

ALBERT-LUDWIGS-UNIVERSITÄT
FREIBURG
INSTITUT FÜR INFORMATIK

Lehrstuhl für Mustererkennung und Bildverarbeitung
Prof. Dr. Hans Burkhardt



Object classification

using

Local Relational Features

Studienarbeit

Margret Keuper

16.2.2006 – 15.6.2006

Erklärung

Hiermit erkläre ich, dass die vorliegende Arbeit von mir selbstständig und nur unter Verwendung der aufgeführten Hilfsmittel erstellt wurde.

Freiburg, den 15.06.2006

15. Februar 2006

**Aufgabenstellung für die Studienarbeit
von Frau Margret Keuper**

Object Classification using Local Relational Features

Haar integrals are an elegant way to generate features which are invariant against arbitrary desired transformation groups [1]. Of particular research interest is the proper choice of the kernel function appropriate for the problem in hand. It was shown in [2] that the so called relational kernel functions are suited for applications where robustness against illumination changes is desired.

This Project should focus on these Relational features, calculated not for the whole objects, but rather around so called interest points. Thus, feature extraction should be viewed as a 2-stage process: 1) Extraction of interest points, and 2) Extraction of features around these points. The two stages are not necessarily independent of each other. Matching of the features extracted from different objects is also a very important design decision for the overall classification accuracy.

The experiments should be carried out on the *Muscle-Coin* Database which contains coins of almost 600 different types. The images depend heavily on the condition of different parts of the coins, thus suggesting the use of relational features.

Literatur

- [1] H. Schulz-Mirbach. Invariant features for gray scale images. In G. Sagerer, S. Posch, and F. Kummert, editors, *17. DAGM - Symposium "Mustererkennung"*, pages 1-14, Bielefeld, 1995. Reihe Informatik aktuell, Springer.
- [2] M. Schael. Invariant grey scale features for texture analysis based on group averaging with relational kernel functions. *Internal Report 1/01, Albert-Ludwigs-Universität Freiburg, Institut für Informatik*, January 2001.

Referent: Prof. Dr.-Ing. H. Burkhardt

Betreuer: Lokesh Setia, M.Sc.

Ausgabedatum: 16. Feb 2006

Abgabedatum: 15. Juni 2006

Bearbeitungszeit: 4 Monate (Halbtags)

.....
Prof. Dr.-Ing. H. Burkhardt

Inhaltsverzeichnis

1	Einleitung und Motivation	3
2	Die Datenbank	4
3	Segmentierung und Vorverarbeitung	6
4	<i>Interest Point</i> Detektoren	9
4.1	Harris Corner Detector	9
4.2	Wavelet-basierte Salient Point Detektoren	10
5	Relationale Kernfunktionen	12
5.1	Invariante Merkmale	12
5.2	Relationale Kerne und ihre Anwendung	12
6	Kantenorientierte Verfahren - die Hausdorff Distanz	14
7	Klassifikation	17
7.1	Kullback-Leibler Divergenz	17
7.2	Der <i>k-Nearest Neighbor</i> Klassifikator	18
7.3	Support Vector Machine	18
8	Implementierung	20
8.1	Interest Point Detektoren	20
8.2	Relationale Kernfunktion	20
8.3	Hausdorff Distanz	21
8.4	Support Vector Machine	21
9	Ergebnisse	22
9.1	Relationale Kernfunktionen	22
9.1.1	Auswertung auf <i>Interest Points</i>	22
9.1.2	Auswertung auf <i>Salient Points</i>	23
9.2	Hausdorff Distanz	26
10	Schluss	28
	Literatur	29
A	Confusion Matrix	30
B	Beispiele: Verwechselte Münzen	31
C	Motivserien	32

1 Einleitung und Motivation

Der Fokus dieser Arbeit soll auf den sogenannten lokalen relationalen Merkmalen liegen. Damit bezeichnet man Grauwertmerkmale, die nicht von den Grauwerten eines Pixels direkt abgeleitet werden, sondern sich aus der Differenz von Grauwerten benachbarter Pixel ergeben.

Diese relationalen Merkmale sollen auf einer Datenbank aus Bildern von Münzen ausgewertet werden. Abbildung 1 zeigt Vorder- und Rückseiten zweier Münzen der gleichen Klasse aus der Datenbank, beides österreichische fünf Schilling Stücke. In den beiden linken Bildern ist die Beleuchtung wesentlich stärker als in den rechten.



Abbildung 1: Vorder- und Rückseite zweier Münzen aus der Datenbank.

Neben kleineren Unterschieden wie der Jahreszahl (auf den Vorderseiten der Münzen), oder durch Verschmutzung und Abnutzung entstandenen Unregelmäßigkeiten, sind die Rotation und die unterschiedliche Beleuchtung die Hauptunterschiede zwischen Bildern einer Klasse. Dies macht die Datenbank interessant für die Arbeit mit lokalen relationalen Merkmalen. Weil diese Merkmale nicht die Grauwerte selbst sondern ihr Verhältnis zueinander beschreiben, können sie streng monotone Variationen in den Grauwerten ausgleichen. Leichte Beleuchtungsunterschiede können solche Variationen hervorrufen.

Beim Betrachten der Datenbankelemente fällt auf, dass ein Großteil der Fläche einer Münze wenig Informationen über deren Klassenzugehörigkeit enthält. Die Informationen sammeln sich um die Kanten im Bild, an den Stellen, an denen in der echten Münze ein Profil heraustritt. Daher erscheint es sinnvoll, die relationalen Merkmale nicht auf den ganzen Münzaufnahmen auszuwerten, sondern nur in der Umgebung einiger wichtiger Punkte. Diese Punkte werden als *Interest Points* oder auch *Salient Points* bezeichnet. Zur Bestimmung dieser *Interest Points* existieren schon mehrere Verfahren, von denen ein passendes ausgewählt werden musste. Zwei der wichtigsten Verfahren, nämlich der *Harris Corner Detector* und der auf Haarwavelets basierende *Salient Point Detector* wurden auf der Datenbank getestet und die Ergebnisse, die erzielt werden konnten, miteinander verglichen.

Letztlich hat sich die Frage gestellt, ob sich die relevanten Informationen zur Klassifikation von Münzen überhaupt in den Grauwerten befinden, oder ob es nicht reicht oder sogar sinnvoller ist, die Kanten eines Bildes zu untersuchen. Das hat zu einem alternativen Ansatz geführt, der Klassifikation der Münzen mittels der Hausdorff Distanz. Diese ist ursprünglich für Punktemengen definiert, lässt sich aber auf Binärbilder übertragen. Die Punktemenge jedes Bildes besteht dabei aus den Pixeln, die den Wert eins haben. Die Hausdorff Distanz bietet den Vorteil, dass sie gegenüber Verschiebungen von Punkten um wenige Pixel sehr robust ist.

2 Die Datenbank

Durch die Währungsumstellung von den europäischen Währungen zum Euro im Januar 2002 mussten große Mengen an Münzen der alten Währungen an die Nationalbanken der teilnehmenden Länder zurückgegeben werden. Diese Gelegenheit wurde von Wohltätigkeitsorganisationen genutzt, um zu Geldspenden aufzurufen. Diese sind als eine unsortierte Menge von unterschiedlichen Währungen gesammelt worden. Im Rahmen der Muscle Benchmarking Initiative wurde von der österreichischen Firma ARC Seibersdorf research GMBH die *Coin Images Seibersdorf (CIS)- Benchmark* aufgestellt und mit den gesammelten Daten ein Wettbewerb ausgeschrieben. Auf der Basis von Dicke, Durchmesser und Grauwertbildern von Vorder- und Rückseite von Münzen sollen diese klassifiziert werden.

Die Datenbank enthält insgesamt 100000 Bilder aus 2270 Klassen, die in eine Trainings- und eine Testmenge unterteilt sind. Dabei wurden alle Trainingsbilder mit der gleichen Kamera aufgenommen, die Testbilder stammen von zwei verschiedenen Kamerasystemen, die aus je zwei Kameras bestehen. Die Dicke und der Durchmesser der Münzen wurde jeweils mit einem Lichtsensor gemessen.

Jede Klasse enthält zwischen einem und 30 Trainingsbeispielen. Die *Ground Truth* ist auf mehrere Arten codiert. Einerseits ist sie in der unteren linken Ecke der Bilder zu lesen, andererseits sind die Trainingsbilder nach Klassennamen gespeichert. Ein Beispiel für einen solchen Namen ist: 1202-03F1. Dabei steht 1202 für den Typ "Belgium 1 Franc". 03F1 ist eine Klasse der möglichen Vorderseiten dieser Münze. Sie muss von der Klasse 01F1 unterschieden werden, aber nicht von 03F0.



Abbildung 2: Drei belgische 1 Franc Münzen. Die linke und mittlere dürfen in die gleiche Klasse sortiert werden. Sie gehören aber nicht der gleichen Klasse an, wie die rechte Münze.

Zusätzlich sind Textdateien gegeben, die über die Klassenzugehörigkeit, Durchmesser und Dicke der Testmünzen informieren. Für die Trainingsmünzen ist in einer Textdatei der durchschnittliche Durchmesser und die durchschnittliche Dicke von Münzen einer Klasse angegeben.

In der Datenbank liegen die Bilder nicht segmentiert vor. Die Vorder- und Rückseiten der Münzen sind vor dem Hintergrund einer schwarzen Textur aufgenommen. Sie sind in beliebiger Richtung rotiert und ihre Position im Bild ist unbestimmt. Für jede Trainingsklasse ist zusätzlich ein Durchschnittsbild aller Elemente angegeben, auf dem die Münze ausgerichtet und zentriert abgebildet ist.

Ein etwas über 10000 Bilder umfassender Teil der Datenbank ist auch getrennt herunterzuladen (Vgl. [11]). Er enthält Bilder von 1100 Testmünzen und 4381 Trainingsmünzen aus insgesamt 345 Klassen. In der Testmenge liegen auch Bilder von Münzen vor, die in keine

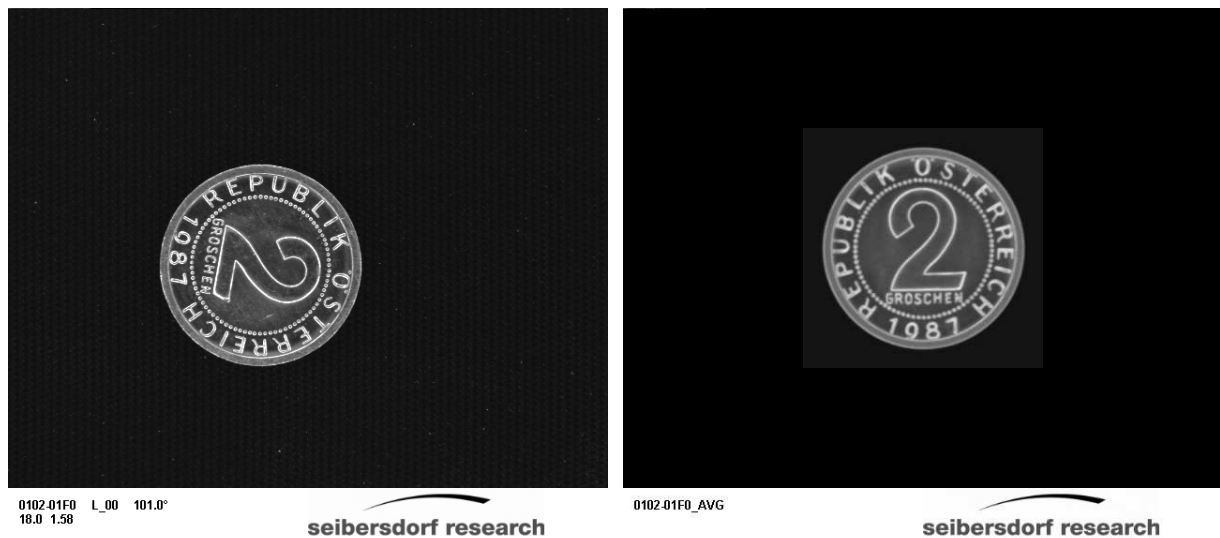


Abbildung 3: (a) Ein Element aus der Datenbank. Die unteren 640x64 Pixel zeigen das Logo von Seibersdorf und Informationen über die Klassenzugehörigkeit der aufgenommenen Münze. (b) Das Durchschnittsbild der Klasse 0102-01F0.

der Trainingsklassen gehören. Sie sollen als solche erkannt und aussortiert werden. Auch gibt es Beispiele, bei denen die Aufnahmen von Vorder- und Rückseite nicht zusammengehören. Auch diese sollen aussortiert werden.

3 Segmentierung und Vorverarbeitung

Da die Münzen fast aller Klassen rund sind, bietet sich zu ihrer Segmentierung die Houghtransformation für Kreise an. Dabei geht man vom Gradienten des Originalbildes aus - in diesem Fall erzeugt mit dem Sobeloperator. Der Sobeloperator ist ein Kantenoperator, der auf den Differenzen von Mittelwerten beruht, das heisst auf eine gewichtete Mittelwertbildung folgt eine Differenzbildung. Damit lässt sich der Sobeloperator folgendermaßen darstellen.

$$\begin{array}{ccc}
 & -1 & 0 & 1 \\
 SOBM = & -2 & 0 & 2 \\
 & -1 & 0 & 1
 \end{array}
 \qquad
 \begin{array}{ccc}
 & 1 & 2 & 1 \\
 SOBN = & 0 & 0 & 0 \\
 & -1 & -2 & -1
 \end{array}
 \qquad (1)$$

In den Dimensionen des Gradientenbilds wird danach eine Akkumulatormatrix erstellt. Anstelle der Houghtransformation wurde eine vereinfachte Variante implementiert, bei der die Richtung des Gradienten unberücksichtigt bleibt. In der Akkumulatormatrix werden dabei alle diejenigen Punkte inkrementiert, welche von einem Punkt im Gradientenbild den Abstand r haben. Der Abstand r entspricht dem Radius des gesuchten Kreises. In der Akkumulatormatrix entsteht bei der Wahl des richtigen Radius ein Maximum an der Stelle, an der das Zentrum eines Kreises im Originalbild liegt.

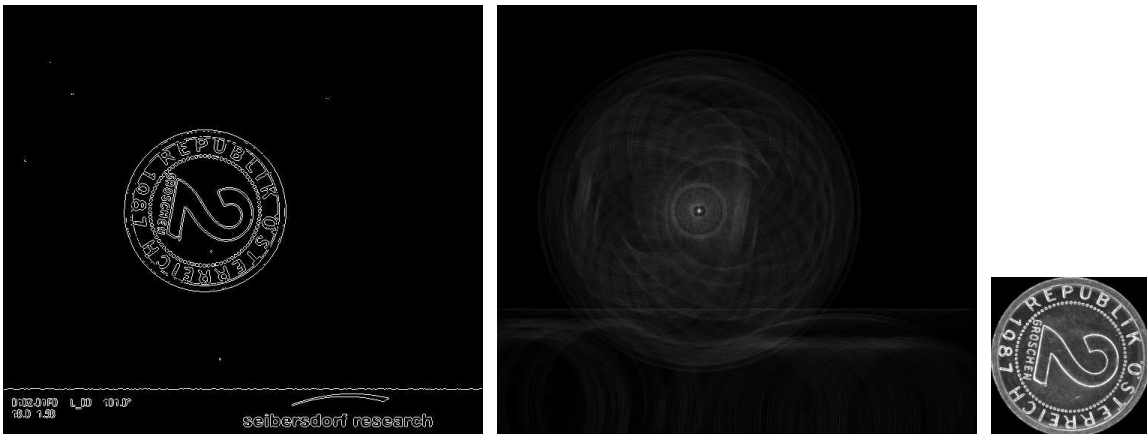


Abbildung 4: (a) Anwendung des Sobeloperators auf das Bild aus Abbildung 3(a). (b) Die Akkumulatormatrix. (c) Die segmentierte Münze.

In der Datenbank sind für alle Klassen von Münzen der minimale und maximale, auftretende Radius angegeben. Diese Werte sind leider nicht besonders genau. Das zeigt sich darin, dass die Minimal- und Maximalwerte nicht immer für Vorder- und Rückseite der selben Münze übereinstimmen. Allerdings erleichtern die Angaben die Suche nach dem richtigen Radius, so dass ein Multiskalenansatz zur Beschleunigung nicht notwendig ist. Der bei der Segmentierung gefundene, genaue Radius andererseits ist ein wichtiges Merkmal zur Klassifikation der Münzen, da diese in den Aufnahmen alle in der gleichen Auflösung vorliegen. Nur Münzen, die den gleichen Radius haben, können derselben Klasse angehören. Die Information über den Radius kann später also als Vorwissen in die Klassifikation mit eingehen.

In den Ecken der Bilder bleibt nach der Segmentierung ein Rand, in dem der Hintergrund sichtbar ist. Um eine Abbildung zu erhalten, auf der nur die Münze zu sehen ist, wird diese nun in Polarkoordinaten transformiert.

$$r = \sqrt{x^2 + y^2} \qquad (2)$$

$$\varphi = \arctan \frac{y}{x} \qquad (3)$$

Für die Arbeit mit der Hausdorff Distanz hat dies auch den Vorteil, dass nicht Rotations-, sondern lediglich Translationsinvarianz in eine Richtung gewährleistet werden muss.



Abbildung 5: Die Münze nach der Transformation in Polarkoordinaten.

Auf den Flächen der Münzen sind oft neben der Prägung auch Abnutzungspuren sowie kleinere Kratzer zu sehen, die für die Klassifikation nicht relevant sind. Daher wurde ein Tiefpassfilter, genauer ein 5×5 Medianfilter auf die Münzen angewendet. Dieses nichtlineare Filter hat die Eigenschaft, dass es scharfe Kanten nicht verschleift, wie es lineare Tiefpassfilter tun. Impulsartige Störungen im Bild wie *salt and pepper noise* können allerdings sehr effektiv mit dem Medianfilter beseitigt werden. Bei der Bestimmung der *Interest Points* (Vgl.



Abbildung 6: Die Münze nach der Medianfilterung.

Kap. 4) hat sich gezeigt, dass speziell der *Harris Corner Detector* auf Datensätzen mit starken Kontrastunterschieden sehr unzuverlässig funktioniert. In kontrastarmen Bildern werden meist keine wichtigen Punkte gefunden. Daher wurde zur Bestimmung der *Interest Points* mit den *Harris Corner Detector* ein Histogrammausgleich durchgeführt. Da es im Datensatz häufig der Fall ist, dass die Bilder an den Stellen, an denen die Prägung der Münzen hervortritt, besonders hell sind, hat sich der klassische Histogrammausgleich mit einer linearen Transformation als ungünstig erwiesen. Die hellen Stellen, die für das Erkennen der Prägung besonders wichtig sind, werden verwaschen, die dunkleren Regionen, in denen vor allem Kratzer und Abnutzungspuren zu sehen sind aber, werden besonders kontrastreich dargestellt. Daher wurde eine Art Histogrammausgleich implementiert, der eine nichtlineare Transformation verwendet. Der helle Bereich wird dabei stärker gespreizt als der dunkle.

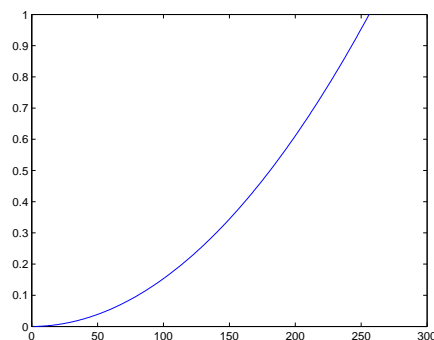


Abbildung 7: Für den Histogrammausgleich verwendete nichtlineare Transformation.

Nach dieser Vorverarbeitung sind in den Bildern, die vorher kontrastarm waren, wesentlich zuverlässiger *Interest Points* entlang der Prägung gefunden worden.

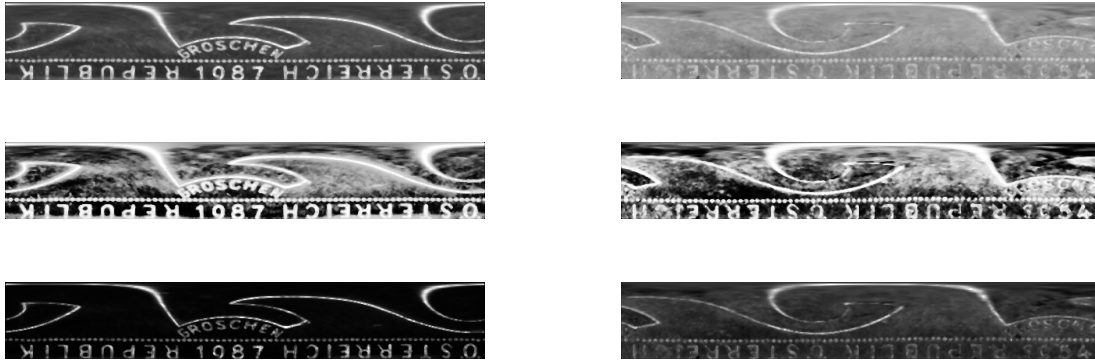


Abbildung 8: Beispiele für die Histogrammtransformationen an zwei Münzaufnahmen, eine mit starkem Kontrast (oben links), und eine mit schwachem Kontrast (oben rechts). Bei der klassischen Histogram Equalization (mittlere Zeile) wird der Kontrast vor allem in den flächigen Regionen der Münze stärker, bei der modifizierten, nichtlinearen Variante, wird die Prägung als das wichtigste Element im Bild hervorgehoben.

Allerdings wurden die so vorverarbeiteten Bilder lediglich zur Bestimmung der *Interest Points* verwendet, nicht aber für die Auswertung der relationalen Kernfunktion.

4 Interest Point Detektoren

Im Bereich der Robotik nimmt die Arbeit mit *Interest Points* einen hohen Stellenwert ein. Ziel ist es hier, korrespondierende Punkte auf unterschiedlichen Aufnahmen derselben Umgebung wiederzufinden und zu *matchen*, um eine 3D Repräsentation einer Szene zu erstellen. Es gibt jedoch auch Motivationen für die Arbeit mit *Interest Points* im Bereich der inhaltsbasierten Bildsuche. Einerseits spart die Arbeit mit wenigen Punkten aus einem Bild Rechenzeit, andererseits kann man sogar hoffen, durch die Auswahl der für die Wahrnehmung wichtigsten Punkte in einem Bild einen besonders diskriminativen Index zu erstellen und damit das Klassifikationsergebnis zu verbessern.

Welche allerdings die Punkte sind, die in einem Bild die meisten Informationen enthalten, ist nicht klar definiert. Der *Harris Corner Detector* geht beispielsweise davon aus, dass sich die wichtigsten Informationen in einem Bild auf den Ecken befinden, weil diese die Gestalt eines im Bild enthaltenen Objektes am besten beschreiben.

4.1 Harris Corner Detector

Ecken in einem Bild sind definiert als die Punkte, an denen der Gradient in mehreren Orientierungen sehr groß ist. Der *Harris Corner Detector* basiert also auf lokalen Ableitungen.

$$M = \exp \frac{x^2+y^2}{2\sigma^2} \cdot \begin{pmatrix} \frac{\delta^2}{\delta x^2} I & \frac{\delta^2}{\delta x \delta y} I \\ \frac{\delta^2}{\delta y \delta x} I & \frac{\delta^2}{\delta y^2} I \end{pmatrix} \quad (4)$$

Die Ecken werden dann mit

$$R = \text{Det}(M) - k \text{tr}^2(M) \begin{cases} R > \epsilon & \text{Ecke} \\ -\epsilon \leq R \leq \epsilon & \text{keine Kante} \\ R < -\epsilon & \text{Kante} \end{cases} \quad (5)$$

bestimmt. Dabei muss für die Entscheidung, ob der Betrag von R an einer Stelle ausreichend groß ist, um als eine Kante oder Ecke in Frage zu kommen, ein Schwellwert ϵ festgelegt werden. Auch die Parameter σ und k müssen festgelegt werden.

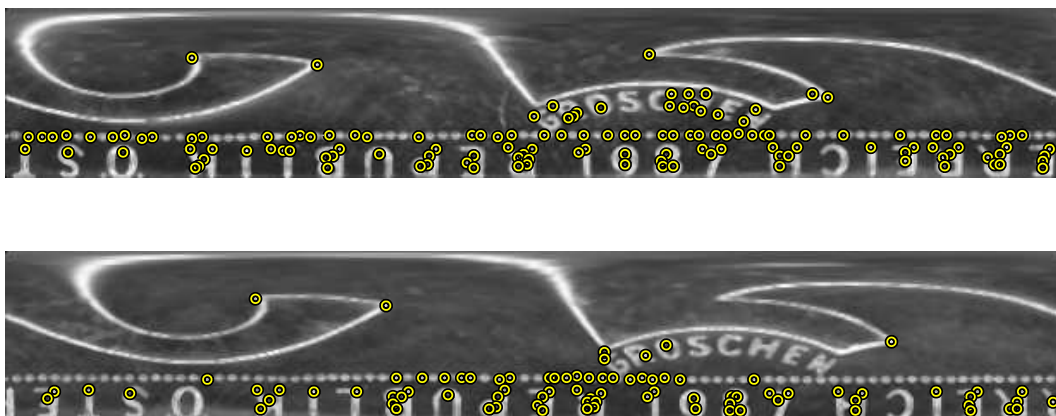


Abbildung 9: Beispiele für *Interest Points*, bestimmt mit dem *Harris Corner Detector*. Die Punkte liegen größtenteils wie erwartet auf den Ecken.

4.2 Wavelet-basierte Salient Point Detektoren

Loupas und Sebe [5] stellen einen *Salient Point Detector* vor, der auf der Wavelettransformation beruht. Diese Repräsentation eines Bildes gibt Auskunft über die Variationen im Bild bei unterschiedlichen Skalierungen. Die so bestimmten *Salient Points* liegen nicht unbedingt auf Ecken sondern zum Beispiel auch in weichgezeichneten Übergangsregionen. Die Autoren hoffen so, eine vollständigere Repräsentation eines Bildes durch *Salient Points* zu gewährleisten.

Wavelets sind wellenförmige, beschränkte, mittelwertfreie Funktionen, die außerhalb eines bestimmten Intervalls verschwinden. Von einem Mutterwavelet $\Psi_{0,0}(t)$ wird durch Skalierung und Verschiebung eine Familie von Wavelets gebildet, die ein Erzeugendensystem der L^2 bilden. Eine orthonormale Basis erhält man, indem man sich auf Zweierpotenzschritte in Skalierung und Translation beschränkt.

$$\Psi_{j,k}(t) = 2^{j/2} \Psi_{0,0}(2^j t - k) \begin{cases} j \in \mathbb{Z} \\ k = 0, \dots, 2^j - 1 \end{cases} \quad (6)$$

Im Folgenden werden nur die Signale betrachtet, für die $j \leq -1$ gilt. Der Waveletkoeffizient einer Funktion f bei der Skalierung 2^j ist dann $W_{2^j} f(n) = \langle f(t), \Psi_{j,n}(t) \rangle$. Für die Auflösung 2^j werden $2^j N$ Waveletkoeffizienten berechnet, wobei N die Anzahl der *Samples* im Signal f ist. Eine wichtige Waveletfamilie sind die Haar-Wavelets:

$$\Psi_{0,0}(t) = \begin{cases} +1 & \text{für } 0 \leq t < \frac{1}{2} \\ -1 & \text{für } \frac{1}{2} \leq t < 1 \end{cases} \quad (7)$$

Sie bilden mit der Skalierungsfunktion $\phi_{0,0}$ ein orthonormales und vollständiges System. Jeder

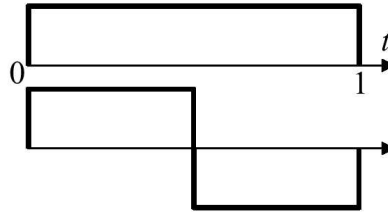


Abbildung 10: Oben: Die Skalierungsfunktion $\phi_{0,0}$, Unten: Das Mutterwavelet der Haarwavelettransformation. Aus [2].

Koeffizient der Haar-Wavelettransformation wird aus 2^{-j} Signalpunkten berechnet, wobei jeder Punkt pro Skalierung in die Berechnung genau eines Waveletkoeffizienten eingeht. Ein Beispiel für die Haartransformation ist in Abbildung 11 gegeben. *Scale 1* ist das ursprüngliche Signal. Der Koeffizient $W_{1/4} f(0)$ ist aus den gleichen *Samples* gebildet, wie $W_{1/2} f(0)$ und $W_{1/2} f(1)$: $C(W_{1/4} f(0)) = \{W_{1/2} f(0), W_{1/2} f(1)\}$. Die Koeffizienten $W_{1/2} f(0)$ und $W_{1/2} f(1)$ sind die *Kinder* von $W_{1/4} f(0)$.

Allgemein sind die Kinder eines Haarkoeffizienten gegeben als:

$$C(W_{2^j} f(n)) = \{W_{2^{j+1}} f(k), 2n \leq k \leq 2n + 1\}, 0 \leq n < 2^j N \quad (8)$$

Der Koeffizient für die Skalierung 2^j berechnet die Variation des Signals bei dieser Skalierung. Seine Kinder geben die Variation in einer bestimmten Teilmenge dieser Punkte an. Die wichtigste Teilmenge von Punkten ist diejenige, die den höchsten Waveletkoeffizienten in der Skalierung 2^{j+1} aufweist. Wendet man diesen Schritt rekursiv an, so kann man den Punkt mit der höchsten *Saliency* zurückverfolgen. Er erhält folgenden *Saliency*-Wert:

$$saliency = \sum_{k=1}^{-j} |C^{(k)}(W_{2^j} f(n))|, 0 \leq n < 2^j N, -J_{max} \leq j \leq -1, \quad (9)$$

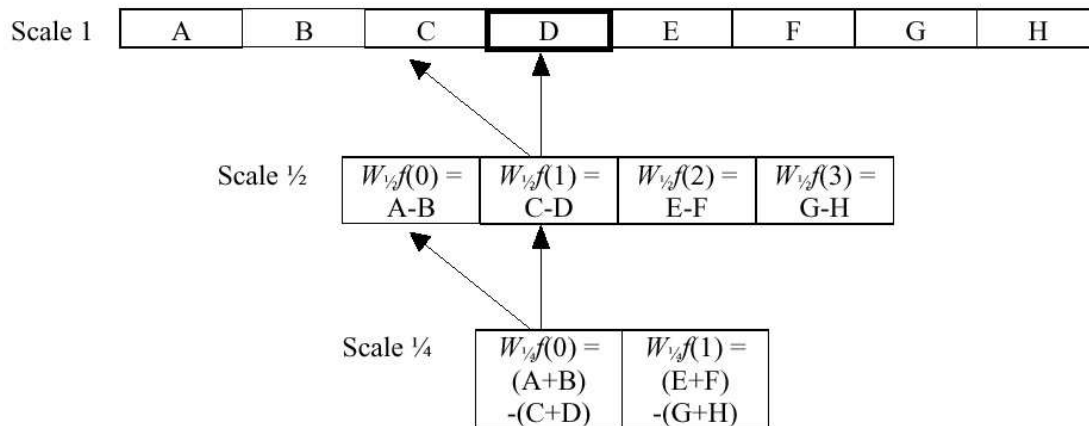


Abbildung 11: Die Waveletkoeffizienten der Haartransformation. Aus[2].

wobei $J_{max} = \log_2 N$. Somit werden Punkte, die globale Variationen repräsentieren, von Wavelets mit grober Auflösung erstellt und haben daher eine hohe *Saliency*. Punkte, die feinere Variationen repräsentieren, werden bei höheren Auflösungen detektiert und haben daher einen geringeren *Saliency*-Wert. Eine detailliertere Darstellung der Wavelet basierten *Salient Point Detektoren* ist in [5] gegeben. Abbildung 12 zeigt für vier Beispiele aus der Datenbank die 120 Punkte mit den höchsten *Saliency*-Werten. Die gefundenen Punkte liegen nicht nur auf den Ecken, sammeln sich aber auf den Kanten im Bild.

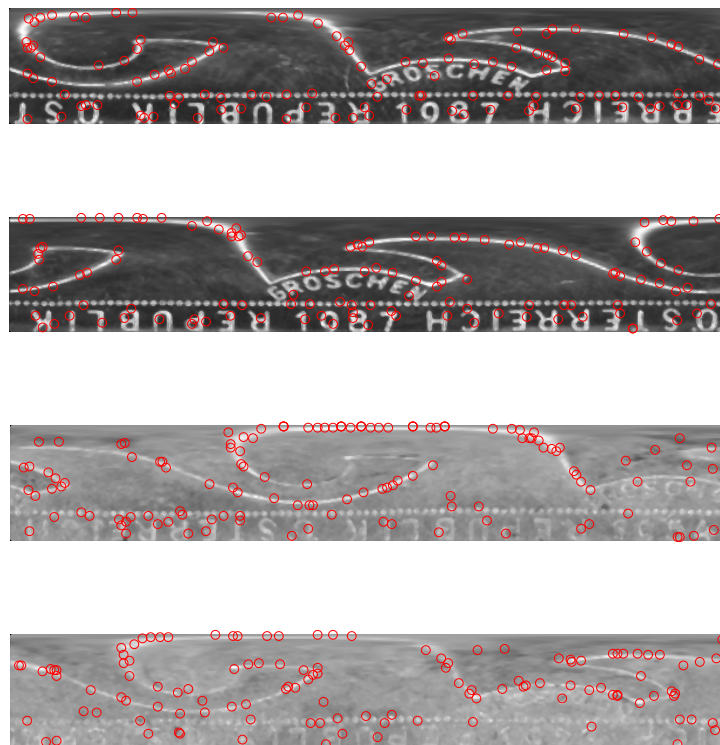


Abbildung 12: Die 120 Punkte mit höchstem Saliency Wert für vier 2 Groschen Münzen.

5 Relationale Kernfunktionen

Relationale Kernfunktionen wurden bisher erfolgreich bei der Klassifikation von Texturen angewendet [8]. Sie sind nicht nur invariant gegenüber zweidimensionalen Euklid'schen Transformationen sondern auch robust gegenüber Grauwerttransformationen mit streng monoton wachsenden Funktionen. Damit bieten sie Invarianz gegenüber Beleuchtungsunterschieden bei diffuser Beleuchtung. Diese Tatsache macht die relationalen Kerne auch interessant für die Klassifikation von Münzen. Die spekularen Beleuchtungsunterschiede, die im Münzdatensatz auftreten, können allerdings mittels der relationalen Kerne nicht ausgeglichen werden - sie sind nicht streng monoton wachsend.

5.1 Invariante Merkmale

Zur Erstellung invarianter Merkmale für zweidimensionale Grauwertbilder stehen verschiedene Methoden zur Verfügung. Schulz-Mirbach[9] beispielsweise stellten eine Möglichkeit vor, solche Merkmale für Grauwertbilder M mittels Integration über einer endlichen, kompakten Transformationsgruppe G zu berechnen.

$$A[f](M) := 1/|G| \int_G f(gM) dg \quad (10)$$

Um rotations- und translationsinvariante Grauwertmerkmale zu erhalten wird hierbei ein Haarintegral über der Gruppe der zweidimensionalen Euklid'schen Transformationen gebildet. Die Berechnung kann, handelt es sich um Kernfunktionen, die lokal ausgewertet werden, in zwei Schritte unterteilt werden: die Auswertung der Kernfunktion für jeden Pixelwert, und das Summieren über die dadurch entstandenen Zwischenwerte.

Eine weitere Möglichkeit, invariante Merkmale zu generieren, bietet die Arbeit mit *fuzzy histograms* (FZH) über den Werten, die sich bei der Auswertung der Kernfunktion ergeben. Mit FZH lässt sich der Nachteil des diskontinuierlichen Verhalten der klassischen Histogramme vermeiden.

In meiner Arbeit sollen die relationalen Kernfunktionen nicht auf dem ganzen Bild sondern lediglich auf den *Salient Points* ausgewertet werden. Dabei ergeben sich je nach Anzahl der *Salient Points* unterschiedlich große Merkmalsmatrizen. Zur Klassifikation sollen dann die statistischen Eigenschaften dieser Merkmale verwendet werden. Dies wird in Kapitel 7.1 genauer beschrieben.

5.2 Relationale Kerne und ihre Anwendung

Die relationalen Kernfunktionen (RKF) haben sich aus den *local binary pattern* (LBP) entwickelt. Diese bilden das Verhältnis von einem Grauwertpixel zu seiner 3x3-Nachbarschaft auf ein binäres Muster ab. Der Nachteil daran ist, dass das resultierende Muster nicht stabil ist. Eine RKF hingegen bildet die Differenz δ zwischen zwei Grauwerten auf eine reelle Zahl ab. So kann ein kontinuierlicheres Merkmal gebildet werden. Für einen Threshold $\epsilon \geq 0$ gibt [8] ein Beispiel für eine einfache RKF an.

$$rel(\delta) = \begin{cases} 1 & \text{für alle } \delta < -\epsilon \\ \frac{1}{2\epsilon}(\epsilon - \delta) & \text{für alle } -\epsilon \leq \delta \leq \epsilon \\ 0 & \text{für alle } \epsilon < \delta \end{cases} \quad (11)$$

Für $\epsilon > 0$ ist dieses Merkmal nicht mehr invariant gegenüber streng monoton wachsenden Grauwerttransformationen, immerhin aber noch recht robust. Es ist invariant gegenüber additiven Grauwerttransformationen.

Diese RKF wird nun auf den Grauwerten ausgewertet, die sich in verschiedenen Radien um einen Bildpunkt herum befinden. Dazu werden mit $p_i = \{r_i, \phi_i\}, i \in \mathbb{N}$ Kreise um den Bildpunkt definiert, wobei $r_i \in \mathbb{R}_+$ der Radius und $\phi_i \in [0, 2\pi[$ die Phase des jeweiligen Kreises i sind. Die *sampling rate* bestimmt, wie viele Grauwerte auf den Kreisen um einen Punkt herum verglichen werden.

Abbildung 13 zeigt ein Beispiel für zwei Kreise $p_1 = \{1, \pi/8\}$ und $p_2 = \{2, 7\pi/8\}$. In diesem Fall wurde als *sampling rate* 16 gewählt, wobei Werte, die nicht auf dem Raster liegen, interpoliert werden müssen.

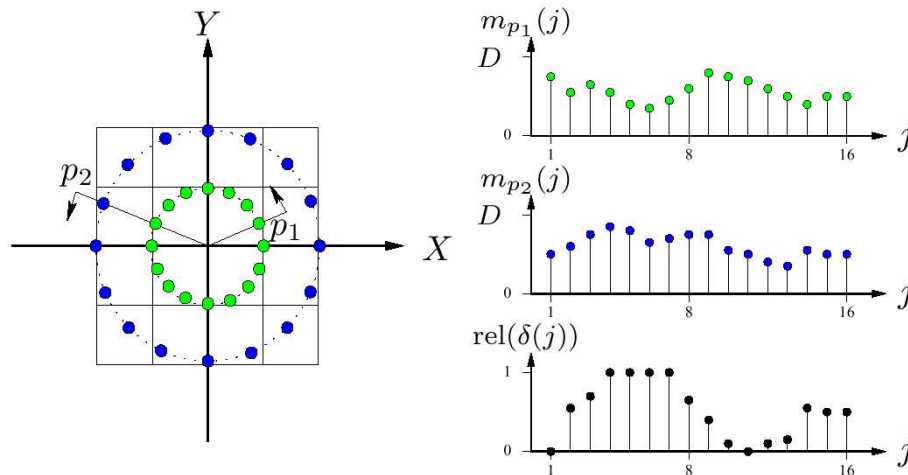


Abbildung 13: Relationale Kernfunktionen. Aus[8].

Die 16 Grauwerte m_{p_i} , die sich pro Kreis ergeben, sind auf der rechten Seite abgebildet, die Differenzen $\delta(j) = m_{p_1}(j) - m_{p_2}(j)$ werden von $rel(\delta)$ auf das Intervall $[0,1]$ abgebildet. Um Rotationsinvarianz zu erhalten wird dann über die Werte von $rel(\delta)$ der Mittelwert gebildet. Bei der Verwendung von n verschiedenen Kreisen mit $n \geq 2$ können so $\alpha = n(n - 1)/2$ Merkmale pro Bildpunkt berechnet werden. Abbildung 14 zeigt ein Beispiel für die Anwendung der relationalen Kernfunktionen. Die leichten Beleuchtungsunterschiede in den beiden Aufnahmen der Münzen werden durch die Kernfunktion ausgeglichen.

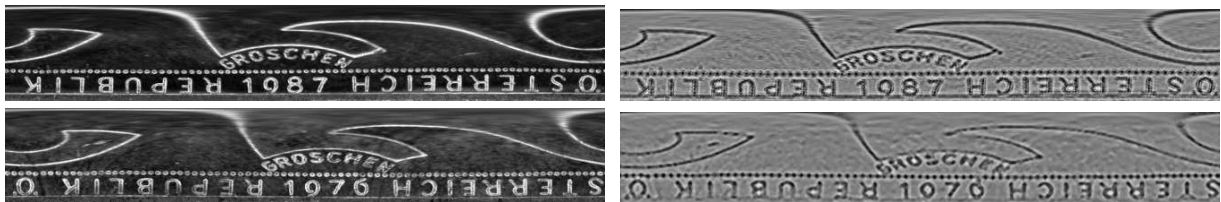


Abbildung 14: Zwei unterschiedlich beleuchtete Münzen in Polarkoordinaten, und die Bilder, die nach Anwendung einer relationalen Kernfunktion daraus entstehen. Angewendet wurde die Kernfunktion aus (11) mit $\epsilon = 25$ auf die Grauwertdifferenzen zwischen $p_1 = \{1, 0\}$ und $p_2 = \{2, 0\}$.

Für jedes Bild ergibt sich somit eine zweidimensionale Merkmalsmatrix mit $\alpha \times m$ Stellen. m ist hierbei die Anzahl der Salient Points, für welche die relationale Kernfunktion ausgewertet wird.

6 Kantenorientierte Verfahren - die Hausdorff Distanz

Für die Bilder aus dem Datensatz stellt sich die Frage, ob die Arbeit auf Grauwerten zu ihrer Klassifikation überhaupt adäquat ist. Ein Großteil der Informationen zu einer Münze scheinen schon in ihrem Kantenbild zu liegen.

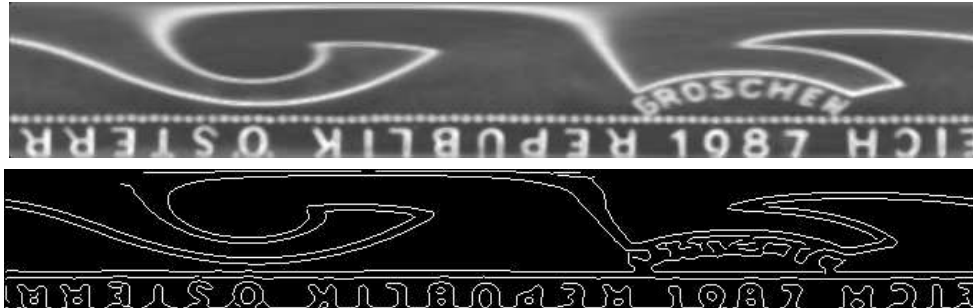


Abbildung 15: (oben) Polarbild aus der Klasse 0102-01F0 (unten) dazugehöriges Kantenbild, erstellt mit dem *Canny Edge Detector*.

Eine Möglichkeit, die Ähnlichkeit zweier übereinandergelegter Binärbilder zu bestimmen, bietet die Hausdorff Distanz. Für zwei Punktemengen $A = \{a_1, a_2, \dots, a_p\}$ und $B = \{b_1, b_2, \dots, b_q\}$ ist diese definiert als

$$H(A, B) = \max(h(A, B), h(B, A)) \quad (12)$$

mit

$$h(A, B) = \max_{a \in A} \min_{b \in B} \|a - b\|, \quad (13)$$

für eine Norm $\|\cdot\|$. $h(A, B)$ heißt die gerichtete Hausdorff Distanz von A nach B. Für den Punkt in A, der am weitesten von allen Punkten in B entfernt ist, wird die Distanz zum nächsten dieser Punkte bestimmt.

Die Hausdorff Distanz misst somit die Unterschiedlichkeit zweier Punktemengen, die eine bestimmte Position zueinander haben. Will man diese Distanz auf Bildern messen und dabei Invarianz gegenüber einer bestimmten Transformationsgruppe G bewahren, ist man interessiert an der Minimalen Hausdorff Distanz

$$M_G(A, B) = \min_{g_1, g_2 \in G} H(g_1 A, g_2 B) \quad (14)$$

Entspricht $\|\cdot\|$ der euklid'schen Norm und G der Gruppe der euklid'schen Transformationen, so ist dies äquivalent zu

$$M_G(A, B) = \min_{g \in G} H(A, gB). \quad (15)$$

Seien $d(x) = \min_{b \in B} \|x - b\|$ und $d'(x) = \min_{a \in A} \|x - a\|$. Dann gilt:

$$H(A, B) = \max(\max_{a \in A} d(a), \max_{b \in B} d'(b)). \quad (16)$$

Der Graph von $d(x)$ heißt auch *Voronoi* Oberfläche von B . Er entspricht für ein Binärbild $B[k, l]$ einer *Distance Transform* $D[x, y]$, da er zu jedem Punkt x seine Distanz zum nächsten Punkt aus B liefert. Punkte in der Voronoi Oberfläche, die besonders hell sind, repräsentieren einen besonders großen Abstand zum nächsten Punkt aus der Punktemenge B . Da die Münzen in Polarkoordinaten vorliegen, brauchen wir die Minimale Hausdorff Distanz



Abbildung 16: Euklid'sche *Distance Transform* zu dem Beispiel aus Abbildung 6.

nur über der Gruppe der Translationen zu betrachten. Für eine Translation t ergibt sich die Hausdorff Distanz als:

$$\begin{aligned}
 H(A, B + t) &= \max(\max_{a \in A} \min_{b \in B} \|a - (b + t)\|, \max_{b \in B} \min_{a \in A} \|a - (b + t)\|) \\
 &= \max(\max_{a \in A} \min_{b \in B} \|(a + t) - b\|, \max_{b \in B} \min_{a \in A} \|a - (b + t)\|) \quad (17) \\
 &= \max(\max_{a \in A} d(a - t), \max_{b \in B} d(b + t)).
 \end{aligned}$$

Für Binärbilder $A[k, l]$ und $B[k, l]$ ergibt sich daraus eine Hausdorff Distanz Matrix F , die an jeder Stelle $[x, y]$ die Hausdorff Distanz zwischen A und B für die Translation $t = [x, y]$ enthält.

$$F[x, y] = \max(\max_{k, l} A[k, l]D[k - x, l - y], \max_{k, l} B[k, l]D'[k - x, l - y]) \quad (18)$$

D und D' sind dabei die *Distance Transforms* von B und A . Eine modifizierte Definition der Hausdorff Distanz ist gegeben mit:

$$H(A, B) = \max(\sum_{a \in A} \min_{b \in B} \|a - b\|, \sum_{b \in B} \min_{a \in A} \|a - b\|). \quad (19)$$

Indem man die Maximumbildung in den gerichteten Hausdorff Distanzen durch die Summation über die Minima ersetzt, erhält man ein Distanzmaß, das weniger anfällig gegenüber Rauschen ist, als die klassische Hausdorff Distanz. Sie kann dazu verwendet werden, die Translation t zu finden, mittels derer eine Münze auf eine andere *gematcht* werden kann. Die Hausdorff Distanz zu dieser Translation bietet ein Ähnlichkeitsmaß für die beiden Münzen, aber auch andere Ähnlichkeitsmaße sind denkbar. So geben auch die Anzahl der Punkte in beiden Bildern, die tatsächlich identisch sind, Auskunft über die Ähnlichkeit, sowie auch die Anzahl der Punkte, die sich nicht überdecken.

[6] gibt einen Algorithmus an, der die Berechnung der Hausdorff Distanz beschleunigt. Er nutzt dabei verschiedene Eigenschaften der Distanz aus. Zum Beispiel die Tatsache, dass die Hausdorff Distanz als Funktion der Translation nicht schneller als linear abnehmen kann, ermöglicht es, für einen Wert $F[x, y] = v$ mit $v > \tau$ die Berechnung aller Werte im Radius $v - \tau$ um $[x, y]$ zu überspringen, falls man den Wert τ als zu hoch erachtet. Die Bestimmung eines solchen Schwellwerts τ beschleunigt die Berechnung enorm. Je niedriger allerdings τ gewählt wird, desto ungenauer ist das erzielte *Matching*.

Abbildung 17 zeigt Beispiele für die Resultate, die mittels der modifizierten Hausdorff Distanz erzielt werden können. Für jeweils zwei Kantenbilder sind die gefundenen *Matchings* abgebildet. Das jeweils linke Binärbild ist in der Überlagerung der beiden Bilder in grün dargestellt, das rechte in rot. Auch bei verrauschten Bildern funktioniert das Verfahren relativ gut. Gegenüber Verschiebungen und Rauschen von wenigen Pixeln ist die Hausdorff Distanz relativ stabil. Beim Erstellen der Beispiele wurde der Wert $\tau = 10$ gewählt.

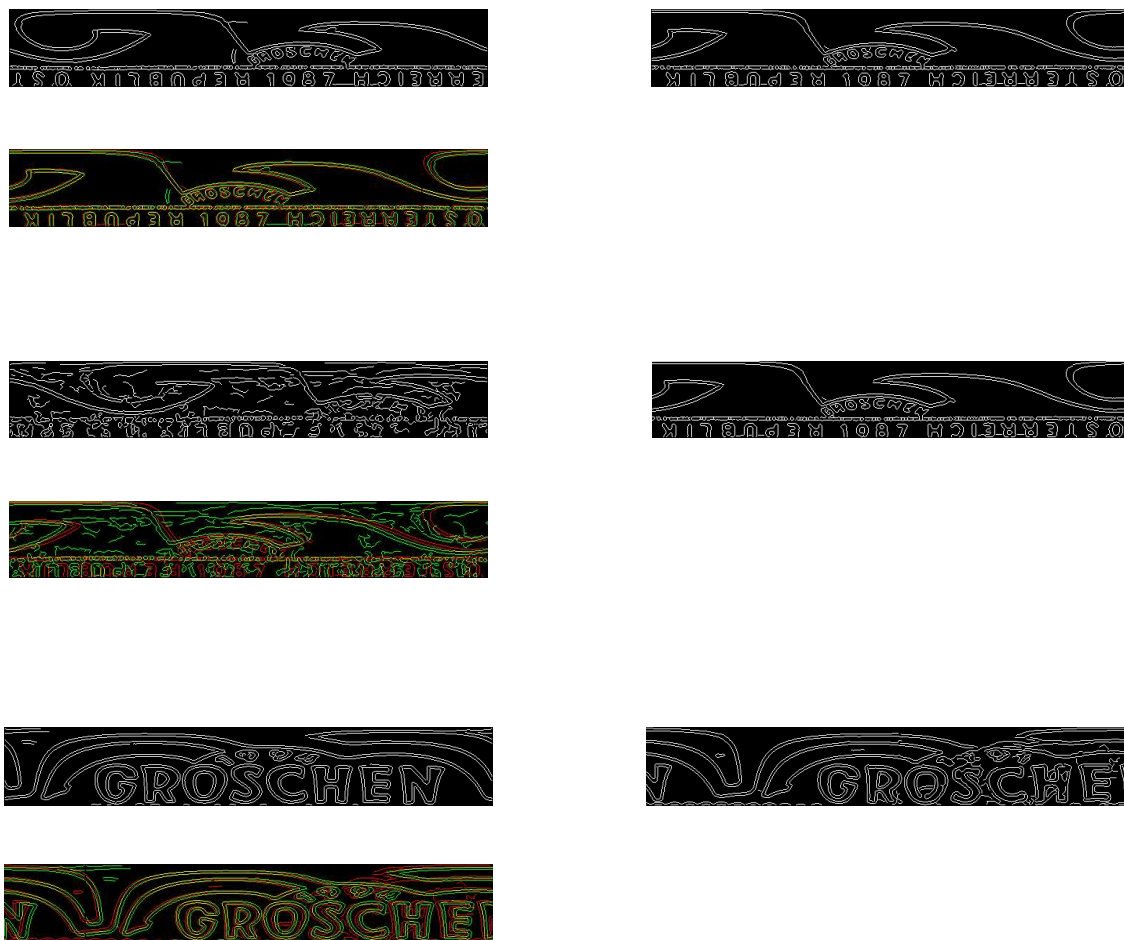


Abbildung 17: Jeweils zwei Kantenbilder, extrahiert mit dem *Canny Edge Detector*, und darunter die durch die modifizierte Hausdorff Distanz gefundene Überdeckung.

7 Klassifikation

7.1 Kullback-Leibler Divergenz

Eine Möglichkeit, basierend auf Wahrscheinlichkeitsdichtefunktionen (PDF, *probability density function*) ein Distanzmaß zu generieren, stellen Moreno, Ho und Vasconcelos [3] vor. Für die Daten wird hierzu ein stochastisches Modell $p(x|\theta_i)$ erstellt, d.h. die Parameter θ_i der PDF eines Datums $X_i = \{x_1, x_2, \dots, x_m\}$ werden geschätzt. Als effektive PDF gerade zur Klassifikation von Mustern aus dem Multimediabereich kann man hier die Gauss'sche Normalverteilung wählen. Dann entsprechen die Parameter θ_i dem Mittelwertvektor und der Kovarianzmatrix der Merkmale eines Datums.

Die Berechnung der Kernfunktion, die das Distanzmaß liefert, wird dann nicht im ursprünglichen Merkmalsraum berechnet, sondern im Raum der PDF.

$$K(X_i, X_j) \implies K(p(x|\theta_i), p(x|\theta_j)) \quad (20)$$

Zur Berechnung der Parameter θ_i für ein X_i wird ein *Maximum Likelihood* Ansatz verwendet. Dazu müssen in jedem der X_i ausreichend viele Informationen vorhanden sein, um die θ_i der PDF korrekt zu schätzen.

Als Kerndistanz in diesem probabilistischen Merkmalsraum wird ein Distanzmaß verwendet, das PDFs vergleicht, in diesem Fall die symmetrische Kullback-Leibler Divergenz. Sie gibt ein Maß, für die Unterschiedlichkeit zweier Wahrscheinlichkeitsverteilungen.

$$D(p(x|\theta_i), p(x|\theta_j)) = \int_{-\infty}^{\infty} p(x|\theta_i) \log\left(\frac{p(x|\theta_i)}{p(x|\theta_j)}\right) dx + \int_{-\infty}^{\infty} p(x|\theta_j) \log\left(\frac{p(x|\theta_j)}{p(x|\theta_i)}\right) dx \quad (21)$$

Eine Matrix von Kerndistanzen, die direkt auf der symmetrischen Kullback-Leibler Divergenz beruht, würde die Mercer-Bedingungen nicht erfüllen, da sie z.B. nicht positiv definit wäre. Diese müssen notwendigerweise erfüllt werden, will man die Kernfunktion nutzen, um eine Matrix zu erstellen, die an eine Support Vektor Machine (Vgl. 7.3) übergeben werden kann. Um eine gültige Kernfunktion zu schaffen, wird der Exponent der skalierten und verschobenen symmetrischen Kullback-Leibler Divergenz gebildet.

$$K(X_i, X_j) \implies K(p(x|\theta_i), p(x|\theta_j)) \implies e^{-A D(p(x|\theta_i), p(x|\theta_j)) + B} \quad (22)$$

Alternativ kann die symmetrische Kullback-Leibler Divergenz auch als Distanzmaß in einem Gausskern verwendet werden.

$$K(X_i, X_j) = e^{-D(p(x|\theta_i), p(x|\theta_j))^2 / (2\sigma^2)} \quad (23)$$

Eine analytische Lösung der Kullback-Leibler Divergenz ist:

$$D(p(x|\theta_i), p(x|\theta_j)) = \text{tr}(\sum_i \sum_j^{-1}) + \text{tr}(\sum_j \sum_i^{-1}) - 2S + \text{tr}((\sum_i^{-1} + \sum_j^{-1})(\mu_i - \mu_j)(\mu_i - \mu_j)^T), \quad (24)$$

wobei S der Dimensionalität des ursprünglichen Signals x entspricht. Damit bietet die symmetrische Kullback-Leibler Divergenz ein Distanzmaß für mehrdimensionale Merkmale.

7.2 Der k -Nearest Neighbor Klassifikator

Bei dem k -Nearest Neighbor Algorithmus handelt es sich um ein einfaches Verfahren zur Klassifikation von Merkmalen. Diese werden als Punkte im Raum dargestellt, für die über eine Metrik ein Distanzmaß vorhanden ist. Für Punkte $x, y \in \mathbb{R}^n$ kann hier zum Beispiel die euklid'sche Distanz $d(x, y) = \sqrt{(x_1 - y_1)^2 + \dots + (x_n - y_n)^2}$ gewählt werden. Für PDFs ist die symmetrische Kullback-Leibler Divergenz ein mögliches Distanzmaß.

Mithilfe einer gelabelten Trainingsmenge kann dann ein Punkt x klassifiziert werden. Es werden die k Trainingspunkte bestimmt, die den geringsten Abstand zu ihm haben. Dann wird x der Klasse zugeordnet, der die Mehrheit seiner k nächsten Nachbarn angehört.

7.3 Support Vector Machine

Die *Support Vector Machine* ist ein optimaler Klassifikator, der auf einem statistischen, lerntheoretischen Ansatz beruht. Aus einer gelabelten Lernstichprobe von l Beobachtungen soll ein Modell erstellt werden, welches die Wahrscheinlichkeit der Fehlklassifikation von Testdaten minimiert. Zur Lösung dieser *Structural Risk Minimization* (SRM) liegt ein Ansatz von Vapnik und Chervonenkis vor, der in [1] genauer erläutert wird. Ein wichtiger Terminus ist hier die VC-Dimension einer Menge von Funktionen. Sie bezeichnet "die maximale Anzahl von Trainingspunkten, welche durch diese Klasse von Funktionen in allen möglichen Zuordnungen separiert werden können." [1] Zur SRM sollte die VC-Dimension der Funktion, die die Klassen trennt, minimiert werden.

Für ein Zweiklassenproblem wird in einem höherdimensionalen Merkmalsraum eine trennende Hyperebene $\langle w, x \rangle + b = 0$ gefunden, so dass die Klassifikation eines Datums x über

$$f(x) = \text{sgn}(\langle w, x \rangle + b) \quad (25)$$

vorgenommen werden kann.

Je breiter der Rand der trennenden Hyperebene ist, desto geringer ist ihre VC-Dimension. Dies ist insofern klar, als die Klassifikationssicherheit bei konstanter Intraklassenstreuung größer ist, je größer der Abstand zwischen den Klassen ist. Die Hyperebene ist durch dieje-

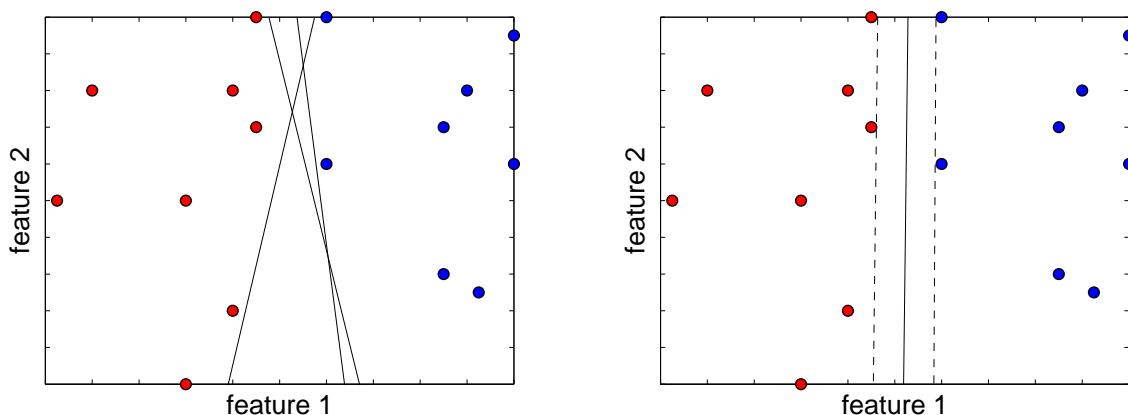


Abbildung 18: Zweidimensionales Zweiklassenproblem (links) mit möglichen trennenden Hyperebenen. (rechts) mit maximal breiter, trennender Hyperebene.

nigen Trainingsdaten genau bestimmt, welche auf ihrem Rand liegen. Sie werden als *Support*

Vektoren (SV) bezeichnet. Das Maximieren der Distanz der Hyperebene w zu den Supportvektoren x_i entspricht der Minimierung von $\|w\|^2$.

$$\begin{aligned} \langle w, w \rangle &\implies \min && \text{mit} \\ \forall i & y_i(\langle w, x_i \rangle + b) &\geq 1 \end{aligned} \quad (26)$$

Probleme, die in ihren ursprünglichen Merkmalsraum F nicht linear trennbar sind, werden mittels einer nichtlinearen Abbildung ψ in einen höherdimensionalen Merkmalsraum F' transformiert, in dem sie linear trennbar sind. Die Klassifikation geschieht dann über

$$f(x) = \text{sgn}(\langle w, \psi(x) \rangle + b). \quad (27)$$

Man nehme zum Beispiel an, ein Problem sei im ursprünglichen Merkmalsraum mit einem Polynom vom Grad p separierbar. Ein p -dimensionales Polynom im ursprünglichen, n -dimensionalen Merkmalsraum F führt dann zu einem $O(n^p)$ -dimensionalen Merkmalsraum F' . Da dort die explizite Berechnung des Skalarprodukts sehr rechenaufwändig ist, weicht man auf die implizite Berechnung mittels Kernfunktionen K aus.

$$K(x_i, x_j) = \langle \psi(x_i), \psi(x_j) \rangle \quad (28)$$

Das Optimierungsproblem aus (26) kann so ersetzt werden durch das Wolf-duale Optimierungsproblem der Maximierung von

$$L'(\alpha) = \sum_{i=1}^l \alpha_i - \sum_{i=1}^l \sum_{j=1}^l y_i y_j \alpha_i \alpha_j \langle x_i, x_j \rangle \quad (29)$$

$$\text{mit } \sum_{i=1}^l y_i \alpha_i = 0 \quad \text{und} \quad 0 \leq \alpha_i \leq c.$$

Dabei stellt c den Wert dar, mit dem eine Fehlklassifikation in der Trainingsmenge bestraft wird. Wird c zu groß gewählt, kann dies zu *overfitting* führen. Die Berechnung der Entscheidungsfunktion erfolgt im ursprünglichen Merkmalsraum F mittels

$$f(x) = \text{sgn}\left(\sum_{x_i \in SV} \alpha_i y_i \langle x_i, x \rangle + b\right). \quad (30)$$

Wichtige Kernfunktionen sind:

$$\begin{aligned} \text{Polynom - Kerne} & \quad K(x_i, x_j) = (\langle x_i, x_j \rangle + 1)^d \\ \text{Gauss - Kerne} & \quad K(x_i, x_j) = \exp(-\|x_i - x_j\|^2 / (2\sigma^2)) \\ \text{Sigmoid - Kerne} & \quad K(x_i, x_j) = \tanh(\kappa \langle x_i, x_j \rangle + \theta) \end{aligned} \quad (31)$$

Es gibt verschiedene Ansätze, die Two-Class-SVM auf Probleme mit mehr als zwei Klassen zu übertragen, nämlich, den *one versus one* und den *one versus rest* Ansatz. Bei *one versus one* wird für jeweils zwei Klassen eine trennende Hyperebene gefunden, bei *one versus rest* wird auf jede Klasse eine SVM angewendet, um sie von den restlichen Klassen zu separieren. Rechenaufwand und Speicherbedarf steigen offensichtlich mit der Anzahl der Klassen.

8 Implementierung

8.1 Interest Point Detektoren

Für beide *Interest Point* Detektoren lagen schon Implementierungen in C++ vor, die verwendet werden konnten. Die Implementierung des Wavelet basierten *Salient Point* Detektors wurde am Lehrstuhl durchgeführt. Für jede Aufnahme einer Münze gibt die vorhandene Funktion eine Karte zurück, die den *Saliency*-Wert jedes Bildpunkts enthält. Für die Auswertung der Kernfunktion wurde bei allen Münzen eine feste Anzahl von *Salient Points* ausgewählt, nämlich die 120 Punkte, mit dem höchsten *Saliency*-Wert. Die Implementierung des *Harris Corner Detector* stammt aus [10]. Diese Funktion gibt keine Karte zurück, sondern bestimmt die *Interest Points* abhängig von einem Schwellwert. Für einen festen Schwellwert werden so in jedem Bild unterschiedlich viele Punkte gefunden. Für die Auswertung mittels der symmetrischen Kullback-Leibler Divergenz stellt dies kein Problem dar, sofern in jedem Bild genügend Punkte gefunden werden, um die PDF zu schätzen. Dies ist mit den Aufnahmen, bei denen eine *Histogram Equalization* (Vgl. Kap. 3) durchgeführt wurde, gewährleistet.

8.2 Relationale Kernfunktion

Die relationale Kernfunktion wurde so implementiert, wie in [8] vorgeschlagen. Sie stellt also eine Rampenfunktion zwischen Null und Eins dar. Die ϵ -Werte sind dabei für je zwei Kreisparameter-Paare individuell zu wählen. Durch Anpassen der ϵ -Werte können später die Klassifikationsergebnisse verbessert werden, in jedem Fall aber sind sie so zu wählen, dass bei der Auswertung keine singulären Matrizen entstehen.

Die Kreise zur Auswertung der Funktion wurden in drei Versuchen bis zu einem Radius von sechs Pixeln gewählt. Unter anderem wurden auch verschiedene Phasen zu den gleichen Radien gewählt. Die Grauwerte, die auf demselben Radius liegen, wurden aber nie miteinander verglichen. Für je zwei Radien wurden ausserdem nur ein Grauwertepaar mit gleicher Phasendifferenz betrachtet, da alle weiteren keine neuen Informationen beitragen. Verschiebt sich in zwei Kreisen die Phase um den selben Wert, so führt dies lediglich zu einer zyklischen Verschiebung im Vektor der Differenzen der Grauwerte auf diesen Kreisen. Als *sampling rate* wurde 32 gewählt. Tabelle 1 zeigt die gewählten Kreisparameter in den ersten drei Versuchen.

Versuch 1	Versuch 2	Versuch 3
$p_1 = \{0, 0\}, p_2 = \{1, 0\},$ $p_3 = \{2, 0\}, p_4 = \{3, 0\},$ $p_5 = \{4, 0\}.$	$p_1 = \{0, 0\}, p_2 = \{1, 0\},$ $p_3 = \{2, 0\}, p_4 = \{3, 0\},$ $p_5 = \{4, 0\}, p_8 = \{2, \pi/2\},$ $p_9 = \{2, \pi\}, p_{10} = \{2, 3\pi/3\},$ $p_{11} = \{3, \pi/2\}, p_{12} = \{3, \pi\},$ $p_{13} = \{3, 3\pi/2\}, p_{14} = \{4, \pi/2\},$ $p_{15} = \{4, \pi\}, p_{16} = \{4, 3\pi/2\}.$	$p_1 = \{0, 0\}, p_2 = \{1, 0\},$ $p_3 = \{2, 0\}, p_4 = \{3, 0\},$ $p_5 = \{4, 0\}, p_6 = \{5, 0\},$ $p_7 = \{6, 0\}, p_8 = \{2, \pi/2\},$ $p_9 = \{2, \pi\}, p_{10} = \{2, 3\pi/3\},$ $p_{11} = \{3, \pi/2\}, p_{12} = \{3, \pi\},$ $p_{13} = \{3, 3\pi/2\}, p_{14} = \{4, \pi/2\},$ $p_{15} = \{4, \pi\}, p_{16} = \{4, 3\pi/2\}.$

Tabelle 1: Gewählte Kreisparameter.

Dazu kam ein vierter Versuch, in dem nur ein Kreisparameterpaar betrachtet wurde, nämlich $p_1 = \{0, 0\}$ und $p_2 = \{3, 0\}$. Hier wurde nicht über die Werte der Relationalen Kernfunktion gemittelt. Jeder der entstandenen Werte wurde als Merkmal genommen. Bei einer *sampling*

rate von 18 entstanden somit auch 18 Merkmale. Diese sind im Gegensatz zu den Merkmalen aus Versuch 1-3 nicht rotationsinvariant.

8.3 Hausdorff Distanz

Prinzipiell wurde die Hausdorff Distanz so implementiert, wie in [6] vorgeschlagen. Von den dort angegebenen Verbesserungen zur Beschleunigung der Berechnung wurde allerdings nur die erste umgesetzt, die schon in Kapitel 6 dargestellt wurde. Da für die Klassifikation von Münzen, dargestellt in Polarkoordinaten, nicht Rotations- und Translationsinvarianz in mehrere Richtungen gewährleistet werden muss, sondern lediglich Translationsinvarianz in eine Richtung, ist dies die vielversprechendste. Um weitere Rechenzeit zu sparen wurden die Testbilder nicht mit allen Trainingsbildern sondern nur mit dem Durchschnittsbild über die Trainingsbilder einer Klasse verglichen. Neben der Hausdorff Distanz wurde für jede Translation auch die modifizierte Hausdorff Distanz bestimmt. Die Translation, bei der diese minimal war, wurde genutzt, um das beste *Matching* zu bestimmen.

Für das beste *Matching* wurde dann die Anzahl der übereinstimmenden Kantenpunkte in beiden Bildern bestimmt, sowie die Anzahl der Punkte im Testbild, für die es keine Übereinstimmung im Trainingsbild gibt. Diese vier Ähnlichkeitsmaße und Kombinationen aus ihnen wurden dann auf ihre Qualität bei der Klassifikation der Münzen getestet.

8.4 Support Vector Machine

Es wurde die LIBSVMTL verwendet, eine objektorientierte C++ Support Vector Machine Library, die auf der LIBSVM von Chih-Jen Lin aufbaut [12]. Hier ist die Möglichkeit gegeben, der SVM vorberechnete Distanzmatrizen oder sogar Kernmatrizen als Eingabe zu übergeben. Die LIBSVM verfügt nicht über diese Möglichkeit. Sie nimmt als Eingabe lediglich die Merkmalsvektoren und berechnet darauf selbst eine der vorgesehenen Kernfunktionen. Da die mit der relationalen Kernfunktion auf den *Salient Points* berechneten Merkmale als Matrix vorliegen, bietet es sich an, eine Distanzmatrix mittels der symmetrischen Kullback-Leibler Divergenz zu erstellen. Diese kann dann zu einer Klassifikation durch den *Nearest Neighbor* Klassifikator genutzt oder an die LIBSVMTL übergeben werden.

Eine Frage, die oft Schwierigkeiten bereitet, ist, wie man zur Klassifikation vorhandenes Vorwissen bei Anwendung der SVM nutzen kann. Dies ist hier relevant, da zum Beispiel Informationen über die Radien der Münzen vorliegen, die in die Merkmalsberechnung bislang nicht eingegangen sind. Hier wurde dieses Problem gelöst, indem für zwei Münzen, deren Radius sich um mehr als ein Pixel unterscheidet, der Distanzwert additiv erhöht wurde. Damit werden unterschiedliche Radien direkt in der Distanzmatrix bestraft, bevor diese an die SVM übergeben wird.

9 Ergebnisse

Während die Segmentierung der Trainingsbilder relativ zuverlässig verlief und der ermittelte Radius von Münzen der gleichen Klasse nur um etwa ein bis zwei Pixel variierte, haben sich schon bei der Segmentierung der Testbilder erste Probleme gezeigt. Sie wurden wesentlich unzuverlässiger richtig ausgeschnitten, oft fehlte der Rand oder auf einer der Seiten war noch im ausgeschnittenen Bild die Hintergrundtextur zu sehen. Zur Klassifikation wurde daher der Testdatensatz von 1100 Bildern auf die 834 Aufnahmen reduziert, bei denen die Segmentierung korrekt funktioniert hatte. Dennoch erschien der Testdatensatz auch bei manueller Auswertung wesentlich schwieriger als der Trainingsdatensatz.

9.1 Relationale Kernfunktionen

Die Klassifikation der Testbilder mit der relationalen Kernfunktion fiel sehr schlecht aus. Berücksichtigte man den Radius der Münzen nicht, so lagen die Klassifikationsraten auf den *Salient Points* zwischen 2 Prozent (*1-Nearest Neighbor* bei Versuch 1) und 26.5 Prozent (SVM bei Versuch 4). Eine 10-fold Cross Validation auf den Trainingsdaten hingegen versprach Ergebnisse von um die 45 %.

Der Unterschied liegt wahrscheinlich darin begründet, dass für die Aufnahmen der Testbilder andere Kamerasysteme verwendet wurden, als für die Aufnahme der Trainingsbilder, die alle von einem Kamerasystem erfasst wurden. Die Beleuchtungsunterschiede in den Bildern beider Mengen gehen über additive Verschiebungen hinaus, und können daher nicht durch die relationale Kernfunktion ausgeglichen werden. Meist sind die Testbilder wesentlich kontrastärmer als die Trainingsbilder. Glanzlichter an den Konturen der Trainingsbilder finden sich in den Testbildern meist nicht wieder. Abbildung 19 zeigt zwei Beispiele für die Unterschiede zwischen Test- und Trainingsbildern.

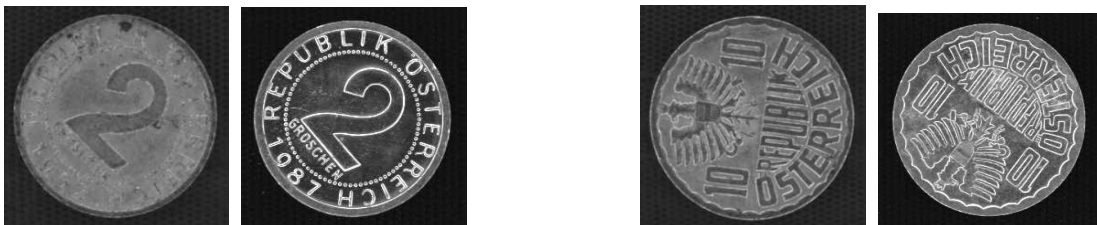


Abbildung 19: Jeweils zwei Bilder aus der selben Klasse, links je ein Testbild, rechts ein Trainingsbild.

Um die relationale Kernfunktion auf den *Interest Points* dennoch zu testen, wurde die Trainingsmenge in neue Test- und Trainingsmengen unterteilt. Jedes zehnte Bild aus der Trainingsmenge wurde der neuen Testmenge zugeordnet. Die Klassifikationsergebnisse fielen auf diesen Mengen wesentlich besser aus.

9.1.1 Auswertung auf *Interest Points*

In Tabelle 2 sind die Ergebnisse aufgelistet, die erzielt werden konnten, wenn man die relationale Kernfunktion in der Umgebung der Punkte auswertete, die mit dem *Harris Corner Detector* bestimmt wurden. Die Klassifikation wurde hier lediglich mit dem *Nearest Neighbor* Klassifikator durchgeführt. Nur auf die Merkmale, die bei der Klassifikation mit dem *Nearest Neighbor* Klassifikator insgesamt am vielversprechendsten erschienen, wurde zum Vergleich danach die SVM angewendet. Zunächst ist augenscheinlich, dass die Ergebnisse, berücksichtigt man die Radien der Münzen zur Klassifikation, wesentlich besser sind als im anderen

		1NN	3NN
Versuch 1	ohne Radius	184/876 (21%)	177/876 (20%)
	mit Radius	428/876 (49%)	431/876 (49%)
Versuch 2	ohne Radius	250/876 (29%)	258/876 (29%)
	mit Radius	450/876 (51%)	444/876 (51%)
Versuch 3	ohne Radius	233/876 (27%)	233/876 (27%)
	mit Radius	414/876 (47%)	411/876 (47%)

Tabelle 2: Ergebnisse der Klassifikation mit *Nearest Neighbor* Klassifikator.

Fall. Dies liegt vermutlich daran, dass sich die Motive auf vielen Münzen sehr ähneln. Auf den Vorderseiten der Münzen der meisten Währungen sind in der Mitte Zahlen abgebildet, die den Wert der Münze bestimmen. Da in den meisten Ländern im Dezimalsystem gerechnet wird, sind auch die Einheiten auf den Münzen entsprechend ähnlich. Es fällt auf, dass Münzenaufnahmen, die die gleiche Zahl zeigen, relativ häufig verwechselt werden, wenn man den Radius nicht berücksichtigt. Auf den Rückseiten haben in vielen Währungen eine ganze Serie von Münzen das gleiche Motiv. Beispiele dazu sind in Tabelle 6 abgebildet. Auch diese Münzen lassen sich vor allem anhand ihres Radius unterscheiden.

Das beste Ergebnis wurde hier in Versuch 2 erzielt, mit 28 Merkmalen also, die in einem Radius von bis zu 4 Pixeln um die *Interest Points* berechnet wurden. Der größere Radius in Versuch 3 verschlechtert das Ergebnis dann. Dies liegt vermutlich daran, dass sich die *Interest Points* meist in den Schriftzügen am Rand der Münzen sammeln. Hier sind die Strukturen relativ fein. Insgesamt erscheint dieses Ergebnis mit einer Klassifikationsrate von 51% allerdings nicht befriedigend.

Die Ergebnisse der Klassifikation mit *1-Nearest Neighbor* und *3-Nearest Neighbor* sind in etwa gleich, in manchen Fällen unterliegt sogar der *3-Nearest Neighbor* Klassifikator.

9.1.2 Auswertung auf *Salient Points*

Für die relationalen Merkmale, berechnet auf den Punkten, die mit dem Wavelet basierten *Salient Point Detector* bestimmt wurden, fielen die Klassifikationsergebnisse aus wie in Tabelle 3 aufgelistet.

		1NN	3NN	SVM
Versuch 1	ohne Radius	130/876 (15%)	136/876 (16%)	
	mit Radius	432/876 (49%)	467/876 (53%)	
Versuch 2	ohne Radius	175/876 (20%)	183/876 (21%)	
	mit Radius	494/876 (57%)	514/876 (59%)	
Versuch 3	ohne Radius	268/876 (31%)	275/876 (32%)	466/876(53%)
	mit Radius	605/876 (69%)	608/876 (69%)	675/876(77%)
Versuch 4	ohne Radius	279/876 (32%)	294/876 (34%)	483/876(55%)
	mit Radius	595/876 (68%)	619/876 (71%)	696/876(79%)

Tabelle 3: Ergebnisse der Klassifikation mit *Nearest Neighbor* Klassifikator und SVM.

Die Ergebnisse, die ohne Berücksichtigung der Radien erzielt wurden, sind etwas schlechter, als bei der Auswertung der RKF auf den *Harris Interest Points*. Unter Berücksichtigung der Radien aber fallen die Ergebnisse mit den Wavelet basierten *Salient Points* besser aus (Vgl. Abbildung 21). Dies hängt womöglich damit zusammen, dass die *Salient Points* sich auch an

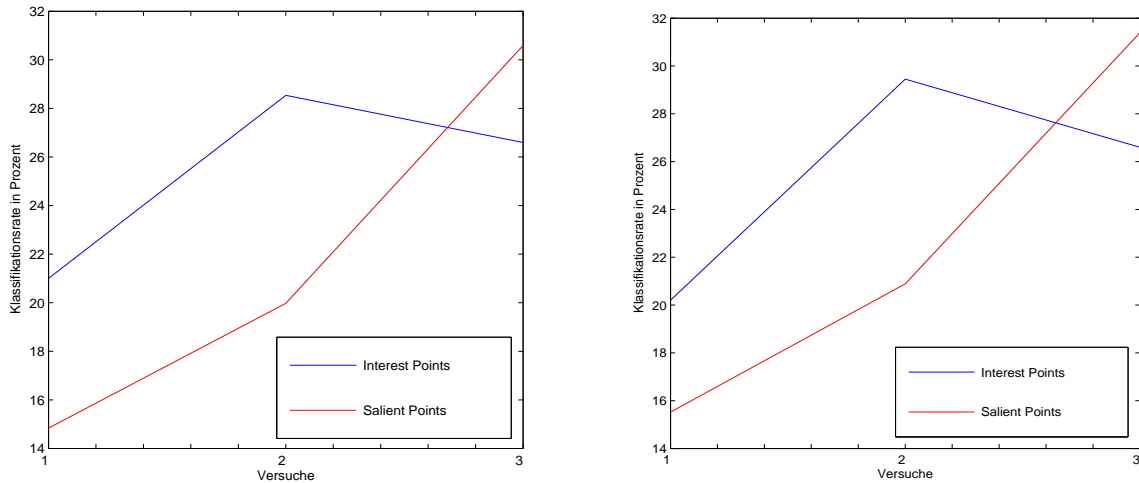


Abbildung 20: Die Ergebnisse der Klassifikation in Versuch 1, 2 und 3 bei unberücksichtigtem Radius. Gegenübergestellt werden die Ergebnisse der Auswertung auf den *Interest Points* bestimmt mit dem *Harris Corner Detector* (blau) und die Ergebnisse auf den Wavelet basierten *Salient Points* (rot). Die linke Graphik zeigt die Klassifikation mit *1-Nearest Neighbor* die rechte die Klassifikation mit *3-Nearest Neighbor* Klassifikator.

den Kanten der aufgedruckten Zahlen und Motive sammeln und so die Innenfläche der Münze besser repräsentieren als die *Interest Points*. Die Münzen werden mit den *Salient Points* vor allem nach ihrem Motiv erkannt. Die *Interest Points* finden sich nur an den wenigen Ecken der Motive im Zentrum und sammeln sich an den Schriftzügen, die oft am Rand der Münzen liegen. Damit können Münzen unterschiedlicher Währungen leichter klassifiziert werden.

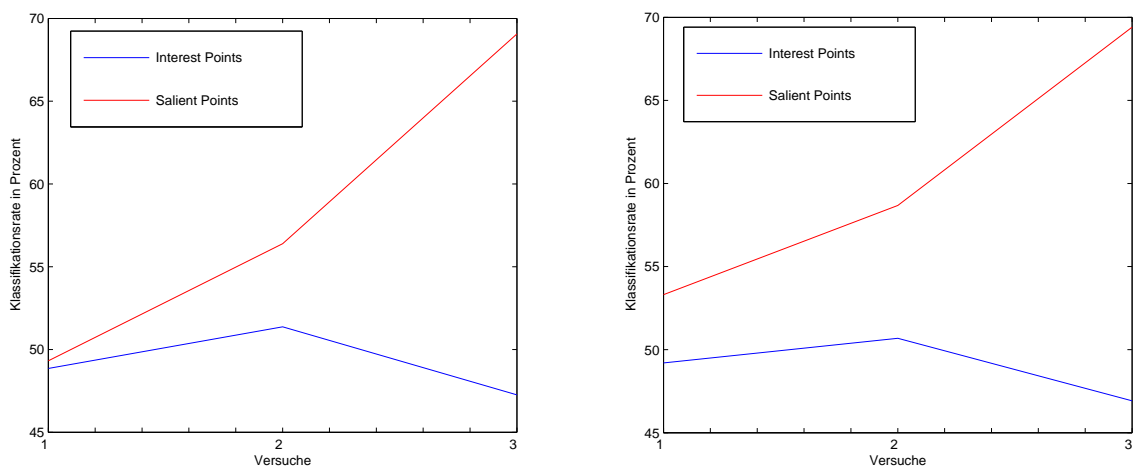


Abbildung 21: Die Ergebnisse der Klassifikation in Versuch 1, 2 und 3 unter Berücksichtigung des Radius auf *Interest Points* (blau) und auf *Salient Points* (rot). (links) Klassifikation mit *1-Nearest Neighbor*. (rechts) Klassifikation mit *3-Nearest Neighbor*.

In jedem Fall zeigt sich, dass der Radius für die relationalen Merkmale auf den *Salient Points* mehr zusätzliche Informationen liefert, als für die Merkmale, die auf den *Interest Points* berechnet wurden.

Beim Betrachten der Ergebnisse, die mit den rotationsinvarianten Merkmalen erzielt wurden, fällt auf, dass die Klassifikation besser wurde, je mehr Merkmale man betrachtete. Dabei ist der Sprung zwischen den Ergebnissen von Versuch 1 zu 2 wesentlich geringer, als der von Versuch 2 zu Versuch 3. Das Betrachten der Grauwertdifferenzen bei unterschiedlichen Phasen bringt also nicht einen so großen Informationsgewinn wie die Hinzunahme größerer Radien. Die besten Ergebnisse erhielt man in Versuch 4, obwohl hier pro *Salient Point* nur 18 Merkmale bestimmt wurden. Dies ist einleuchtend, da in allen anderen Merkmalen zusätzliche Rotationsinvarianz vorliegt, die für die Münzaufnahmen in Polarkoordinaten nicht notwendig ist. Die Ergebnisse in Versuch 3 waren allerdings nicht wesentlich schlechter, sondern bei der Klassifikation mit dem *1-Nearest Neighbor* Klassifikator sogar geringfügig besser. Von Münzen, die in Versuch 3 unter Berücksichtigung der Radien falsch klassifiziert wurden, sind im Anhang einige Beispiele abgebildet, zusammen mit der Aufnahme, die der *1-Nearest Neighbor* Klassifikator als Ähnlichste bestimmt hat. Es fällt auf, dass die Grauwerte der als ähnlich klassifizierten Münzen trotz der relationalen Kernfunktion meist im selben Bereich liegen.

Als bester Klassifikator hat sich wie erwartet die SVM herausgestellt. Bei bekanntem Radius verbessert die SVM das Klassifikationsergebnis um etwa 8 Prozent.

Abbildung 22 zeigt die Confusion Matrix, die sich bei Versuch 3 unter Berücksichtigung des Radius ergibt, wenn zur Klassifikation die SVM verwendet wird. Zur besseren Lesbarkeit sind die Werte logarithmisch skaliert eingetragen.

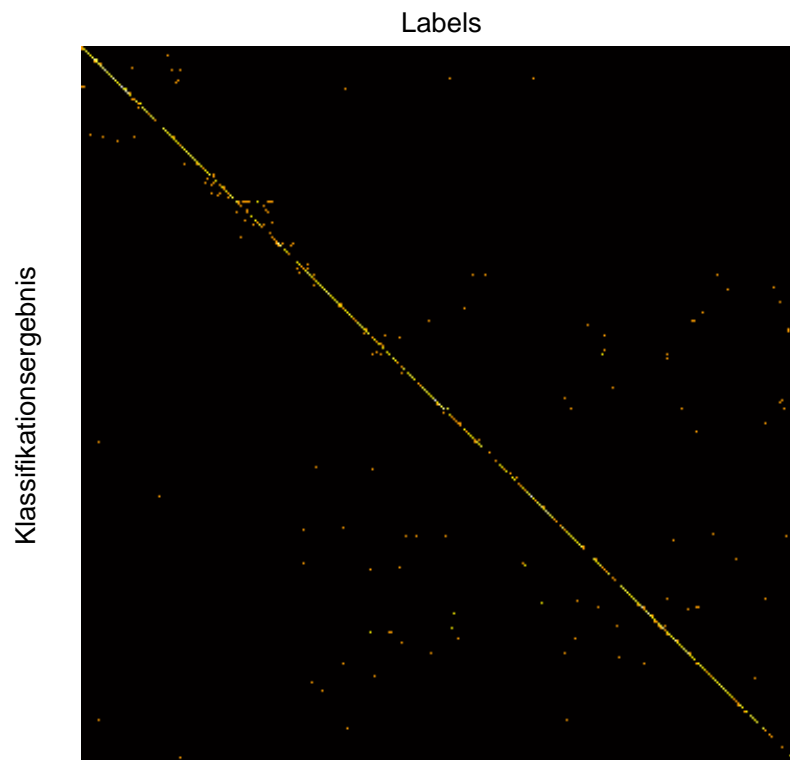


Abbildung 22: Confusion Matrix.

Die Confusion Matrix für Versuch 4 unter Berücksichtigung des Radius und bei Anwendung der SVM findet sich im Anhang (Abbildung 24).

9.2 Hausdorff Distanz

Zunächst stellte sich die Frage, welches Distanzmaß auf Basis der Hausdorff Distanz das beste zur Klassifikation von Münzen ist. Auf einer kleineren Menge von 1000 Bildern aus 35 Klassen aus der Menge der Trainingsbilder wurden die verschiedenen Distanzmaße getestet. Jedes Bild wurde mit allen anderen verglichen und mit dem *1-Nearest Neighbor* Klassifikator klassifiziert.

Als Distanzmaße wurden die Hausdorff Distanz (hausdorff), die modifizierte Hausdorff Distanz (mod), das Verhältnis der Anzahl der Kantenpixel, die in den Bildern übereinstimmen (u) und der Anzahl der Pixel, für die es keine Übereinstimmung gibt (s) und Kombinationen aus diesen Distanzmaßen verwendet. Die Ergebnisse sind in Tabelle 9.2 angegeben. Die bes-

τ	s/u	mod	mod· s/u	hausdorff	hausdorff· s/u	hausdorff· s^2/u
10	87.7%	65.7%	75.3%	71.3%	88.2%	88.5%
30	89.4%	67.2%	76.8%	73.8%	89.8%	89.9%
50	90.0%	67.2%	77.8%	74.7%	90.6%	90.7%

Tabelle 4: Klassifikation mittels verschiedener Distanzmaße auf der Basis der Hausdorff Distanz.

ten Ergebnisse konnten mit $\text{hausdorff} \cdot s^2/u$ erzielt werden. Für alle Distanzmaße zeigt sich, dass das Klassifikationsergebnis mit wachsendem Schwellwert τ besser wird. Je kleiner das τ ist, desto ungenauer werden die Bilder aufeinander *gematcht*.

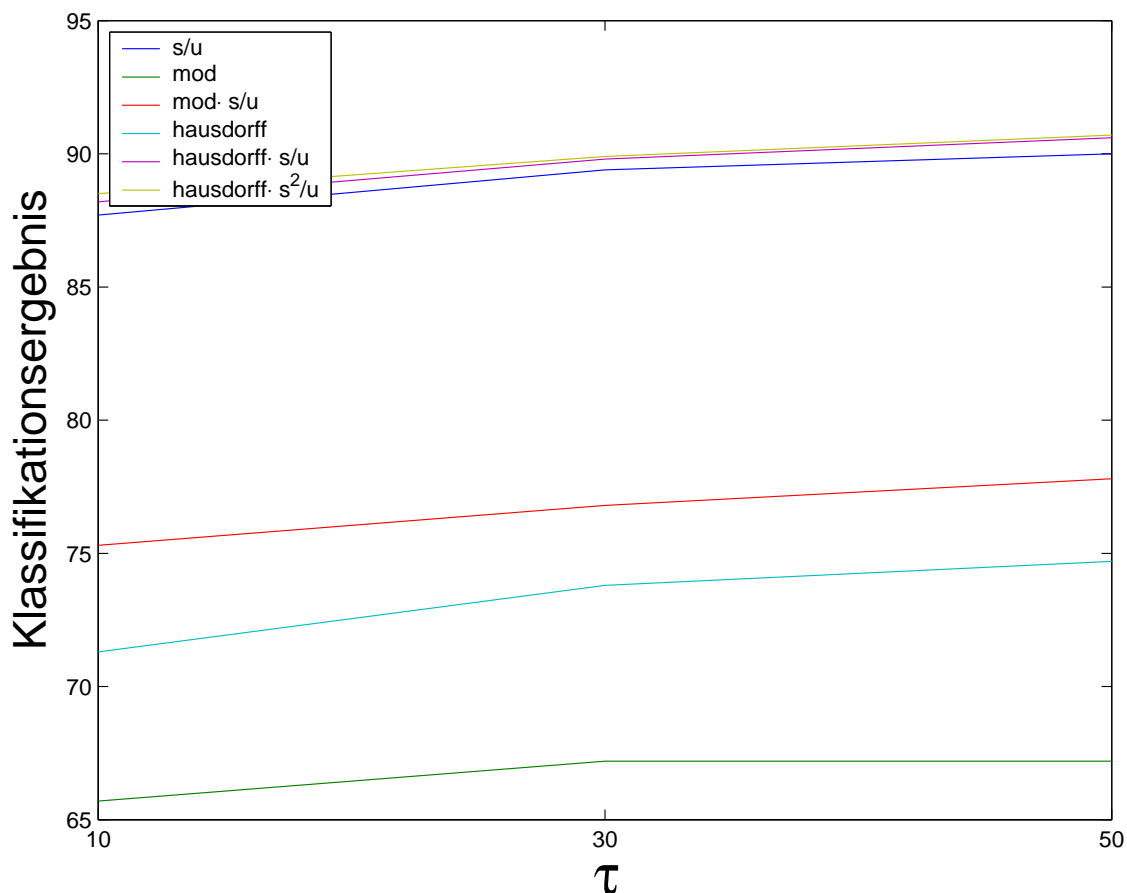


Abbildung 23: Klassifikation mit verschiedenen Distanzmaßen in Abhängigkeit von τ .

Mit dem Distanzmaß $\text{hausdorff}\cdot s^2/u$ wurde dann die ursprüngliche Testmenge klassifiziert. Auch hier zeigt sich, dass die Ergebnisse besser sind, je größer man den Parameter τ wählt.

$\tau=15$	$\tau=25$	$\tau=35$
66.19%	70.5%	72.66%

Tabelle 5: Klassifikation der Testmenge.

Dass die Ergebnisse bei den Testbildern schlechter ausfallen, als auf der Menge der 1000 Trainingsbilder, liegt wahrscheinlich vor allem an der reduzierten Anzahl von Klassen in dieser Menge. Einen Einbruch der Klassifikationsraten zwischen Test- und Trainingsmenge gibt es bei der Klassifikation mit der Hausdorff Distanz also nicht. Im Gegenteil: die neue Testmenge, die für die Auswertung der relationalen Merkmale aus einem Zehntel der ursprünglichen Trainingsmenge gebildet worden war, wurde mit der Hausdorff Distanz schlechter klassifiziert, als die ursprüngliche Testmenge. Für $\tau = 15$ wurden von der neuen Testmenge 53.31% der Bilder richtig klassifiziert. Die schlechteren Ergebnisse können allerdings damit zusammenhängen, dass aus der ursprünglichen Testmenge nur diejenigen Bilder getestet wurden, die korrekt segmentiert worden waren. In der Trainingsmenge lagen auch Bilder vor, bei denen die segmentierten Münzen im Radius um bis zu zwei Pixel vom tatsächlich korrekten Radius abwichen. Der Rand fehlte also teilweise.

10 Schluss

Es hat sich herausgestellt, dass die relationalen Kernfunktionen durchaus Merkmale liefern, die auch auf der Münzdatenbank leistungsfähig sind. Allerdings wird der Ausgleich der Beleuchtungsunterschiede doch nicht in dem Maße erreicht, wie es wünschenswert wäre. Die ursprünglichen Beleuchtungsunterschiede bleiben auch nach der Anwendung der relationalen Kernfunktion ein wichtiges Klassifikationskriterium. Bei zu starken Beleuchtungsunterschieden, wie sie zum Beispiel auch zwischen Test- und Trainingsmenge vorlagen, erleiden die Klassifikationsraten daher auch einen starken Einbruch.

Die Wavelet basierten *Salient Points* scheinen den auf den Ecken befindlichen *Interest Points* in der Repräsentation der Münzen überlegen zu sein. Sie repräsentieren besser die Motive in der Mitte der Fläche, unabhängig davon, wie eckig das jeweilige Motiv ist. Auch erscheinen die Klassifikationsraten bei Verwendung der *Salient Points* noch steigerungsfähig. Die Ergebnisse sind mit wachsendem Radius besser geworden; die Vermutung liegt also nahe, dass man in einem noch größeren Radius um die Punkte herum weitere diskriminative Merkmale zur Klassifikation finden kann. Dies ist in der Umgebung der *Interest Points* nicht der Fall. Hier wurden die Ergebnisse mit größerem Radius ja schlechter.

Bessere Ergebnisse würden sich wohl auch mit einer Anpassung der Parameter, vor allem des ϵ in der relationalen Kernfunktion, ergeben. Dies ist zeitaufwendig, da man für jedes Kreisparameterpaar einen optimalen Wert finden müsste. Insgesamt ist jedoch wichtig zu bemerken, dass die besten Ergebnisse erzielt wurden, indem man auf die überflüssige Rotationsinvarianz verzichtete.

Die kantenbasierten Verfahren haben es als einzige geleistet, auch auf der schwereren und anders beleuchteten Testmenge gute Ergebnisse zu erzielen. Die modifizierte Hausdorff Distanz ist ein gutes Maß, mithilfe dessen zwei Binärbilder zuverlässig übereinander gematcht werden können. Als Distanzmaß zur Bestimmung der Klassenzugehörigkeit sind dann andere Distanzmaße auf der Basis der Hausdorff Distanz geeigneter.

Literatur

- [1] Burkhardt, H.: Grundlagen der Bilderzeugung und Bildanalyse. Vorlesung am Institut für Informatik. Albert-Ludwigs-Universität Freiburg.
- [2] Burkhardt, H.: Algorithmen zur Digitalen Bildverarbeitung. Vorlesung am Institut für Informatik. Albert-Ludwigs-Universität Freiburg.
- [3] Burkhardt, H.: Bildverarbeitungspraktikum I. Veranstaltung am Institut für Informatik. Albert-Ludwigs-Universität Freiburg.
- [4] Harris, C. und Stephens, M.: A Combined Corner and Edge Detector, Proc. of 4th Alvey Vision Conference, S. 147-151, 1988.
- [5] Loupas, E. und Sebe, N.: Wavelet-based Salient Points for Image Retrieval, Research Report RR 99.11, RFV-INSA Lyon, 1999.
- [6] Huttenlocher, D.P., Klanderman, G.A., Rucklidge, W.J.: Comparing Images Using the Hausdorff Distance.
- [7] Moreno, P.J., Ho, P.P., Vasconcelos, N.: A Kullback-Leibler Divergence Based Kernel for SVM Classification in Multimedia Applications. HP Laboratories Cambridge, 2004.
- [8] Schael, M.: Invariant Texture Classification using Group Averaging with Relational Kernel Functions. In: Texture 2002 the 2nd international workshop on texture analysis and synthesis, S. 129-134, 2002.
- [9] Schulz-Mirbach, H.: Invariant features for gray scale images. In G. Sagerer, S. Posch, and F. Kummert, Hrsg., 17. DAGM - Symposium "Mustererkennung", S. 1-14, Bielefeld, 1995. Reihe Informatik aktuell, Springer.
- [10] <http://www.robots.ox.ac.uk/vgg/research/affine/index.html>
- [11] <http://muscle.prip.tuwien.ac.at/index.php>
- [12] <http://lmb.informatik.uni-freiburg.de/lmbsoft/libsvmtl/>

A Confusion Matrix

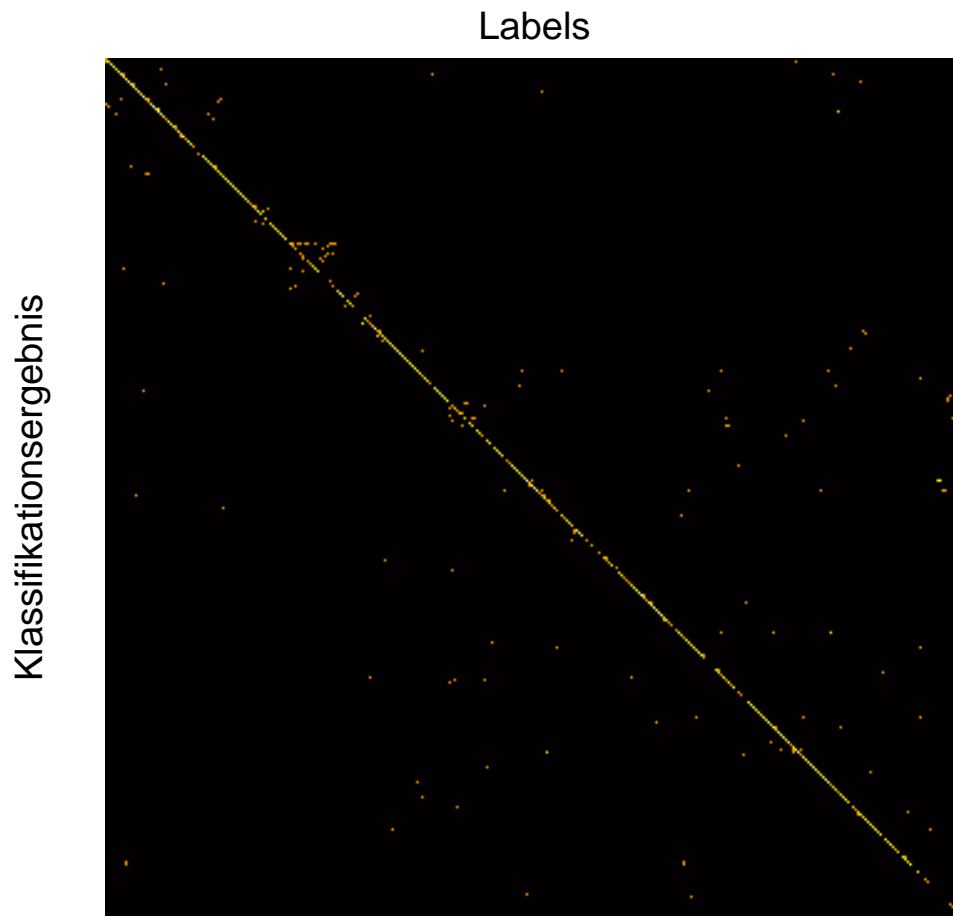


Abbildung 24: Confusion Matrix aus Versuch 4 unter Berücksichtigung des Radius bei Anwendung der SVM.

B Beispiele: Verwechselte Münzen

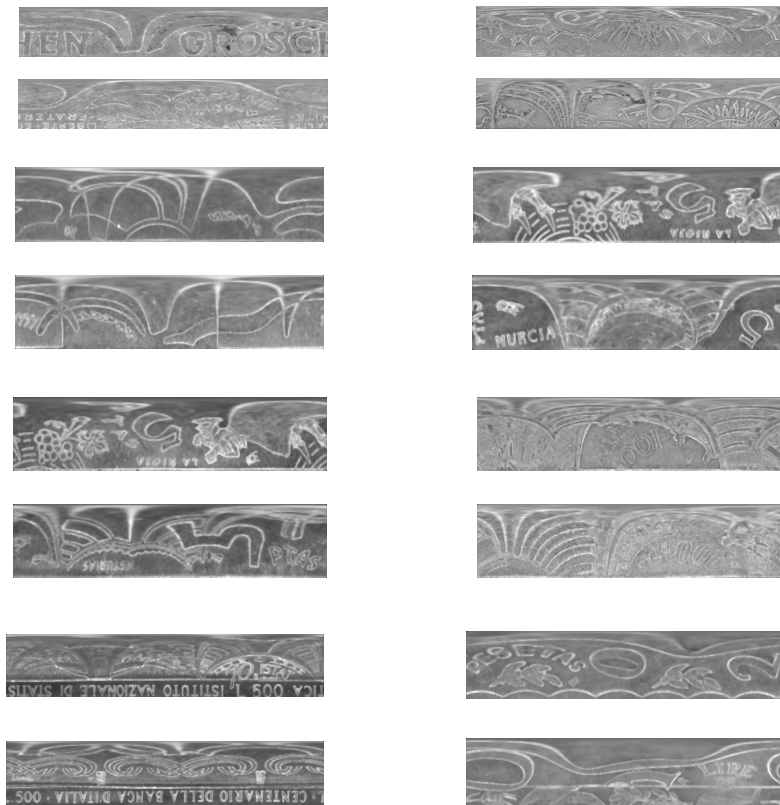


Abbildung 25: Acht Testbilder (jeweils oben) und ihr *Nearest Neighbor* aus der Menge der Trainingsbilder(jeweils darunter).

C Motivserien


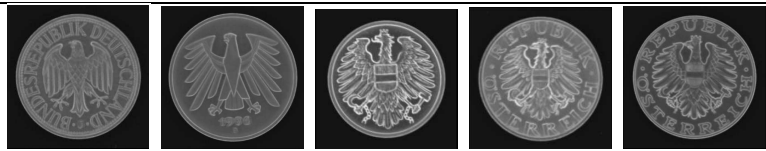


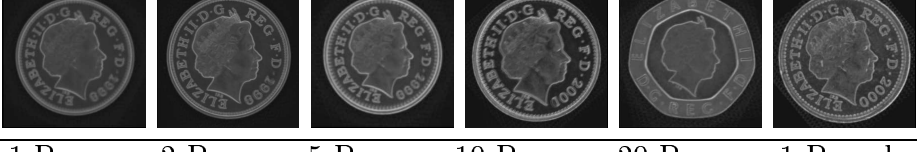



Münzen	Motiv
	Eichenlaub
1Pfennig 2Pfennig 5Pfennig 10Pfennig 10Pfennig	
	Adler
1DM 5DM 2Groschen 5Groschen 10Schilling	
	E II R
	E II R
	E II R
1 Shilling 1 Penny 2 Pence 5 Pence 10 Pence 20 Pence 1 Pound	
	Löwe
10Haleru 20h 50h 1CZK 2CZK 5CZK 10CZK	
	Königin Beatrix
	Königin Juliana
5ct 10ct 25ct 1Gulden 2G	

Tabelle 6: Beispielsatz von Durchschnittsmünzen verschiedener Klassen aus der Datenbank. Jeweils eine ganze Serie von Münzen trägt das gleiche Motiv.