und Eigenwerte:

Die Ausgangsbilder werden nach (3.16) in die Kovarianzmatrix \mathbf{K}_W gebracht. Dabei wird die Streuung innerhalb der k Klassen in der Kovarianzmatrix berücksichtigt. Um die Eigenvektoren und die Eigenwerte zu erhalten muss das einfache Eigenwertproblem gelöst werden. Um dies zu vereinfachen, wird wie in Kapitel 3.3 \mathbf{K}_W in die Form $\mathbf{A}\mathbf{A}^T$ gebracht. Dieses Vorgehen wird in (3.17) beschrieben.

p_k ist der Wert für die Wahrscheinlichkeit, dass die Klasse k in der Trainingsmenge vorkommt. Berechnet wird p_k mit $p_k = \frac{N}{N^k}$.

⁶Gewichtungsfaktoren werden auch als Eigenmerkmale oder Eigenfeatures bezeichnet.

$$\mathbf{K}_W = \sum_{i=1}^k \frac{p_k}{N^k} \sum_{N_{j=1}^{N^k}} (\boldsymbol{\Gamma}_j^i - \boldsymbol{\mu}^i)(\boldsymbol{\Gamma}_j^i - \boldsymbol{\mu}^i)^T \quad (3.16)$$

$$\mathbf{K}_W = \mathbf{A}\mathbf{A}^T \quad (3.17)$$

mit $\mathbf{A} = [\phi_1 \phi_2 \dots \phi_N]$,

$$\text{dabei gilt } \phi_i = \frac{\sqrt{p_k}}{\sqrt{N^k}} (\boldsymbol{\Gamma}_i^k - \boldsymbol{\mu}_i^k)$$

Nun wird die Singulärwertzerlegung wie in (3.8) auf \mathbf{A} angewendet um die Matrix $\boldsymbol{\Sigma}$ mit den Eigenwerten und den dazugehörigen Eigenvektoren zu erhalten. Nun werden für die weiteren Berechnungen alle Eigenvektoren der Matrix \mathbf{U} und die komplette Diagonale mit den Eigenwerten von $\boldsymbol{\Sigma}$ verwendet. Dadurch ergibt sich in den gewichteten Merkmalsvektoren $\tilde{\boldsymbol{\Gamma}}$ aus (3.24) keine Reduzierung der Dimensionen.

2. Berechnung der Grenzen für den Face-, Noise- und Nullspace:

Um die Grenzen zwischen den einzelnen Räumen festzulegen, werden die Grenzen zwischen Face- und Noisespace mit dem Faktor a und zwischen Noise- und Nullspace mit b bezeichnet. b wird mit der Anzahl c der Eigenwerte bestimmt, deren Wert Null ist. Folglich gilt für den Grenzwert von Noise- zu Nullspace $b = N - c$. Berechnet werden muss lediglich der Grenzwert zwischen dem Face- und dem Noisespace. Der Faktor a bezeichnet den i -ten Eigenwert, der im Facespace liegt. An der Stelle $a + 1$ befindet man sich im Noisespace. Gleiches gilt für den Grenzwert b .

In (3.18) wird der Median der Eigenwerte λ_{med}^W der Kovarianzmatrix \mathbf{K}_W gesucht. Dabei werden alle von Null verschiedenen Eigenwerte in die Berechnung mit einbezogen.

$$\lambda_{med}^W = \text{median}\{\forall \lambda_i^W | i \leq b\} \quad (3.18)$$

Formel (3.19) beschreibt die Findung des ersten Eigenwertes, der zum Noisespace gehört. Daraus lässt sich die Grenze a ableiten. σ ist eine Konstante, deren Werte sich um Eins bewegt. In dieser Arbeit wird σ mit dem Wert 1 belegt.

$$\lambda_{a+1}^W = \max\{\forall \lambda_i^W | \lambda_i^W \leq \lambda_{med}^W + \sigma(\lambda_{med}^W - \lambda_b^W)\} \quad (3.19)$$

3. Berechnung der regulierten Eigenmerkmale und die Projektion der Ausgangsbilder in den neuen Unterraum:

Die Eigenvektoren der unterschiedlichen Räume Face-, Noise- und Nullspace werden unterschiedlich gewichtet. Dabei werden N Gewichtungsfaktoren ψ für die \mathbf{K}_W berechnet (3.23). Die Anzahl der Bilder aus der Trainingsmenge N ist gleich der Anzahl berechneter Eigenvektoren. Um ψ berechnen zu können, werden in (3.20) und (3.21) die Faktoren α und β berechnet. Mit Hilfe dieser Faktoren wird für jeden Raum mit den Grenzen a und b der Faktor $\tilde{\lambda}_i^W$ berechnet (3.22), der in (3.24) verwendet wird.

$$\alpha = \frac{\lambda_1^w \lambda_m^w (m-1)}{\lambda_1^w - \lambda_m^w} \quad (3.20)$$

$$\beta = \frac{m\lambda_1^w - \lambda_m^w}{\lambda_1^w - \lambda_m^w} \quad (3.21)$$

$$\tilde{\lambda}_k^w = \begin{cases} \lambda_k^w, & k \leq m \\ \frac{\alpha}{k+\beta}, & m \leq k \leq r \\ \frac{\alpha}{r+1+\beta}, & r \leq k \leq n \end{cases} \quad (3.22)$$

$$\psi_i = \frac{1}{\sqrt{\tilde{\lambda}_i^W}}, i = 1, 2, \dots, N \quad (3.23)$$

Die Eigenvektoren \mathbf{w} werden alle mit dem entsprechendem Wert ψ gewichtet und bilden als Spaltenvektoren die Projektionsmatrix $\tilde{\mathbf{W}}_W$. Alle Spaltenvektoren \mathbf{I} der k Klassen werden nun nach (3.24) in den gewichteten Merkmalsvektor $\tilde{\mathbf{I}}$ transformiert. Diese gewichteten Merkmalsvektoren werden als Spaltenvektoren mit regulierten Eigenfeatures bezeichnet.

4. Verwendung der regulierten Eigenfeatures und Berechnung der Projektionsmatrix \mathbf{W}_{ERE} :

Nun werden aus den gewichteten Spaltenvektoren eine neue Kovarianzmatrix nach (3.25) erstellt. Hier kann wieder die Schreibweise von $\tilde{\mathbf{K}}_T$ als Matrixmultiplikation genutzt werden, um wie nach (3.8) die Eigenvektoren zu erhalten. Aus diesen Eigenvektoren wird eine weitere Projektionsmatrix \mathbf{W}_W erstellt. Dabei werden h Eigenvektoren verwendet, deren Eigenwerte am größten sind. Dabei entspricht h maximal N . Aus den beiden Projektionsmatrizen $\tilde{\mathbf{W}}_W$ und \mathbf{W}_T ergibt sich die eigentliche Projektionsmatrix \mathbf{W}_{ERE} (3.26).

$$\tilde{\mathbf{T}}_j^i = \tilde{\mathbf{W}}_W^T \mathbf{T}_j^i, \text{ mit } i = 1, 2, \dots, k \text{ und } j = 1, 2, \dots, N_i^k \quad (3.24)$$

$$\tilde{\mathbf{K}}_T = \sum_{i=1}^k \frac{p_k}{N^k} \sum_{N_j^k=1}^{N^k} (\tilde{\mathbf{T}}_j^i - \boldsymbol{\mu})(\tilde{\mathbf{T}}_j^i - \boldsymbol{\mu})^T \quad (3.25)$$

$$\mathbf{W}_{ERE} = \tilde{\mathbf{W}}_W^T \mathbf{W}_T, \mathbf{W}_{ERE} \in \mathbb{R}^{(n*m) \times h} \quad (3.26)$$

Um nun Bilder mit der Trainingsmenge zu vergleichen, wird wie bei der Eigenface-Methode vorgegangen. Das zu vergleichende Bild wird mittels \mathbf{W}_{ERE} in den Unterraum projiziert, in dem auch alle Bilder der Trainingsmenge vorliegen. Der Unterschied zur Eigenface- und Fisherface-Methode besteht in der Distanzfunktion. Die ERE-Methode bedient sich nicht dem euklidischen Abstandsmaß, sondern verwendet die Kosinusdistanz, definiert in (3.28). Dabei wird das Bild zum Vergleich als \mathbf{T}_{image}^{ERE} bezeichnet, dargestellt durch $y_1, y_2, \dots, y_{(N-h)}$. Für die Bilder der Trainingsmenge gilt $i = 1, 2, \dots, N$:

$$\mathbf{T}_i^{ERE} = \mathbf{W}_{ERE}^T \mathbf{T}_i, \text{ mit } i = 1, 2, \dots, N \quad (3.27)$$

$$d(\mathbf{T}_i^{ERE}, \mathbf{T}_{image}^{ERE}) = \frac{\mathbf{T}_{image}^{ERE T} \mathbf{T}_i^{ERE}}{\|\mathbf{T}_{image}^{ERE}\|_2 \|\mathbf{T}_i^{ERE}\|} \quad (3.28)$$

3.6 Testdurchführung

3.6.1 Kreuzvalidierung

Um die Gültigkeit von multivarianten Modellen zu prüfen, wird die Kreuzvalidierung eingesetzt [Kes07]. Um die Qualität eines Modells zu bewerten, werden die vorhandenen Modelldaten in zwei Gruppen aufgeteilt: Trainingsmenge und Testmenge. Die Testmenge wird auf das Modell, erstellt aus der Trainingsmenge, angewendet. Die Ergebnisse werden mit den tatsächlichen Werten verglichen um diese bewerten zu können. Für die Bewertung der Ergebnisse gibt es verschiedene Klassifikatoren, die in 3.6.2 beschrieben werden. Die Aufteilung in Trainings- und Testmenge wird so lange wiederholt, bis alle Modelldaten Elemente der Testmenge waren. Man unterscheidet bei der Kreuzvalidierung zwischen einfacher, stratifizierter und Leaving-One-Out-Kreuzvalidierung (LOO-Kreuzvalidierung).

1. einfache Kreuzvalidierung:

Die einfache Kreuzvalidierung wird auch als t -fache Kreuzvalidierung genannt. Oft wird die sogenannte 10-fache Kreuzvalidierung angewendet, um ein Modell bewerten zu können. Dabei wird die Trainingsmenge in t Testmengen aufgeteilt, um jeweils die t -te Testmenge gegen die restliche Trainingsmenge zu testen. Entsprechend müssen t Testläufe durchgeführt werden. Die Auswahl der Testmenge aus der Trainingsmenge berücksichtigt nicht die Verteilung innerhalb dieser Menge. Die Fehlerquote der einfachen Kreuzvalidierung ist die durchschnittliche Fehlerquote der t Testläufe.

2. stratifizierte Kreuzvalidierung:

Stratifiziert bedeutet im Zusammenhang der Kreuzvalidierung Erstellung der t Testmengen mit möglichst gleicher Verteilung. Diese Art der einfachen Kreuzvalidierung kann die Varianz der Abschätzung verringern und somit die Fehlerquote verbessern.

3. Leaving-One-Out-Kreuzvalidierung:

Diese Methode bezeichnet eine besondere Art der t -fachen Kreuzvalidierung. t ent-

spricht der Anzahl von zu verwendeten Bildern. Für jeden Testdurchlauf wird lediglich ein Bild als Testmenge verwendet. Dies führt dazu, dass die Leaving-One-Out-Kreuzvalidierung nicht stratifiziert und keine optimale Verteilung angewendet werden kann. Ein weiterer Nachteil ist der weitaus höhere Rechenaufwand, da sich nicht nur die Anzahl der Testdurchläufe, sondern auch die Trainingsmenge vergrößert. Der Vorteil dieser Methode ist die optimale Ausnutzung der vorhandenen Daten in der Testdurchführung.

3.6.2 Klassifikator

Als Klassifikator wird die KNN-Methode⁷ verwendet. Diese Methode klassifiziert einen Merkmalsvektor durch die Betrachtung seiner nächsten Nachbarn im Merkmalsraum. Dabei steht K für die Anzahl der nächstliegenden Nachbarn. Voraussetzung ist ein vorhandenes Abstandsmaß. Entsprechend bedienen sich die vorgestellten Methoden der euklidischen Distanz, bzw. der Kosinusdistanz. Die Klassifikation mit der KNN-Methode kann mittels verschiedener Kriterien durchgeführt werden. Bild 3.4 zeigt exemplarisch das Problem einer Klassifikation; ein zu klassifizierendes Objekt soll anhand ihre Nähe zu einer Klasse zugeordnet werden.

1. Mehrheitskriterium:

Die K nächsten Nachbarn werden ihren jeweiligen Klassen zugeordnet. Deren Anzahl im Verhältnis zu den K Nachbarn bestimmen die Wahrscheinlichkeiten, zu welcher Klasse das zu klassifizierende Objekt gehört. Dabei kann die Wahl von K entscheidend sein. Bei nicht deutlich voneinander getrennten Klassen, kann ein zu groß gewähltes K zu einer schlechten Klassifikation führen. Ebenso ist eine ungerade Anzahl zu betrachtender Nachbarn zu empfehlen, um Gleichstände zu vermeiden.

2. Abstandsgewichtete KNN-Klassifikation:

Im beschriebenen Mehrheitskriterium und in der Positionsgewichtete KNN-Klassifikation spielen die Abstandsverhältnisse der K Nachbarn keine Rolle. Bei sehr durchmischten Klassen kann über die Gewichtung der Distanzverhältnisse eine bessere

⁷KNN ist die Abkürzung für die englische Bezeichnung *K-Nearest-Neighbour*

Klassifikation erzielt werden. Bei dieser Methode der Klassifikation muss K gerade gewählt werden.

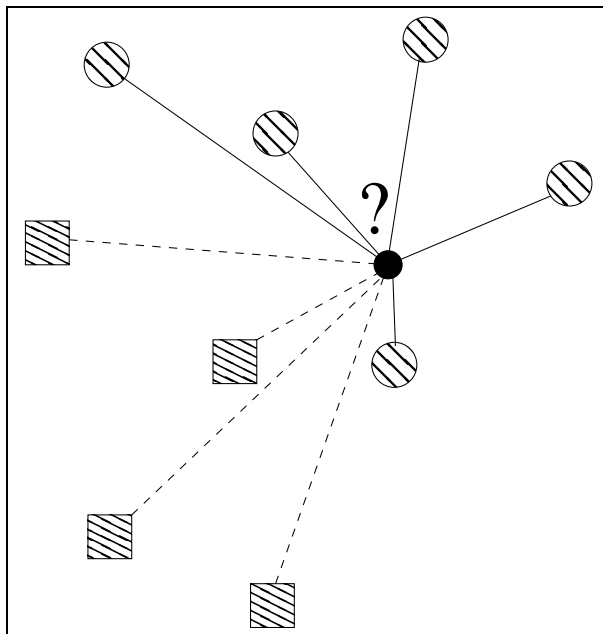


Bild 3.4: Zu lösendes Problem der Klassifizierung.

3.6.3 Receiver-Operating-Characteristics(ROC)-Analyse

Entwickelt wurde diese Methode der Analyse für die Rauschunterdrückung in der Radartechnik bereits im 2. Weltkrieg. Dieses Verfahren ermöglicht das Analysieren und Visualisieren von Diagnosesystemen und wird seit den 60er Jahren des letzten Jahrhunderts in der Medizin eingesetzt. Die ROC-Analyse [Gre03] wird für die Beurteilung der Performance eines Klassifikators eingesetzt [KH95]. Die möglichen Klassifikatoren, angewendet auf die in Abschnitten 3.3, 3.4 und 3.5 beschriebenen Methoden, werden in Abschnitt 3.6.2 aufgezeigt. Seit Mitte der 80er Jahre wird die ROC-Analyse für die Beurteilung von Interpretationen radiologischer Bilddaten genutzt. Eine Diagnose ist eine binäre Einzelentscheidung mit den Aussagen *ist krank* und *ist nicht krank*. In der medizinischen Diagnose stehen die Bezeichnungen *positiv* für die Aussage *der Patient ist krank*, und *negativ* für

der Patient ist nicht krank. Daraus ergeben sich die in Tabelle 3.1 dargestellten bewerteten Diagnoseergebnisse.

Testergebnis	kranker Patient	gesunder Patient
positiv	richtig positiv (rp)	falsch positiv (fp)
negativ	falsch negativ (fn)	richtig negativ (rn)

Tabelle 3.1: Gegenüberstellung Testergebnis und Zustand des Patienten

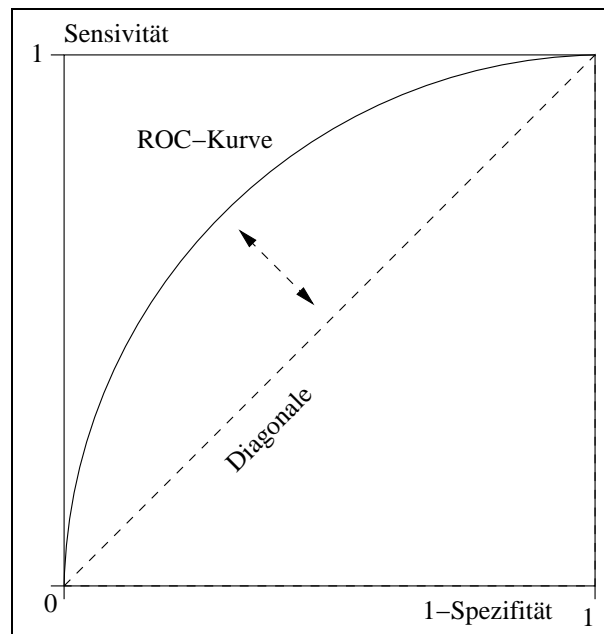


Bild 3.5: Beispiel einer ROC-Kurve.

Die ROC-Analyse beruht auf dem Prinzip, den Anteil richtig positiver Ergebnisse gegenüber den Anteil falsch positiver Ergebnisse zu stellen. Falsch positive Ergebnisse werden als bedingt durch spezifisches Rauschen (Fehlinterpretationen) angesehen und werden aus diesem Grund für verschiedene Schwellwerte berechnet. Diese Schwellwerte werden auch als *Cut-Off-Punkte* bezeichnet. Diese beiden Anteile werden als *Richtig-Positiv-Rate* ((3.29) für Sensitivität) und für *Falsch-Positiv-Rate* ((3.30) 1-Spezifität) betrachtet. Als

Spezifität wird die Richtig-Negativ-Rate bezeichnet (3.31).

$$\text{Richtig-Positiv-Rate} = \frac{rp}{rp + fn} \quad (3.29)$$

$$\text{Falsch-Positiv-Rate} = \frac{fp}{fp + rn} \quad (3.30)$$

$$\text{Richtig-Negativ-Rate} = \frac{rn}{rn + fp} \quad (3.31)$$

In einer ROC-Kurve werden die Wertepaare *Sensivität* und *1-Spezifität* gegenübergestellt. Auf der x-Achse (Ordinate) werden die Werte der Falsch-Positiv-Rate abgebildet, auf der y-Achse (Abzisse) wird die Sensivität aufgetragen. Die Punkte der ROC-Kurve ergeben sich durch das Auftragen der Ergebnisse für alle Cut-Off-Werte eines Messbereichs. Das Maß der Güte von Testergebnissen spiegelt sich in der Fläche unterhalb der ROC-Kurve wider. Dieser Wert wird als AUC-Wert⁸ bezeichnet und nimmt bei geeigneten Methoden einen Wert zwischen 0.5 und 1 an. Bild 3.5 zeigt die Diagonale des dargestellten Wertebereichs der beiden Achsen und eine beispielhafte ROC-Kurve. Die Fläche unterhalb der Diagonalen beträgt 0.5. Ein diagnostischer Test weist Trennschärfe auf, sobald die Kurve stark von der Diagonalen abweicht. Die ideale ROC-Kurve hat eine Fläche von 1. Dies bedeutet hinsichtlich der zu bewertenden Methodik einer Diagnose keine falschen Testergebnisse. Gleichbedeutend ist dies mit einer Sensivität und Spezifität von jeweils 1. Tabelle 3.2 gibt einen möglichen Überblick, welche Aussagen hinter dem AUC-Wert stehen können.

AUC-Wert	Leistungsfähigkeit
< 0.5	Methode der Diagnose ist ungeeignet
= 0.5	Diagnose ist nicht aussagekräftig
> 0.5 und ≤ 0.7	schwache Aussagekraft der Diagnose
> 0.7 und ≤ 0.9	mittelmäßige bis gute Aussagekraft der Diagnose
> 0.9 und < 1.0	sehr gute Aussagekraft der Diagnose
= 1.0	perfekte Aussagekraft der Diagnose

Tabelle 3.2: Überblick der Aussagen verschiedener AUC-Werte(Quelle [Gre03]).

⁸AUC ist die Abkürzung für die englische Bezeichnung *area under curve*

Für die Berechnung der AUC-Werte der in Kapitel 4 beschriebenen Ergebnisse, wird die Software *RocKit*⁹ verwendet. Für die Berechnungen werden die Ergebnisse der Methoden auf die Wahrscheinlichkeit hin untersucht, mit der ein Ergebnis positiv ist. Im idealen Fall würde ein negatives Beispiel durch einen Leaving-One-Out-Test eine Wahrscheinlichkeit positiv zu sein, von 0% aufweisen. Durch dieses Vorgehen werden für alle Bilder die der Kreuzvalidierung unterzogen werden, Werte zugeordnet, aus denen eine ROC-Kurve mittels der Software RocKit erstellt werden kann. Auf diese Kurven beruhen die AUC-Werte.

⁹RocKit ist eine Software die für die Darstellung einer ROC-Kurve und die Berechnung des AUC-Wertes verwendet werden kann. Die Internetseite zum Download dieser Software ist unter <http://www-radiology.uchicago.edu/krl> erreichbar

Kapitel 4

Experimente und Ergebnisse

4.1 Versuchsaufbau und -durchführung

Zur Durchführung der Experimente werden die Ausgangsdaten, Normierungen und Methoden aus Kapitel 3 verwendet. Aus der DDSM-Bilddatenbank werden Mammogramme geladen und deren markierte Regionen als Ausgangsdaten verwendet. Daraus ergibt sich eine Menge benigner und maligner Herdbefunde. Diese werden je nach Durchführung der Experimente unterschiedlich normiert. Auf diese Bildmenge werden die implementierten Methoden angewendet. Dabei wird eine Kreuzvalidierung in Form des *Leaving-One-Out-Testing* durchgeführt, um die Ergebnisse der *ROC-Analyse* zu unterziehen. Die Fläche unterhalb der ROC-Kurve liefert den Wert, der das Ergebnis der Methode und der angewendeten Normierung repräsentiert.

Die Durchführung der Versuche gliedert sich in zwei Schritte: Vorversuch und eigentlich Testdurchführung. Diese beiden Schritte unterscheiden sich im Versuchsaufbau in der verwendeten Kreuzvalidierung. Der Vorversuch verwendet eine 10-fache Kreuzvalidierung. Dieses Ergebnis liefert eine erste Einschätzung der zu erwartenden Ergebnisse einer Methode und deren Normierung. Wenn die Aussagekraft eines Vorversuchs ausreichend ist, wird dieser Versuch der Testdurchführung mittels *Leaving-One-Out-Testing* unterzogen. Die Vorversuche werden dazu verwendet, die Normierungsverfahren auf die Testmenge abzustimmen. Wirken sich Normierungen auf die zu erwarteten Ergebnisse aus, werden

die Bilddaten auf alle Methoden angewendet und für einen Vergleich der ROC-Analyse unterzogen.

Die Auswahl der Bilddaten aus der DDSM-Datenbank geschieht willkürlich. Dies bedeutet, dass keine exemplarische Testmenge erstellt wird, die sich bewusst auf das Ergebniss auswirkt. Testdurchführungen mit exemplarischer Bildmenge, werden in Abschnitt 4.2 gekennzeichnet. Die Beschränkungen in der Auswahl bestehen in der Anzahl der benignen und malignen Mammogramme, damit bei der Testdurchführung ein tendenziell ausgeglichenes Verhältnis von bös- zu gutartigen Herdbefunden vorhanden ist. Die verwendeten Normierungen auf die Ausgangsdaten werden im Abschnitt 3.2 beschrieben.

Für die Testdurchführung werden verschiedene Konstellationen verwendet. Es werden unterschiedliche Anzahlen von ROIs verwendet, verschiedene Bildgrößen und Grauwertnormierungen. Jede ROI wird auf ein quadratisches Seitenverhältnis gebracht. Dabei wird die markierte Region nicht beschnitten, sondern zentriert um die Markierung erweitert.

1. Anzahl der *Regions of Interest*:

Die Anzahl der willkürlich gewählten ROIs wird auf 206 und 409 festgelegt. Von den 206 ausgewählten ROIs sind 114 malign und 92 benign. Die 409 verwendeten ROIs gliedern sich in 202 maligne und 207 benign Herdbefunde. Die exemplarisch erstellte Bildmenge zur Testdurchführung von 95 ROIs gliedert sich in 50 benign und 45 maligne Herdbefunde.

2. verwendete Bildgrößen:

(a) Einheitliche Bildgrößen:

Als Normierung auf einheitlichen Bildgrößen werden Skalierungen aller ROIs auf eine gemeinsame Länge des Bildvektors I bezeichnet. Dabei haben die ausgeschnittenen Regionen der Mammogramme zusätzlich gleiche Seitenverhältnisse. Für die Versuchsdurchführung werden die Bildgrößen 80×80 und 160×160 Pixel verwendet.

(b) Gruppierung nach Bildgrößen:

Die Ausgangsdaten werden nach ihren Bildgrößen in drei Gruppen *Groß, Mit-*

tel und Klein unterteilt. Die Grenzen der einzelnen Gruppen werden so gewählt, dass die Anzahl der Bilder in den Gruppen möglichst gleichbleibend verteilt ist. Die Regionen, deren Seitenverhältnisse kleiner gleich 420×420 Pixel sind, werden in die Gruppe der kleinen ROIs eingeordnet und auf eine Bildgröße von 80×80 Pixeln skaliert. ROIs deren Bildgröße kleiner gleich 600×600 und größer 420×420 Pixeln sind, werden in die Gruppe *Mittel* eingeteilt und auf die Größe 120×120 skaliert. Als *Groß* werden alle ROIs angesehen, deren Bildgröße größer 600×600 Pixel ist. Diese ROIs werden auf die Bildgröße 160×160 Pixel skaliert.

3. Grauwertnormierung:

(a) 8bit-Stretching:

Durch das 8bit-Stretching wird der komplette Grauwertbereich von $[0 \dots 255]$ verwendet. Da die verwendeten Aufnahmen der DDSM-Bilddatenbank im 12bit-Modus vorliegen, werden die Bilddaten in 8bit umgerechnet, bevor das Stretching den kompletten Grauwertbereich $[0 \dots 255]$ abdeckt.

(b) Einfache Kontrastverstärkung:

Neben dem Stretching im gesamten verwendeten Grauwertbereich, wird der Kontrast in den Mammographieaufnahmen verstärkt. Dabei werden Grauwerte ab dem Wert 125 um 10 erhöht, unterhalb des Schwellwertes von 125 um 10 verringert.

4.2 Ergebnisse

Die Bilden 4.1 und 4.2 stellen die ROC-Kurven der drei Methoden Eigen- und Fisherfaces, sowie der ERE-Methode dar. Die ROC-Kurven aus Bild 4.1 zeigen die Ergebnisse der Methoden, angewendet auf Bilddaten der DDSM-Bilddatenbank mit einheitlicher Größennormierung auf die Seitenverhältnisse 80×80 Pixel. Die Trainingsmenge besteht bei allen Testläufen der beiden Bilden aus 206 ROIs. Als weitere Normierung wurde das 8bit-Stretching angewendet. Die Methode Fisherfaces erzielt einen AUC-Wert von 0.5256,

die ERE-Methode einen AUC-Wert von 0.5617 und für die Methode Eigenfaces wird ein AUC-Wert von 0.5779 berechnet. Ergebnisse für die Unterteilung der Trainingsmenge in drei Gruppen anhand der Größen verwendeter ROIs, stellt die Bild 4.2 dar. Die Fisherface-Methode erzielt eine AUC-Wert von 0.5280, der AUC-Wert der ERE-Methode beträgt 0.5305 und die Eigenface-Methode erreicht einen Wert für die Fläche unterhalb der ROC-Kurve von 0.6239.

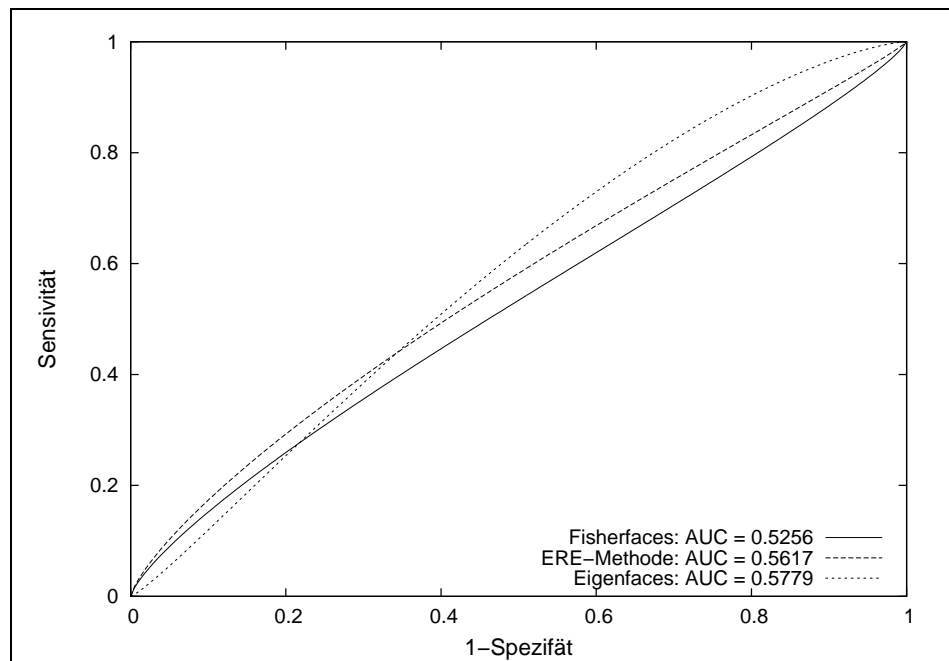


Bild 4.1: Testergebnis mit folgenden Rahmenbedingungen: einheitlicher Skalierung auf die Größe 80×80 Pixel, 206 verwendete ROIs, angewendetes 8bit-Stretching.

Die Tabellen 4.1 und 4.2 zeigen die AUC-Werte der implementierten Methoden. Die Tabellen beinhalten Ergebnisse unterschiedlicher Normierungen. In Tabelle 4.1 sind die Bilddaten auf eine einheitliche Größe von 80×80 Pixeln und mit dem 8bit-Stretching normiert. Im Gegensatz zu den Testdaten aus Bild 4.1 wird der Kontrast zusätzlich verstärkt. Dieser Unterschied zwischen den Testdaten ist auch in den Ergebnissen aus Bild 4.2 und Tabelle 4.2 vorhanden. Beide Tabellen 4.1 und 4.2 repräsentieren die AUC-Werte der implementierten Methoden unter Verwendung unterschiedlicher Anzahlen von Eigenfeatures. Die Eigenfaces erzielen bei einheitlicher Größennormierung und Kontrastverstärkung in Ta-

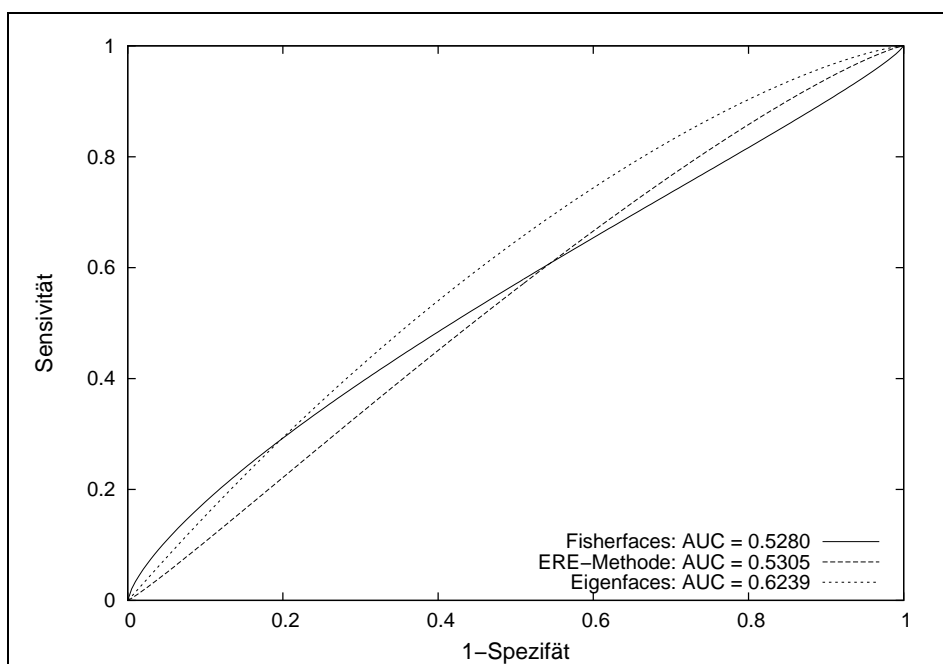


Bild 4.2: Testergebnis mit folgenden Rahmenbedingungen: Gruppierung der Testmenge nach unterschiedlichen Bildgrößen, 206 verwendete ROIs, angewendetes 8bit-Stretching.

belle 4.1 bei 10 und 40 verwendeter Eigenfeatures jeweils den AUC-Werte 0.6017. Fließen alle vorhandenen Eigenfeatures in die Distanzfunktion mit ein, erreicht die Eigenface-Methode einen AUC-Wert von 0.5941. Die Fisherface-Methode erreicht bei 10 verwendeter Eigenfeatures den AUC-Wert 0.5623, die ERE-Methode 0.5496. Die Werte unterhalb der ROC-Kurve bei 40 verwendeten Eigenfeatures betragen für die Fisherfaces und die ERE-Methode 0.5161 und 0.5559. Werden alle vorhandenen Eigenfeatures verwendet, erhält man die Werte 0.5155 und 0.5406 für die Fisherface- und ERE-Methode.

Die Ergebnisse aus Tabelle 4.2 begründen sich in einer nach Größen unterteilten Trainingsmenge. Die Testdaten haben im Vergleich mit der Trainingsmenge aus Bild 4.2 zusätzlich einen erhöhten Kontrast. In der Reihenfolge 10, 40 und alle vorhandenen Eigenfeatures, erzielen die Methoden folgende AUC-Wert: Die Eigenface-Methode erreicht die Werte 0.5916, 0.6000 und 0.5988; die Fisherface-Methode hat AUC-Wert von 0.5152, 0.5016 und 0.5527; die Werte der Fläche unterhalb der Kurve betragen für die ERE-Methode 0.5713, 0.5735 und 0.5422.

Anzahl der Eigenfeatures	AUC-Werte		
	Eigenfaces	Fisherfaces	ERE-Methode
10	0.6017	0.5623	0.5496
40	0.6017	0.5161	0.5559
alle	0.5941	0.5155	0.5406

Tabelle 4.1: Testergebnis mit folgenden Rahmenbedingungen: einheitlicher Skalierung auf die Größe 80×80 Pixel, 206 verwendete ROIs, angewendetes 8bit-Stretching und Kontrastverstärkung. Dabei werden für die Distanzfunktionen unterschiedliche Anzahlen von Eigenfeatures verwendet.

Anzahl der Eigenfeatures	AUC-Werte		
	Eigenfaces	Fisherfaces	ERE-Methode
10	0.5916	0.5152	0.5713
40	0.6000	0.5016	0.5735
alle	0.5988	0.5527	0.5422

Tabelle 4.2: Testergebnis mit folgenden Rahmenbedingungen: Gruppierung der Testmenge nach unterschiedlichen Bildgrößen, 206 verwendete ROIs, angewendetes 8bit-Stretching. Dabei werden für die Distanzfunktionen unterschiedliche Anzahlen von Eigenfeatures verwendet.

Bild 4.3 stellt die Ergebnisse der implementierten Methoden gegenüber, angewendet auf 409 auffälliger Region. Davon sind 202 ROIs malign und 207 ROIs benign. Skaliert werden alle ROIs auf die einheitliche Größe von 80×80 Pixel. Die Fisherface-Methode erzielt einen AUC-Wert von 0.4717, die ERE-Methode hat 0.4913 als AUC-Wert. Die Eigenface-Methode erhält eine Fläche unter der ROC-Kurve vom Wert 0.5839.

Tabelle 5.4 stellt die AUC-Werte der drei implementierten Methoden gegenüber, deren Ergebnisse auf Bilddaten beruhen, die einheitlich auf die Größe 160×160 Pixel skaliert sind. Angewendet sind 8bit-Stretching und eine Kontrastverstärkung. Die AUC-Werte für die Eigenface-, Fisherface und ERE-Methode betragen 0.5119, 0.4895 und 0.4897.

Die Verwendung einer exemplarisch aufgebauten Bilddatenbank beinhalten die Ergebnis-

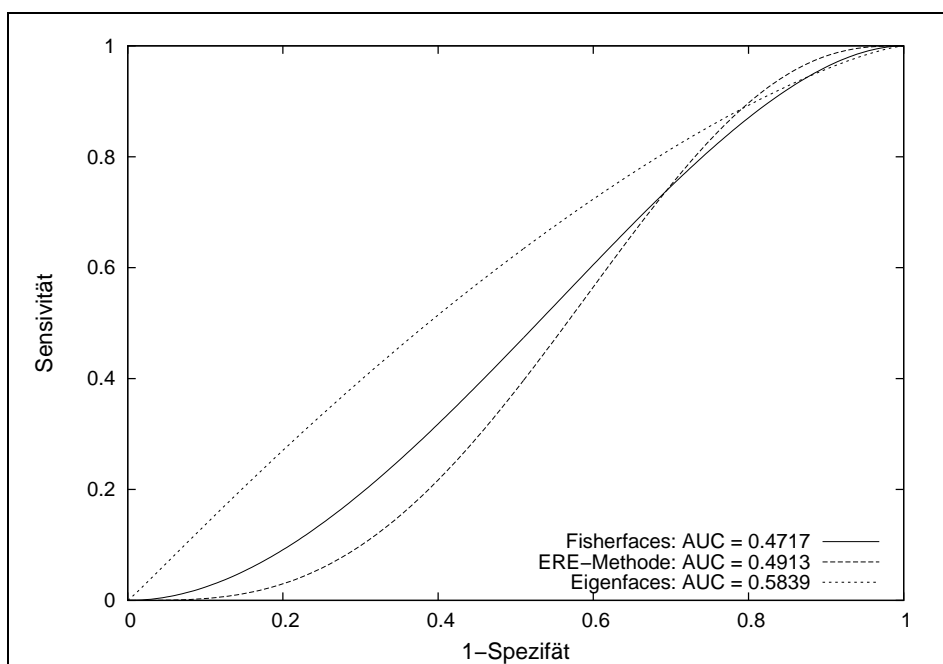


Bild 4.3: Testergebnis mit folgenden Rahmenbedingungen: einheitlicher Skalierung auf die Größe 80×80 Pixel, 409 verwendete ROIs (202 malign/207 benign), angewendetes 8bit-Stretching.

se aus Tabelle 4.4. Die AUC-Werte für die Methoden Eigen- und Fisherfaces, sowie für die ERE-Methode betragen bei einheitlicher Skalierung der Mammogramme 0.5001, 0.4887 und 0.4986. Bei der Unterteilung der Trainingsmenge in drei Gruppen anhand ihrer Bildgröße, wird für die Fisherface-Methode ein AUC-Wert von 0.5604 ermittelt. Die ERE-Methode erzielt einen Wert von 0.5344 und für die Eigenface-Methode beträgt der Wert 0.5264.

AUC-Werte		
Eigenfaces	Fisherfaces	ERE-Methode
0.5119	0.4895	0.4897

Tabelle 4.3: Testergebnis mit folgenden Rahmenbedingungen: einheitlicher Skalierung auf die Größe 160×160 Pixel, 206 verwendete ROIs, angewendetes 8bit-Stretching und Kontrastverstärkung.

Größennormierung	AUC-Werte		
	Eigenfaces	Fisherfaces	ERE-Methode
einheitliche Skalierung	0.5001	0.4887	0.4986
Gruppierung nach Größen	0.5264	0.5604	0.5344

Tabelle 4.4: Testergebnis mit folgenden Rahmenbedingungen: einheitlicher Skalierung auf die Größe 80×80 Pixel und Unterteilung der Trainingsmenge nach Größen der ROIs, die Anzahl verwendeter ROIs beträgt 95 (45 maligne/50 benigne), angewendetes 8bit-Stretching und Kontrastverstärkung.

Kapitel 5

Diskussion

5.1 Fisherface-Methode

Für die Rahmenbedingungen, dass die Trainingsmenge aus willkürlich ausgewählten ROIs besteht, beträgt der höchste AUC-Wert der Fisherface-Methode 0.5623 (siehe Tabelle 4.1). Dabei werden für die Messung der Euklidischen Distanz zehn Eigenfeatures verwendet. Im Vergleich mit den AUC-Werten von 0.5161 und 0.5155 unter Verwendung von 40 und mehr Eigenfeatures zeigt sich, dass sich die Aussagekraft dieser Methode um den Wert 0.05 deutlich unterscheidet. Diese Ergebnisse wurden mittels 8bit-Stretching und einer Kontrastverstärkung erzielt. Zum Vergleich werden die Ergebnisse aus Tabelle 4.2 herangezogen. Bei gleicher Anzahl verwendeter Eigenfeatures erreicht die Fisher-Face-Methode einen Wert von 0.5152. Hier zeigt sich, dass die Einteilung der Bilddaten nach ihren Größen bei der Anwendung der Fisherface-Methode keine Reduzierung der Falsch-Positiv-Rate bedeutet. Eine Reduzierung der Falsch-Positiv-Rate wird erst nach Verwendung aller Eigenfeatures in Tabelle 4.2 erreicht. Hier wird bei der Gruppierung der ROIs ein AUC-Wert von 0.5527 erreicht, was nur im Vergleich mit einheitlicher Größenskalierung und Verwendung aller Eigenfeatures ein besseres Ergebnis bedeutet. Unter Verwendung einheitlicher Skalierung der ROIs erzielt die Fisherface-Methode die besten Ergebnisse. Hinzu kommt, dass die Kontrastverstärkung zusätzlich einen Gewinn bringt. So werden die Bilddaten durch die Verstärkung des Kontrastes bei der Berechnung der Ko-

varianmatrizen am deutlichsten voneinander getrennt. Dadurch wird der Vergleich über die Euklidische Distanz aussagekräftiger. Jedoch sind die erzielten Ergebnisse nicht ausreichend, einen zuverlässigen Diagnosevorschlag zu stellen. Nach Tabelle 3.2 besitzt der Diagnosevorschlag dieser Methode eine schwache Aussagekraft und tendiert sogar zu keiner Aussagekraft.

5.2 ERE-Methode

Die Methode *Eigenfeature Regularization and Extraction* erzielt bei der Gruppierung der Trainingsmenge seinen höchsten AUC-Wert mit 0.5735. Auch hier haben Diagnosevorschläge der Trainingsmengen mit verstärktem Kontrast eine höhere Aussagekraft. Trotzdem bewegen sich die Flächen unterhalb der ROC-Kurve im Wertebereich $[0.53 \dots 0.58]$ und lassen auf eine schwache Aussagekraft der Methode schließen. Festzustellen ist die geringe Auswirkung unterschiedlicher Grauwertnormierungen. Ohne erhöhten Kontrast liefert die Methode einen minimal besseren AUC-Wert. Dies bestätigen die Werte aus Bild 4.1 von 0.5617 und Tabelle 4.1 mit dem maximal erreichten AUC-Wert von 0.5559. Erst die Unterteilung der gesamten Trainingsmenge in drei kleinere Trainingsmengen verstärkt den Unterschied in den AUC-Werten zwischen unterschiedlichen Grauwertnormierungen. Beträgt der AUC-Wert in Bild 4.2 noch 0.5305, hat er sich in Tabelle 4.2 auf 0.5735 erhöht. Hier wirkt sich die Minimierung der Trainingsmenge positiv auf die Falsch-Positiv-Rate aus, wenn der Kontrast erhöht wird.

5.3 Eigenface-Methode

Die AUC-Werte der Eigenface-Methode unterscheiden sich bei einheitlicher Skalierungsgröße und unterschiedlicher Grauwertnormierung deutlich. Beträgt die Fläche unterhalb der ROC-Kurve in Bild 4.1 0.5779, vergrößert sich die Fläche bei erhöhtem Kontrast auf 0.6017 (Tabelle 4.1). Die unterschiedliche Anzahl verwendeter Eigenfeatures hat minimale Auswirkungen auf die Aussagekraft der Diagnosevorschläge. Dies verdeutlichen die AUC-Werte der beiden Tabellen 4.1 und 4.2. Vergleicht man die Flächen unter den

ROC-Kurven der kontrastverstärkten Trainingsmengen, fällt auf, dass die Unterteilung der Trainingsmengen in Gruppen keine signifikante Auswirkung auf die AUC-Werte hat. Den höchsten AUC-Wert erzielt die Methode nach Einteilung der Bilddaten in drei Gruppen mit 8bit-Stretching. Der vergleichbare AUC-Wert aus Tabelle 4.2 ist um den Wert 0.0239 geringer.

5.4 Vergleich der Methoden

Die Skalierung der ROIs auf eine einheitliche Größe hat einen Informationsverlust zur Folge. Die Ergebnisse aus der Tabelle zeigen, dass größere Bilddaten die Ergebnisse der angewendeten Methoden verbessern. Mit einem AUC-Wert von 0.5119 erzielt die Eigenface-Methode das höchste Ergebnis. Die Fisherface-Methode und ERE-Methode erzielen einen AUC-Wert unterhalb von 0.5. Dies bedeutet, dass die Methoden keine Aussagekraft besitzen. Die Bild 4.3 zeigt die Überlegenheit der Eigenface-Methode gegenüber den anderen implementierten Methoden. Bei 409 verwendeten ROIs in der Trainingsmenge, haben die Fisherface- und ERE-Methode keine Aussagekraft. Wie bei größerer Skalierung liegen die AUC-Werte der Methoden unterhalb von 0.5. Die Fläche der ROC-Kurve verwendeter Eigenfaces tendiert auch bei dieser Versuchsdurchführung gegen 0.6 und beträgt 0.5839.

Ein weiterer Versuch die Falsch-Positiv-Rate zu minimieren und die Sensivität zu erhöhen, besteht in der Anwendung der Methoden auf eine exemplarische Bilddatenbank, erstellt mit Mammographieaufnahmen der DDSM-Bilddatenbank. Die Anzahl der Trainingsbilder ist mit 45 malignen und 50 benignen ROIs stark reduziert. Die Bilder wurden nach ihrer subjektiven Eigenschaft ausgewählt, ob die Formen deutlich erkennbar sind. Auch hier erzielen die Methoden, wie in Tabelle 4.4 ersichtlich, keine verbesserten Ergebnisse. Die Reduzierung der Bildmenge um vermeindliche Außreißer wirkt sich nicht positiv auf die Sensivität aus. Als Außreißer werden an dieser Stelle Bilder bezeichnet, die sich grafisch von den ROIs unerscheiden. Diese Unterschiede äußern sich im verstärkten Bildrauschen, schwachen Kontrasten und kaum erkennbaren Formen des verdichteten Gewebes. Stattdessen liegen die Projektionen der Bilder in den niederdimensionalen Merkmalsräumen der Methoden zu nah beieinander, als dass die Klassifizierungen eine hohe Sensivität aufweisen. Die vermeindlichen Außreißer tragen zu einer höheren Sensivität bei.

Stellt man alle drei Methoden gegenüber, hat die Eigenface-Methode die stärkste Aussagekraft. Ebenso weisen die AUC-Werte der angewendeten Eigenface-Methode die geringsten Schwankungen in den unterschiedlichen Versuchsdurchführungen auf. Die größte Reduzierung der Falsch-Positiv-Rate wird mittels Gruppierung nach Größen bei der Eigenface-Methode erzielt. Berücksichtigt man die Implementierung der Fisher- und ERE-Methode, scheint sich die Anwendung der PCA zur Datenreduktion negativ auf die Ergebnisse der Methoden auszuwirken. Beinhaltet die Eigenface-Methode eine einmalige Hauptkomponentenanalyse, werden diese Daten in den anderen beiden Methoden als Ausgangsdatensatz verwendet. Die Erstellung der Kovarianzmatrix über die gesamte Streuung der Bildinformationen, beinhaltet die aussagekräftigste Information. Dies spiegelt sich in den erzielten Ergebnissen wider. Keine andere Methode liefert bei einer bestimmten Versuchsdurchführung aussagekräftigere Diagnosevorschläge als die Eigenface-Methode. Die Fisherface-Methode verwendet zur Lösung des Eigenwertproblems die Kovarianzmatrizen, die die Streuung innerhalb und zwischen den Klassen beschreiben. Die ERE-Methode verwendet als Kovarianzmatrizen lediglich die Streuung innerhalb der Klassen. Vergleicht man die Ergebnisse zwischen diesen beiden Methoden, hat die ERE-Methode die konstant höheren AUC-Werte. Erklären lässt sich dies mit den Eigenschaften der Bildinformationen. Die Verwendung der Streuungsinformationen innerhalb und zwischen den Klassen trennen die Daten nicht eindeutig genug. Aufgrund der Ähnlichkeit und Menge an Merkmalen die ein Bild beschreiben, kommt es durch die Beschreibung der Bilddaten über die Streuung innerhalb der Klassen zur verstärkten Vermischung anstatt zur Trennung der Bildinformationen. Ziel der Methoden Eigen- und Fisherfaces sowie der ERE-Methode ist es, den Informationsgehalt der Bilder in einem niederdimensionalen Raum abzubilden und gleichzeitig wichtige Informationen herauszustellen. Dies gelingt nach den Ergebnissen aus 4 der Eigenface-Methode am besten. Aus diesem Grund hat sie die stärkste Aussagekraft. Die Diagnosevorschläge der Fisherfaces und der ERE-Methode sind nur bedingt mit der Aussagekraft der Eigenfaces vergleichbar. Zwar haben alle drei Methoden das Ergebnis, eine schwache Aussagekraft zu haben, allerdings erzielt die Eigenface-Methode als einzige implementierte Methode einen AUC-Wert um 0.6. Die Fisherface-Methode hat insgesamt die schlechteste Aussagekraft. Die Verwendung der Streuung innerhalb und zwischen den Klassen liefert keine Verbesserung der Ergebnisse. In [BHK97] wird die Verbesserung der Methode gegenüber der Eigenfaces beschrieben, bezogen auf den Be-

reich der Gesichtserkennung. Dabei liegt der Vorteil von Fisherfaces gegenüber den Eigenfaces darin, dass die Falsch-Positiv-Rate bei unterschiedlichen Beleuchtungsverhältnissen und Positionierungen der Gesichter minimiert wird. Das Kriterium Beleuchtung ist übertragen auf die Mammographie vergleichbar mit dem Kontrast einer ROI. Dieser Kontrast ist abhängig von den Rahmenbedingungen der Bildentstehung und des Brustgewebes. Die Positionierung des Herdes in der ROI ist abhängig von der Definition der ROI und von der Normierung der Bilddaten. Die Eigenface-Methode fordert einheitliche Lichtverhältnisse und Positionierung der Gesichtsbilder einer Person. Da die Kriterien bei angewandeter Eigenface-Methode [TP91] nicht erfüllt werden können, wird von der Fisherface-Methode eine Verbesserung der Sensivität erwartet. Die Ergebnisse aus Kapitel 4 zeigen, dass die Methode der Fisherfaces keine erhöhte Sensivität erreicht. [JMK08] zeigt wie [BHK97] eine deutlich geringere Fehlrate in der Gesichtserkennung gegenüber der Eigenface-Methode. Zusätzlich zeigen die von X.Jiang et al. dargestellten Ergebnisse eine niedrigere Fehlrate gegenüber der Fisherface-Methode. Als ein Grund für die verbesserten Ergebnisse wird die Verwendung des Informationsgehalt vom Nullspace (siehe Kapitel 3.5) genannt. Die Ergebnisse aus [JMK08] werden durch die Ergebnisse aus Kapitel 4 dahingehend bestätigt, dass die ERE-Methode der Fisherface-Methode überlegen ist. Jedoch wird keine verbesserte Sensivität gegenüber der Eigenface-Methode erreicht.

Kapitel 6

Zusammenfassung

Ziel dieser Arbeit war die Implementierung und Evaluierung von Methoden zur Diagnose von Herdbefunden in Mammogrammen. Dabei sollen Modelle vorhandener Herdbefunde erstellt werden, die als Basis der Charakterisierung von auffälligen Regionen in Mammogrammen dienen. Die implementierten Methoden stammen aus dem Bereich der Gesichtserkennung. Als Grundlage für die Implementierung dienten die Veröffentlichungen [TP91] für die Eigenface-Methode, [BHK97] für die Fisherface-Methode und [JMK08] für die ERE-Methode. Die Testdurchführung wurde mit Mammographieaufnahmen der DDSM-Bilddatenbank getätigt. Die markierten Regionen in den Mammographieaufnahmen wurden ausgeschnitten und verschieden normiert. Die unterschiedlichen Normierungen der ROIs beziehen sich auf die Größe der Bildausschnitte und auf den verwendeten Grauwertbereich. Für die Evaluation der Ergebnisse und zu Demonstrationszwecken wurde eine Applikation erstellt. Diese Applikation wird im Anhang beschrieben. Die Fragestellung, die es zu beantworten galt, war, ob die Methoden aus dem Bereich der Gesichtserkennung auf Mammographieaufnahmen übertragbar sind. Die Ergebnisse haben gezeigt, dass die Methoden eine geringe Sensivität aufweisen. Auch eine für Testzwecke erstellte Auswahl von Mammogrammen konnte keine Minimierung der Falsch-Positiv-Rate erzielen. Die implementierten Methoden wurden der ROC-Analyse unterzogen. Die Flächen unterhalb der ROC-Kurven dienen als Maß für die Aussagekraft der Methoden. Die erzielten Ergebnisse zeigen, dass alle implementierten Methoden eine schwache Aussage-

kraft haben. Dabei wurden die Erwartungen an die Fisherface- und ERE-Methode nicht erfüllt. Präsentieren [BHK97] und [JMK08] Zahlen, die die Überlegenheit gegenüber der Eigenface-Methode dokumentieren, hat die Eigenface-Methode angewendet auf Herdbefunde in Mammogrammen die höchsten AUC-Werte erreicht. Um die Qualität als entscheidendes Kriterium der schwachen Aussagekraft ausschließen zu können, wurden Grauwertnormierung in der Auswertung von Ergebnissen berücksichtigt. Es zeigte sich, dass die qualitativen Unterschiede der Mammogramme nicht ausschlaggebend für die Ergebnisse sind.

Kapitel 7

Ausblick

Die Charakterisierung von Herdbefunden in Mammogrammen mittels den implementierten Methoden haben nur eine schwache Aussagekraft. Unter Berücksichtigung, dass die Eigenface-Methode die besten Ergebnisse erzielt, kann diese Charakterisierung ein weiteres Element zur Entscheidungsfindung des Diagnosevorschlags eines CADx-Systems sein. Die Normierungen der ROIs beseitigen kein Bildrauschen, verstärken nicht explizit Kanten und beseitigen keine Gewebeüberlagerungen. Diese Normierungen können für eine Verbesserung der Methoden beitragen. Daher sollten die implementierten Methoden mit solchen normierten Bilddaten getestet werden. Ebenso können die Größennormierungen der auffälligen Regionen verfeinert werden. Das quadratische Seitenverhältnis muss nicht das optimale Seitenverhältnis für die Größennormierung sein. In diesem Zusammenhang ist es auch möglich, eine genaue Segmentierung des Herdes als Grundlage zur Größennormierung zu verwenden. Von dieser Methode wird eine Minimierung der Bildinformationen erwartet, die bei angepasster Bildgröße ungewollt hinzukommt.

Zu beachten ist, dass die Testdurchführungen nicht auf die einzelnen Methoden abgestimmt wurden. Es wurden verschiedene Versuchsdurchführungen getätigt, aber keine Normierung verstärkt hinterfragt und auf die verschiedenen Methoden zur Datenanalyse angepasst. Dies ist ein weiterer Ansatz für eine verbesserte Anwendung der Methoden auf Mammographieaufnahmen. Ebenso zu prüfen ist die Aussage, dass die Eigenface-Methode der Fisherface- und ERE-Methode überlegen ist. An dieser Stelle kann ein an-

derer Versuchsaufbau und Klassifizierung die Aussagekraft der Fisherfaces und der ERE-Methode erhöhen.

Literaturverzeichnis

- [BB07] BALL, JE ; BRUCE, LM.: Digital mammographic computer aided diagnosis (CAD) using adaptive level set segmentation. In: *Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE* (2007), S. 4973–4978
- [BHK97] BELHUMEUR, P.N. ; HESPANHA, J.P. ; KRIEGMAN, D.J.: Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection. In: *IEEE Transactions on pattern analysis and machine intelligence, Vol. 19, No. 7* (1997)
- [DHR07] DESTOUNIS, S. ; HANSON, S. ; ROEHRIG, J.: Computer-aided-detection marker value and breast density in the detection of invasive lobular carcinoma. (2007), August, S. 99–104
- [FU01] FREER, T.W. ; ULISSEY, M.J.: Screening mammography with computer-aided detection: Prospective study of 12.860 patients in a community breast center. In: *Radiology 220* (2001), S. 781–786
- [GCF07] *Global Cancer Facts and Figures 2007. 2007*
- [Gre03] GREINER, Matthias: *Serodiagnostische Tests: Evaluierung und Interpretation in der Veterinärmedizin und anderen Fachgebieten*. Springer, 2003. – ISBN 978-3-540-00401-1
- [HBK⁺98] HEATH, Michael ; BOWYER, Kevin ; KOPANS, Daniel ; KEGELMEYER, W. P. ; MOORE, Richard ; CHANG, Kyong ; MUNISHKUMARAN, S.: Current status

- of the Digital Database for Screening Mammography. In: *Digital Mammography* (1998), S. 457–460
- [HBK⁺01] HEATH, Michael ; BOWYER, Kevin ; KOPANS, Daniel ; MOORE, Richard ; KEGELMEYER, W. P.: The Digital Database for Screening Mammography. In: *Proceedings of the Fifth International Workshop on Digital Mammography* (2001), S. 212–218. ISBN 1–930524–00–5
- [JMK08] JIANG, X. ; MANDAL, B. ; KOT, A.: Eigenfeature Regularization and Extraction in Face Recognition. In: *IEEE Transactions on pattern analysis and machine intelligence*, Vol. 30, No. 3 (2008)
- [Kes07] KESSLER, Waltraud: *Multivariate Datenanalyse: Für die Pharma-, Bio- und Prozessanalyse*. WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim, 2007. – ISBN 9783–527–31262–7
- [KH95] KANUNGO, T. ; HARALICK, R.: *Receiver operating characteristic curves and optimal bayesian operating points*. citeseer.ist.psu.edu/kanungo95receiver.html. Version: 1995
- [Nis07] NISHIKAWA, Robert M.: Current status and future directions of computer-aided diagnosis in mammography. In: *Computerized Medical Imaging and Graphics 31* (2007), S. 224–235
- [NOV04] *Mammographie-Fibel für MAMMOMAT 1000/3000 Nova*. 2004. – Bestell-Nr. A91100-M1600-C924-1
- [OMBF06] OLIVER, Arnau ; MARTI, Joan ; BOSCH, Anna ; FREIXENET, Jordi: A new approach to the classification of mammographic masses and normal breast tissue. In: *The 18th International Conference on Pattern Recognition (ICPR'06)*, 2006
- [RAD07] RANGAYYAN, R.M. ; AYRES, F.J. ; DESAUTELS, J.E.L.: A review of computer-aided diagnosis of breast cancer: toward the detection of subtle signs. In: *Journal of the Franklin Institute 344* (2007), S. 312–348

- [RK206] *Krebs in Deutschland, Häufigkeiten und Trends*. 2006. – in Zusammenarbeit mit dem Robert Koch Institut
- [SFR02] SCHMID, E. ; FRIEDLAND, W. ; REGULLA, D.: *Biologische Experimente und Modellrechnungen zur relativen biologischen Wirksamkeit von Mammographie-Röntgenstrahlung*. 2002
- [Tho08] THOMAS, A.: *Das Mammakarzinom: Neue sonographische Techniken zur Dignitätseinschätzung von Herdläsionen*, Fachbereich Humanmedizin, Freie Universität Berlin, Diss., 2008. <http://www.diss.fu-berlin.de/2008/135/index.html>. – Habilitationsschrift Charité Universitätsmedizin Berlin
- [TP91] TURK, M. ; PENTLAND, A.: Eigenfaces for Recognition. In: *J. Cognitive Neuroscience* 3 (1991), Nr. 1
- [WEM⁺67] WINSBERG, F. ; ELKIN, M ; MACY, J. ; BORDAZ, V. ; WEYMOUTH, W.: Detection of radiographic abnormalities in mammograms by means of optical scanning and computer analysis. In: *Radiology vol. 89* (1967), S. 211–215

Anhang A

Verzeichnisse

A.1 Notationstabelle

Kovarianzmatrix	K
Projektionsmatrix	W
Datenmatrix	D
Diagonalmatrix	Σ
Bild	I
Bilderanzahl	N
Anzahl der Eigenwerte die den Wert Null haben	c
Spaltenvektor	Γ
Mittelwertvektor	μ
Streuungsvektor	ϕ
Eigenwert	λ
Eigenvektor	w
Konstante	σ
Anzahl von Bildklassen	k
Anzahl von Bildern der Bildklasse k	N^k
Wahrscheinlichkeitswert	p_k
Anzahl von Testläufen	t
Gewichtungsfaktor	ψ
Anzahl verwendeter Eigenvektoren	h

A.2 Literaturrecherche

Die Methoden zur Anwendung von Herdbefunden in Mammogrammen, beschrieben in Kapitel 3, wurden in diesem Zusammenhang noch nicht publiziert. Die Eigenfacemethode wurde von Oliver et al. für die Klassifikation von möglichen Massen verwendet und in [OMBF06] beschrieben. Zum Thema Mammographie und computer-assistierende Systeme bietet sich die Zeitschrift *Der Radiologe* an. Sie ist eine monatlich erscheinende Fachzeitschrift, die neueste Publikationen angrenzender Fachgebiete zur Radiologie beinhaltet. In dieser Zeitschrift finden sich Artikel zum Thema Computer-Systeme in der Mammographie, Auswertungen von Studien und zum Thema Radiologie selbst. Im Artikel *Computerassistierte Diagnose in der Screeningmammographie*¹ werden verschiedene Verfahren der computer-assistierten Diagnose vorgestellt und die Einsatzmöglichkeiten diskutiert. Die Potentiellen Möglichkeiten, eine Übersicht und eine Einordnung von CAD-Systemen in der Mammographie geben die beiden Artikel *Der Stellenwert von CAD-Systemen*² und *CAD for mammography: the technique, results, current role and further developments*³. Ersterer gibt einen Überblick der heute kommerziell eingesetzten Systeme im Bereich der Mammographie, der Artikel aus *European Radiology* gibt zusätzlich einen Überblick vorhandener Literatur zu diesem Thema und gibt einen Ausblick auf die zu erwartende Entwicklung im Bereich der computer-assistierten Diagnose, bzw. Detektion. Der Artikel *Mammographie*⁴ erklärt die Schwierigkeiten konventionellen Röntgendiagnostik und dient als Einstieg in das Thema Mammographie. Weiterführend ist der Artikel *Stufe-3-Leitlinie Brustkrebs-Früherkennung in Deutschland*⁵. Darin wird der Zusammenhang von Heilungsmethode und Früherkennung aufgezeigt und die Notwendigkeit eines

¹Aus der Zeitschrift *Der Radiologe*, Verlag Springer Berlin/Heidelberg, ISSN 0033-832X (Print), Seiten 72-80, Januar 1996

²Aus der Zeitschrift *Der Radiologe*, Verlag Springer Berlin/Heidelberg, SSN 0033-832X (Print), Seiten 270-274, April 2002

³Aus der Zeitschrift *European Radiology*, Verlag Springer Berlin/Heidelberg, 0938-7994 (Print), Seiten 1449-1460, Januar 2006

⁴Aus der Zeitschrift *Der Radiologe*, Verlag Springer Berlin/Heidelberg, ISSN 0033-832X (Print), Seiten 1114-1123, November 2000

⁵Aus der Zeitschrift *Der Radiologe*, Verlag Springer Berlin/Heidelberg, ISSN 0033-832X (Print), Seiten 495-502, Juni 2003

Mammographie-Screening verdeutlicht. Eingegangen wird in diesem Artikel auf die Früherkennung explizit in Deutschland, deren Stellenwert und Potential. In diesem Zusammenhang steht auch der Artikel *Auswertung der bisherigen Mammographie-Screening-Studien in Europa und in Nordamerika*⁶, der bisherige Studien hinterfragt und sich mit ihrer Gültigkeit befasst. Das Radiologische Institut des Südharz-Hospitals Nordhausen veröffentlichte im März 2008 einen Artikel in *PubMed Central*⁷ zum Thema computer-assistierte Detektion im Bereich der Mammographie mit dem Titel *The performance of computer-aided detection when analyzing prior mammograms of newly detected breast cancers with special focus on the time interval from initial imaging to detection*. Darin wird die Aussagekraft von Computersystemen im Einsatz der Detektion von möglichen Herdbefunden und Mikrokalk analysiert.

Im Januar 2008 erschien beim Georg Thieme Verlag das Buch *Practical Digital Mammography*⁸. Dieses Buch gibt eine Einführung in die Thematik Mammographie, von der Theorie bis hin zu der verwendeten Technik. Ein weiterer Schwerpunkt ist die digitale Mammographie, deren Vergleich zur herkömmlichen Methode, die Verwendung von Röntgenfilm, und die Möglichkeiten dieser Technologie.

A.3 Patentrecherche

Eine Recherche nach möglichen Patenten, die für die Methoden aus Kapitel 3 angemeldet sind, hat keine Treffer geliefert. Es scheinen keine Patente zu geben, die Methoden der multivarianten Datenanalyse zur Diagnose im Bereich der Mammographie beinhalten. Desweiteren gibt es Patente zu computer-assistierten Diagnose, Detektion und Darstellungsmöglichkeiten von digitalen Mammogramme zur Unterstützung von Radiologen. Im folgenden wird eine Auswahl von Patenten vorgestellt, die dem Thema computer-assistierte Diagnose/Detektion zugeordnet werden können.

⁶Aus der Zeitschrift *Der Radiologe*, Verlag Springer Berlin/Heidelberg, ISSN 0033-832X (Print), Seiten 344-351, April 2001

⁷PubMed Central (PMC) ist eine freie Bibliothek des U.S. National Institutes of Health (NIH) von Publikationen zum Thema Biomedizin und weiteren Biowissenschaften.

⁸Das Buch *Practical Digital Mammography* von Beverly Hashimoto hat folgende ISBN-10 Kennziffer: 3-13-148041-6.

Das Patent *US 2002/0181797 A1* beinhaltet eine Methode zur Verbesserung von diagnostischen Befunden digitaler Bilder durch Verwendung von erhöhten Kontrasten. Automatisierte Detektion haben die beiden Patente *P-Nr.: 5579360* und *P-Nr.: 5598481* gemein. Ersteres Patent wird unter dem Titel *Mass detection by computer using digital mammograms of the same breast taken from different viewing directions* geführt. In dieser Methode werden mögliche Massen in jeweils zwei Aufnahmen, gleiches Objekt aufgenommen von verschiedenen Positionen, detektiert. Im darauffolgenden Schritt müssen potentielle Massen aus einer Aufnahme mit einer anderen Masse in der zweiten Aufnahme korrespondieren. Trifft dies nicht zu, wird die Masse nicht weiter berücksichtigt. Ziel dieser Methode ist die Reduzierung der Falsch-Positiv-Rate⁹ Das Patent *P-Nr.: 5598481* beinhaltet eine Methode zur Detektion auffälliger Regionen in Mammogrammen unter Verwendung linearer Musteranalyse und Kantendetektion. Dabei wird unter anderem die Wavelet-Transformation¹⁰ angewendet. *Computer-assisted diagnosis method using correspondence checking and change of salient features in digital images* ist der Titel des Patents *P-Nr.: 6453058*. Darin wird eine Methode zur computer-assistierten Diagnose beschrieben. Verwendet werden in dieser Methode zwei digitalen Aufnahmen, die jeweils in einen relationalen Attributengraph zerlegt werden. Jeder Knoten des Graphen korrespondiert zu einem detektierten auffälligen Merkmal einer Region und die Äste des Graphen decken sich mit der Anordnung der markierten Merkmale der Aufnahmen. Mittels dieser beiden Graphen kann eine computer-assistierte Diagnose gestellt werden. Das Patent *P-Nr.: 6625303* mit dem Titel *Method for automatically locating an image pattern in digital images using eigenvector analysis* beschreibt eine automatisierte Methode zur Auffindung von spezifizierten Strukturen in digitalen Bildern. Unter anderem beinhaltet das Patent eine Klassifizierung gefundener Merkmale von spezifizierten Strukturen mittels Eigenvektoren, um ungewollte Merkmale identifizieren zu können. *P-Nr.: 5729620* beinhaltet ein Patent zu einem computer-assistierten System und die Methode dafür. Das computer-assistierte System besteht aus den nötigen Schritten ein Röntgenbild auf Film zu digitalisieren und Merkmale auffälliger Regionen zu detektieren. Nach diesen Schritten wird das digitalisierte Bild auf einem Monitor präsentiert. Der Radiologe hat so die Möglichkeit, die ursprüngliche Mammographieaufnahme und das digitalisierte Bild mit den detektierten Region gegenüberge-

⁹Die Falsch-Positiv-Rate wird in Kapitel 3 im Zusammenhang der ROC-Analyse beschrieben.

¹⁰Als Wavelet-Transformation wird eine bestimmte lineare Zeit-Frequenz-Transformation bezeichnet.

stellt zu betrachten. Dadurch wird eine verbesserte Diagnose erwarten.

A.4 Demo-Applikation

Die Applikation *MAMMOx* veranschaulicht Teile der Ergebnisse, die mit den drei implementierten Methoden Eigen- und Fisherfaces sowie der ERE-Methode erzielt werden. Aus einem Quellverzeichnis werden die beiliegenden XML-Dokumente ausgewertet und die markierten Regionen der Mammographieaufnahmen ausgeschnitten. Die so entstandenen Bilder werden als Trainingsmenge in den Speicher geladen. Auf dieser Trainingsmenge lassen sich die implementierten Methoden anwenden. Nach angewandeter Eigenface-Methode lassen sich in der Applikation die ersten zwölf Eigenfaces anzeigen. Mit dieser Trainingsmenge lässt sich ein explizit ausgewähltes Bild vergleichen. Angezeigt werden die Bilder der Trainingsmenge, die dem Vergleichsbild am nächsten liegen. Zusätzlich ermöglicht die Applikation die geladenen Bilder der Trainingsmenge und das einzelne Vergleichsbild anzuzeigen. Bei angewandeter Eigenface-Methode lässt sich das Vergleichsbild im ursprünglichen Zustand gegenübergestellt mit dem rekonstruierten Vergleichsbild anzeigen. Dafür wird das Vergleichsbild in den Unterraum projiziert, der mit der Eigenface-Methode, und angewendet auf die Trainingsmenge, aufgebaut wurde. Mittels den vorhandenen Eigenwerten und den zugehörigen Eigenvektoren wird das Bild rekonstruiert und dargestellt. Der Benutzer hat in der Applikation die Möglichkeit, die ersten 20 Eigenwerte in ihrem Einfluss auf die Rekonstruktion zu gewichten. Dadurch kann der Informationsgehalt bestimmter Eigenvektoren und Eigenwerte visualisiert werden.

Neben diesen Ansichten zur Demonstration der Methoden, können über die Applikation Leaving-One-Out-Tests und Kreuzvalidierungen angestoßen werden. Die erzielten Ergebnisse der Testläufe werden in Textdateien geschrieben und gespeichert. Diese Dateien werden zur Auswertung der Testläufe verwendet. Bild A.1 zeigt die Applikation. Dargestellt wird in dieser Abbildung das geladene Vergleichsbild und die Ergebnisse der drei ähnlichsten Bilder der Trainingsmenge. Desweiteren sind die Auswahlmöglichkeiten der implementierten Methoden und Testverfahren ersichtlich. Die Demo-Applikation *MAMMOx* ist auf der beiliegenden CD vorhanden. Zusätzlich befindet sich im Ordner der für Windows kompilierten Applikation eine Software-Dokumentation.

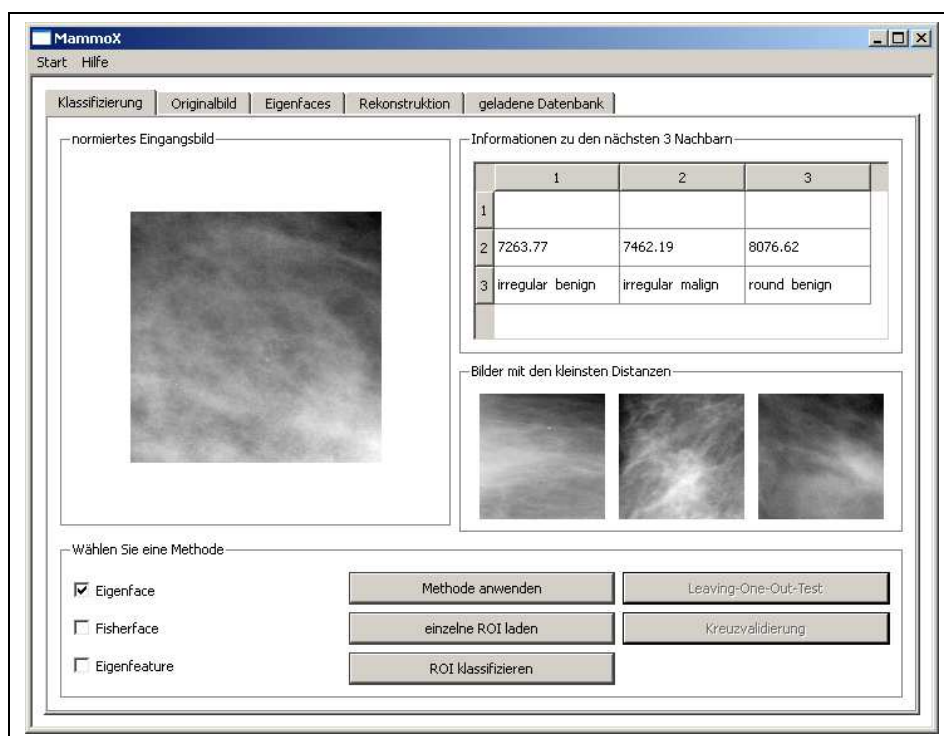


Bild A.1: Grafische Oberfläche der Applikation MammoX.

A.5 Verwendete Bibliotheken

Für die Implementierung der Methoden unter Windows wurden folgende Bibliotheken verwendet:

1. vxl-1.9.0
2. boost-1.34.1
3. dcmtk-3.5.4
4. qt-win-opensource-src-4.3.3
5. gil (boost)