

Equivariante Kerne in der Mustererkennung

Marco Reisert

Abteilung für Röntgendiagnostik
Medizinische Physik , Uniklinik Freiburg
marco.reisert@uniklinik-freiburg.de

Abstract: Einer der wichtigsten Einflussfaktoren im Entwurf eines Mustererkennungssystems stellt a-priori Wissen dar. A-priori Wissen liefert Information welche über die reinen Trainingsdaten hinausgeht. Eine wichtige Untergruppe ist das sogenannte Transformationswissen. Man weiss, welche Gruppe von Transformationen die Bedeutung der betrachteten Muster unverändert lässt. In der Literatur existieren viele ad-hoc Ansätze wie solches Wissen miteinbezogen werden kann. In dieser Dissertation wurde ein wohlfundiertes mathematisches Rahmenwerk ausgearbeitet, welches klärt, in welcher Weise sich solches Wissen optimal in ein System integrieren lässt. Aus praktischer Perspektive wird eine Methode vorgestellt, wie sich explizit Equivarianzbedingungen in kernbasierte und parametrische Modellbeschreibungen integrieren lassen. In diesem Artikel werden die Grundbegriffe und Hauptresultate präsentiert und ein Anwendungsbeispiel zur rotationsinvarianten Objektdetektion besprochen.

1 Einführung

Typischerweise kann der Mustererkennungsprozeß als eine Klassifikations- oder Regressionsfunktion ausgedrückt werden. Kernbasierte Modelle [SS02] dieser Funktionen kamen Mitte der 90er Jahren auf. Diese neue Klasse von Ansätzen ermöglichte es Forschern, nichtlineare Beziehungen mit einer Effektivität zu analysieren, die bis dahin lediglich linearen Algorithmen vorbehalten war. Sowohl vom Gesichtspunkt der Berechenbarkeit als auch von einem konzeptuell, mathematischen Punkt aus, sind kernbasierte Algorithmen sowohl effizient als auch wohlfundiert, wobei sie nicht unter den typischen Problemen nichtlinearer Algorithmen leiden, wie zum Beispiel das Steckenbleiben in lokalen Minimas oder dem sogenannten 'overfitting'. In dieser Dissertation [Rei08] wurde die konzeptuelle Einfachheit der kernbasierten Modelle genutzt, um ein theoretisch wohlfundiertes Rahmenwerk des Invarianzbegriffs in der Musteranalyse aufzustellen. Die lineare Natur hilft auf überraschend einfache Weise Optimalitätsaussagen zu treffen. Diese sind eng verwandt mit dem Prinzip der Gruppenintegration, welche seine Ursprünge in der klassischen Invariantentheorie hat.

Von einem abstrakten Standpunkt aus, kann der Mustererkennungsprozess als Regressions oder Klassifikationsfunktion $f : X \rightarrow Y$ formuliert werden. Hierbei ist X der Raum der Muster und Y der Ergebnisraum. Zum Beispiel könnte Y den Raum der möglichen Klassifikationsergebnisse repräsentieren. Der Begriff der Invarianz bedeutet nun, dass sich das

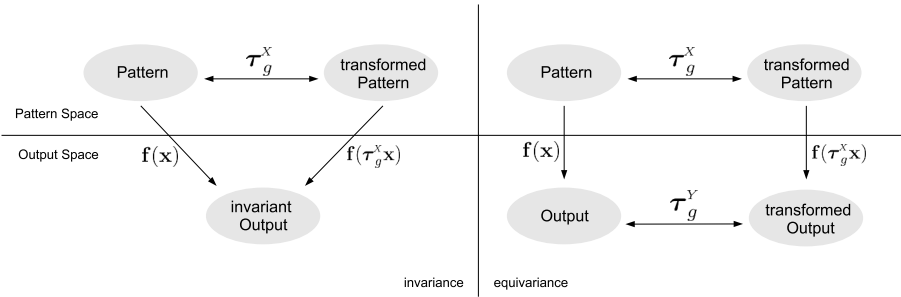


Abbildung 1: Invarianz und Equivarianz im Vergleich

Ergebnis des Erkennungsprozesses bei bestimmten Transformation des Eingangsmusters nicht ändert, formal

$$f(\tau_g \mathbf{x}) = f(\mathbf{x}) \quad \forall \mathbf{x} \in X, \forall g \in \mathcal{G}.$$

wobei τ_g die Wirkung der Transformation $g \in \mathcal{G}$ auf das Muster \mathbf{x} beschreibt und \mathcal{G} eine kompakte Gruppe ist. Man stelle sich folgendes Beispiel vor: ein Protein soll anhand seiner dreidimensionalen Struktur einer funktionalen Klasse zugeordnet werden. Natürlich ist die Klasse unabhängig von der Lage des Proteins im Raum, somit muss auch die Klassifikationsfunktion diese Anforderung erfüllen. In diesem Fall wäre die Invarianzgruppe \mathcal{G} die euklidische Bewegungsgruppe.

Auf der anderen Seite steht der Begriff der Equivarianz (Physiker sprechen auch oft von Kovarianz). Die Equivarianz steht mehr im Bezug zum Regressionsproblem. Man könnte sagen, dass das Ergebnis der Regressionsfunktion das Transformationsverhalten des Eingangsmuster 'erbt'. Formal bedeutet dies

$$f(\tau_g^X \mathbf{x}) = \tau_g^Y f(\mathbf{x}) \quad \forall \mathbf{x} \in X, \forall g \in \mathcal{G},$$

wobei τ_g^X die Gruppenwirkung auf dem Musterraum bezeichnet und τ_g^Y die Wirkung auf dem Ergebnisraum. Zur Anschauung stelle man sich folgendes Beispiel vor. Für eine gegebene Proteinstruktur möchte man eine Wahrscheinlichkeitskarte für die Position eines bestimmtes aktives Zentrum lernen. Natürlich hängt auch diese Karte von der tatsächlichen Lage des Proteins im Raum ab, das heisst falls das Protein seine Lage ändert muss auch dementsprechend die Lage der Wahrscheinlichkeitskarte sich ändern. Beide Begriffe, die der Invarianz und Equivarianz sind in Abbildung 1 veranschaulicht.

Diese Arbeit beschäftigt sich nun mit der präzisen Formulierung von Invarianz und Equivarianz in kernbasierten Räumen. In diesem Artikel werden auf der einen Seite die theoretischen Kerngedanken ausgearbeitet. Als ein Hauptresultat wird das 'Representer Theorem' formuliert, welches zeigt, dass die Lösung eines bestimmen invarianten Lernproblems sich mit dem Prinzip der Gruppenintegration formulieren lässt. Auf der anderen Seite wird als Anwendungsbeispiel ein Konzept zur Objektdetektion vorgestellt, welches auf der Idee der verallgemeinerten Hough Transformation basiert. Hierbei wird das Funktional, welches einen lokalen Bildausschnitt auf eine örtliche Wahrscheinlichkeitsverteilung für das Objektzentrum abbildet, mittels einer rotations-equivarianten Kernmaschine modelliert.

2 Equivarianz in kernbasierten Räumen

Im Bereich des maschinellen Lernens sind kernbasierte Algorithmen heutzutage nicht mehr wegzudenken. Der wohl bekannteste Vertreter ist wohl die sogenannte Support-Vektor Maschine. Sie hat durch ihre Einfachheit und Robustheit als allgemeines Regressions- und Klassifikations-Tool Einzug in viele Labors gehalten. Die Idee der kernbasierten Methoden ist recht einfach: Man nehme eine Methode, welche zur Grundlage einen linearen Vektorraum hat (als einfaches Beispiel betrachte man die 'lineare Regression'). Nun bringe man das Problem in die Form, dass es nur noch von Innenprodukten in dem zu Grunde liegenden Vektorraum abhängt. Statt nun das normale Innenprodukt auszuwerten ersetzt man dieses mit einer 'fast' beliebigen nichtlinearen Funktion, dem sogenannten Kern. Dieser Kern kann nun wiederum als inneres Produkt in einem verzerrten hochdimensionalen Raum aufgefasst werden. Die Hoffnung ist, dass in diesem Raum das gegebene Problem leichter zu lösen ist.

Typischer Weise lösen kernbasierte Algorithmen 'skalare' Probleme. Zum Beispiel im Falle einer Regression heisst dies, dass das Ergebnis der Regressionsfunktion *eine* Zahl ist. Wir werden uns mit dem allgemeineren Fall eines vektorwertigen Ergebnisses beschäftigen. Dies schafft die Notwendigkeit sich mit dem Begriff der Equivarianz auseinander zu setzen. Dazu führen wir den sogenannten *Reproducing Kernel Hilbert Space* ein:

Definition 1 (Reproducing Kernel Hilbert Space) Sei H ein Hilbertraum von Funktionen $f : X \rightarrow Y$. Man nennt H einen reproducing kernel Hilbert space (RKHS), falls die Abbildung $f \rightarrow \langle y, f(x) \rangle_Y$ stetig auf H ist.

Der einzige Unterschied dieser Definition zum skalaren Fall ist, dass im skalaren Fall $Y = \mathbb{R}$ der Raum der reellen Zahlen ist. Die Definition des RKHS erlaubt nun die Einführung des oben erwähnten Kernes. Dazu müssen wir aber erst erwähnen, dass die geforderte Stetigkeit die Existenz eines sogenannten Auswertefunktional zulässt. Das heißt, unter der Annahme dass $f \in \text{RKHS}$ existiert eine lineare Abbildung $E_x \in L(Y, H)$, so dass $\langle y, f(x) \rangle_Y = \langle E_x y, f \rangle_H$ gilt. Hierbei ist das innere Produkt auf der linken Seite im Vektorraum Y definiert, das heißt im Bild der Funktionen f . Das innere Produkt auf der rechten Seite ist auf dem Raum der Funktionen selbst definiert. Dies bedeutet, dass ein inneres Produkt mit Auswertefunktional $E_x y$ mit einer Funktion f eben genau diese Funktion an der Stelle x auswertet (und nebenbei noch ein inneres Produkt mit y im Ergebnisraum liefert). Man kann sich nun fragen, was das innere Produkt zwischen zwei solchen Auswertefunktionalen bedeutet. Nun, dies genau ist der Kern des RKHS. Formal lässt sich der Kern wie folgt definieren:

$$\langle E_{x_1} y_1, E_{x_2} y_2 \rangle_H = \langle y_1, (E_{x_2} y_2)(x_1) \rangle_Y = \langle y_1, \mathbf{K}(x_1, x_2) y_2 \rangle_Y.$$

Hierbei ist nun $\mathbf{K} : X \times X \rightarrow L(Y)$ der Kern des damit assoziierten RKHS ist. In der Tat ist der Kern eine lineare Abbildung auf dem Ergebnisraum Y . Hier wird der Unterschied zum skalaren Fall deutlich, wo der Kern lediglich eine skalare Zahl ist.

2.1 Equivariante Kerne

Im folgenden wollen wir nun Kerne betrachten, deren zugrundeliegende Funktionenräume die Equivarianzeigenschaft erfüllen. In [RB07, Rei08] konnte gezeigt werden, dass diese Kerne folgende Eigenschaft erfüllen.

Proposition 2 (Equivarianter Kern) Sei $H^{\mathcal{G}}$ ein RKHS von \mathcal{G} -equivarianten Funktionen $f : X \rightarrow Y$, dann gehorcht sein Kern der folgenden Gleichung

$$\mathbf{K}(\tau_g^X \mathbf{x}_1, \tau_h^X \mathbf{x}_2) = \tau_g^Y \mathbf{K}(\mathbf{x}_1, \mathbf{x}_2) \tau_{h^{-1}}^Y.$$

für alle $g, h \in \mathcal{G}$ und $\mathbf{x}_1, \mathbf{x}_2 \in X$. Wir nennen so einen Kern \mathcal{G} -equivariant.

Das wahrscheinlich einfachste Beispiel für einen equivarianten Kern ist das normale innere Produkt in einem komplexen Vektorraum. Hierbei ist die Gruppenwirkung gerade die Multiplikation mit komplexen Einheitszahlen. Formal drückt sich dies wie folgt aus:

$$\mathbf{K}(e^{i\phi_1} \mathbf{x}_1, e^{i\phi_2} \mathbf{x}_2) = \langle e^{i\phi_2} \mathbf{x}_2, e^{i\phi_1} \mathbf{x}_1 \rangle = e^{i\phi_1} \mathbf{K}(\mathbf{x}_1, \mathbf{x}_2) e^{-i\phi_2}.$$

Praktisch stellt sich natürlich die Frage wie man solche equivarianten Kerne nun konstruiert. In dieser Arbeit hat sich herausgestellt, dass die Gruppenintegration hier eine fundamentale Rolle spielt. Nimmt man an, man habe einen Kern gegeben, der nicht die Equivarianzeigenschaft erfüllt, dann lässt sich aus diesem mittels dem folgenden Theorem ein equivarianter Kern konstruieren.

Theorem 3 (GIM-Kern) Sei $\mathbf{K}_0 : X \times X \rightarrow L(Y)$ ein unitärer matrixwertiger Kern. Dann ist

$$\mathbf{K}(\mathbf{x}_1, \mathbf{x}_2) = \int_{\mathcal{G}} \mathbf{K}_0(\mathbf{x}_1, \tau_g^X \mathbf{x}_2) \tau_g^Y dg$$

ein \mathcal{G} -equivariant Kern. Wir nennen \mathbf{K} einen Gruppen-Integrations-Matrix Kern, kurz GIM-Kern.

Es gibt noch andere Möglichkeiten equivariante Kerne zu konstruieren, aber GIM-Kerne zeigen ein ausgezeichnetes Verhalten, was sich im folgenden Theorem manifestiert, welches als theoretisches Hauptresultat der Arbeit zu sehen ist.

Theorem 4 (Equivariant Representer Theorem) Sei $\Omega : \mathbb{R}^+ \mapsto \mathbb{R}$ eine strikt monoton fallende Funktion $c : (X \times Y \times Y)^T \mapsto \mathbb{R}$ und $(\mathbf{x}_i, \mathbf{y}_i)$ mit $i = 1, \dots, T$ eine Menge von Trainingsbeispielen. Weiter, sei \mathbf{K}_0 ein beliebiger matrixwertiger Kern und H sein RKHS. Dann ist jeder equivariante Minimierer $\mathbf{f} \in H$ des Funktionals

$$R(\mathbf{f}) = c(\mathbf{x}_1, \mathbf{y}_1, \mathbf{f}(\mathbf{x}_1), \dots, \mathbf{x}_T, \mathbf{y}_T, \mathbf{f}(\mathbf{x}_T)) + \Omega(\|\mathbf{f}\|^2)$$

von folgender Form

$$\mathbf{f}(\mathbf{x}) = \sum_{i=1}^T \mathbf{K}(\mathbf{x}, \mathbf{x}_i) \mathbf{a}_i,$$

wobei \mathbf{K} der GIM-kernel von \mathbf{K}_0 ist.

2.2 Andere Equivariante Kerne

Neben den bereits eingeführten GIM-Kernen wurden auch zwei anderen Typen von Kernen untersucht. Erstens, **Normalisierungs Kerne**. Das Prinzip der Normalisierung ist das am häufigsten benutzte Invarianzprinzip. Man nehme an, man habe eine Funktion $N : X \rightarrow \mathcal{G}$ die ein Muster \mathbf{x} in eine Standardlage $\mathbf{x}_0 = \boldsymbol{\tau}_{N(\mathbf{x})}^x \mathbf{x}$ versetzt (häufig wird das Muster entlang der Hauptachse des Trägheitstensors ausgerichtet). So eine Funktion erfüllt $N(\boldsymbol{\tau}_g \mathbf{x}) = g^{-1} N(\mathbf{x})$. Man kann damit leicht zeigen, dass

$$\mathbf{K}_N(\mathbf{x}_1, \mathbf{x}_2) = (\boldsymbol{\tau}_{N(\mathbf{x}_1)}^y)^{-1} K_0(\boldsymbol{\tau}_{N(\mathbf{x}_1)}^x \mathbf{x}_1, \boldsymbol{\tau}_{N(\mathbf{x}_2)}^x \mathbf{x}_2) \boldsymbol{\tau}_{N(\mathbf{x}_2)}^y$$

ein equivarianter Kern ist. Hierbei ist K_0 wieder ein unitärer, skalarer Basiskern. Wir nennen diesen Kern N-Kern. Weiter haben wir noch sogenannte **Maximum Matching Kerne** betrachtet. Statt eine Standardlage im Vorhinein zu spezifizieren, werden die Muster während der Kernauswertung in allen möglichen Lagen verglichen. Nimmt man wieder an, man habe einen Basiskern K_0 gegeben, dann ist

$$\mathbf{K}_{\text{MM}}(\mathbf{x}_1, \mathbf{x}_2) = K_0(\mathbf{x}_1, \boldsymbol{\tau}_{g^*}^x \mathbf{x}_2) \boldsymbol{\tau}_{g^*}^y, \quad \text{wobei } g^* = \underset{g \in \mathcal{G}}{\operatorname{argmax}} K_0(\mathbf{x}_1, \boldsymbol{\tau}_g^x \mathbf{x}_2)$$

ein equivarianter Kern. Wir nennen so einen Kern Maximum-Matching Kern, kurz MM-Kern.

3 Rotationsinvariante Objektdetektion mittels der verallgemeinerten Houghtransformation

Die verallgemeinerte Houghtransformation ist eine generische Methode um geometrisch ähnliche Objekte in verrauschten Szenen mit beliebigem Hintergrund zu detektieren und zu erkennen. Der Methode liegt die Idee zugrunde dass alle Pixel in der unbekanntem Szene für mögliche Objektpositionen des gesuchten Objekts abstimmen. Die Positionen, die viele Stimmen erhalten haben sind dann Kandidaten für die Position des gesuchten Objekts. Das einfachste Beispiel ist die Detektion eines weißen Kreises mit definiertem Radius r auf schwarzem Hintergrund. Während der Detektion stimmt jeder weiße Pixel für alle Positionen, die den Abstand r zum aktuellen Pixel haben. Es ist klar, falls sich ein Kreis in der Szene befindet, wird das Zentrum des Kreises viele Stimmen erhalten. Allerdings ist dies ein sehr einfaches Beispiel, für komplexe Szenen in beliebigen Grauwertbildern wird das Vorgehen modifiziert. Bevor ein Pixel seine Stimmen abgibt wird erst die Nachbarschaft des Pixels analysiert und entsprechend der Gestalt der Nachbarschaft die Stimmenverteilung berechnet. In der Literatur ist diese Idee unter anderem als 'Implicit Shape Model' bekannt [LLS04]. Implizit, weil die Form und Gestalt des gesuchten Objekts in eben dieser Abbildung, welche kleine Teile (d.h. die lokalen Nachbarschaften der Pixel des Objekts) auf Wahrscheinlichkeitskarten für die Objektposition abbildet, impliziert kodiert ist. Und genau diese Abbildung möchten wir nun mit den eingeführten Methoden modellieren. Die Eingangsmuster \mathbf{x} beschreiben das Aussehen der lokalen Nachbarschaften von bestimmten Pixeln. Diese werden dann mittels einer Funktion auf die Stimmenverteilung $\mathbf{y} = \mathbf{f}(\mathbf{x})$

abgebildet. Nun möchten wir, dass das Objekt unabhängig von seiner Orientierung in der Bildebene gefunden wird. Was bedeutet dies nun für das Abstimmungsfunktional f ? Es ist ähnlich wie im einführenden Proteinbeispiel zur Equivarianz: wenn sich das Objekt dreht, dann drehen sich entsprechend auch die einzelnen 'Objektbestandteile' \mathbf{x} mit und somit müssen sich die Stimmenverteilungen \mathbf{y} mitdrehen um konsistent für die Objektposition abstimmen zu können. Das heisst das Abstimmungsfunktional f muss equivariant bzgl. Drehungen sein.

Es sind noch einige Zutaten nötig um diese Idee nun effizient umsetzen zu können. Erstens, man muss nicht unbedingt alle Pixel im Bild eine Stimme abgeben lassen. Pixel in kontrastarmen Regionen liefern keine Information. Somit wird zuvor eine bestimmte Teilmenge mit einem sogenannten Interest-Point Detektor ausgewählt. Die Idee ist einfach, es werden bestimmte Varianzkarten des Bildes erstellt und lokale Maxima dieser Varianzkarten werden als 'interessante Punkte' definiert. Hier gibt es in der Literatur eine Vielzahl von Vorschlägen, wir haben uns für [Lin98] entschieden. Die Beschränkung auf eine Teilmenge von Pixeln hat zudem den Vorteil die Laufzeit des Algorithmus deutlich zu verkürzen.

Zweitens muss die Nachbarschaft eines Pixels sinnvoll repräsentiert werden, hier sind die SIFT-Deskriptoren [Low04] die wohl bekanntesten Vertreter (Vertreter des 'Implicit Shape Model' benutzen genau solche lokalen Deskriptoren). Allerdings haben wir noch eine zusätzliche Anforderung an die lokalen Merkmale, sie müssen eine effiziente Berechnung der equivarianten Kerne zulassen. Eine zwingende Voraussetzung dafür ist, dass die Merkmale selbst equivariant sind. Das heisst, angenommen wir haben einen Merkmalsvektor $\Psi(\mathbf{x})$ der lokalen Nachbarschaft zugeordnet, dann muss sich aus diesem Merkmalsvektor auch der Merkmalsvektor der gedrehten Nachbarschaft berechnen lassen, das heisst, $\Psi(\tau_g \mathbf{x}) = \tau_g \Psi(\mathbf{x})$. Die direkte Anwendung der SIFT-Deskriptoren scheidet hierfür aus, da sie selbst rotationsinvariant sind, somit enthalten die Merkmale selbst nicht mehr die originale Rotationsinformation. Die Invarianz wird hier durch Normalisierung bzgl. lokal extrahierten Hauptrichtungen erreicht. Es wurde sich allerdings beim Entwurf equivarianten Merkmale stark an den SIFT-Deskriptoren orientiert.

Schliesslich muss auch die interne Repräsentation der Stimmenverteilungen sinnvoll gewählt werden. Zum einen muss sie kompatibel zur Berechnung der equivarianten Kerne sein, und zum anderen muss es möglich sein die Verteilung effizient für die Objektposition 'abstimmen' zu lassen. Hier wurde eine parametrische Beschreibung in Polarkoordinaten gewählt, wobei die Winkelkomponente im Fourierbereich dargestellt wird. Die Wahl einer parametrischen Beschreibung ist in diesem Zusammenhang neu. In der Literatur werden diese Verteilungen typischer Weise nicht-parametrisch beschrieben (als sogenannte 'Codebook'-Ansätze bekannt). Nur diese nicht-parametrischen Beschreibung hat es in der Vergangenheit möglich gemacht mittels Normalisierung die Equivarianz des Abstimmungsfunktionals zu erreichen, da parametrische Entwicklungen von equivarianten Funktionen in der hier vorgestellten Art und Weise noch nicht bekannt waren.

3.1 Die lokalen Merkmale

Um jeden erkannten Interest-Point wird eine Umgebung von 29×29 Pixeln extrahiert. Für jede dieser Umgebungen \mathbf{x} soll nun ein equivarianter Merkmalsvektor $\Psi(\mathbf{x})$ berechnet werden. Wir bezeichnen die Komponenten des Merkmalsvektors mit $\psi_j^l(\mathbf{x})$ wobei der Index l zur Fourierdarstellung der Winkeldimension dient und j eine zusätzliche Merkmalsdimension ist. Die Equivarianzbedingung formuliert sich dann als $\psi_j^l(\tau_\phi^x \mathbf{x}) = e^{il\phi} \psi_j^l(\mathbf{x})$. Wir folgen dem Geiste der SIFT-Deskriptoren und legen den Merkmalen die Gradienten der Grauwerte des Bildes zu Grunde. Das Merkmal ist ein dreidimensionales Verbundhistogramm über, erstens, die Distanz zum Interest-Point z und, zweitens, dem relativen Winkel zwischen Gradient und dem Differenzvektor zum Interest-point, und schliesslich, dem absoluten Winkel des Differenzvektors. Die ersten beiden Histogrammdimensionen sind mit dem Index j assoziiert und die dritte Dimension wird im Fourierbereich dargestellt und bezieht sich auf den Index l .

3.2 Implementation der Kerne und des Trainings

Die Implementation des N-kernes ist relativ einfach, da er sich weitestgehend an vorhandenen Ideen aus der Literatur orientiert. Hier können zum Beispiel die Orientierungsschätzungen, die der Berechnung der SIFT oder GLOH-Merkmalen zu Grunde liegen als Normalisierer N benutzt werden. Um eine schnelle Berechnung der GIM-Kerne zu gewährleisten kann eine FFT benutzt werden. Die FFT führt genau die gewünschte Gruppenintegration aus. Ähnlich kann die FFT auch benutzt werden um die beste Übereinstimmung zu finden, was beim MM-Kern notwendig ist. Zum Training der Abstimmungsfunktion wurde eine normale 'gekernelte' lineare Regression verwendet, hierbei setzen sich die Trainingsbeispiele aus den lokalen Nachbarschaften \mathbf{x}_i gepaart mit den Abstimmungskarten \mathbf{y}_i zusammen. Dabei hat jede individuelle Abstimmungskarte \mathbf{y}_i genau an der Stelle des Zentrum des Objekts eine 1 und sonst 0en. Man bedenke, dass die Position der 1 immer relativ zu der zugehörigen Position des Interest-Points zu sehen ist.

4 Experimente

Kommen wir nun zu einem Anwendungsbeispiel. Es wurde sich als Aufgabe gestellt auf Luftbildern von Flughäfen die Flugzeuge zu erkennen. Die Bilder sind etwa von einer Größe 1200×1000 (siehe Abbildung 3). Ein Flugzeug hat etwa eine Länge von 50 Pixeln. Alle Bilder in unseren Experimenten zeigen einen komplexen Hintergrund, also viel Kandidaten für Fehldetektionen. Auch die Lichtverhältnisse sind sehr variabel. Abbildung 2 zeigt die 6 (!) Bilder, die zum Training verwendet wurden. Sie wurden so gewählt, dass möglichst alle Schattenverhältnisse präsent sind.

Abbildung 3 ist eine Beispiel für die Fähigkeiten des System. Die Berechnungszeit hängt sehr von der Komplexität der Szene ab (d.h. von der Anzahl der Interest-Points). Für das

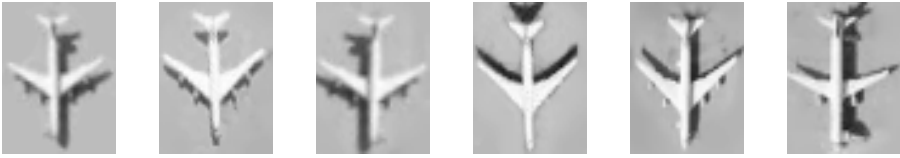


Abbildung 2: **Trainingsbilder.** Diese sechs Bilder wurden zum Training benutzt.

gegebene Beispiel, eines von mittlere Komplexität, braucht der Algorithmus auf einem Standard-PC etwa 10 Sekunden. In Abbildung 3 sind zusätzlich auch Orientierungsschätzungen der Objekte dargestellt. Diese wurden auf ähnliche Weise wie die Positionen, mit einer Abstimmung über die Orientierung erzeugt.

Zur Quantifizierung der Ergebnisse werden in Abbildung 4a) die drei eingeführten equi-varianten Kerne verglichen. Der GIM-kern schneidet am besten ab, dicht gefolgt vom N-kern und MM-kern. Es ist verwunderlich, dass der MM-kern leicht schlechter abschneidet als der N-kern. Man hätte erwartet dass ein 'matching' -Ansatz zuverlässiger ist als eine Normalisierung bzgl. der Orientierung. Aber insgesamt, sind die Unterschiede nicht sehr signifikant.

Schliesslich soll nun noch der Unterschied zwischen den klassischen nicht-parametrischen Abstimmungstechniken [LLS04] und den hier vorgeschlagenen parametrischen Techniken quantifiziert werden. Und zusätzlich werden auch noch die hier eingeführten EGH-Merkmale mit den GLOH-Merkmalen [MS05] verglichen. Die Ergebnisse sind in Abbildung 4b) zusammengefasst. Man kann einen deutlichen Unterschied zwischen dem parametrischen und nicht-parametrischen Ansatz erkennen. Der parametrische Ansatz schneidet deutlich besser ab. In Abbildung 5 sind beispielhaft zwei Abstimmungskarten für den parametrischen und nicht-parametrischen Ansatz gezeigt. Offenbar hat der nicht-parametrische Ansatz Probleme in Bereichen, die viele zusätzliche Interest-Points beinhalten, die nicht zu einem Flugzeug gehören.

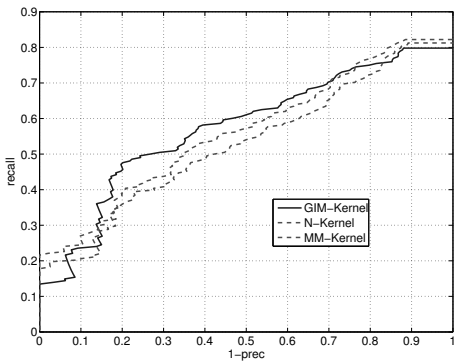
Schliesslich kann man noch feststellen, dass die Unterschiede zwischen den GLOH und EGH nicht signifikant sind.

5 Zusammenfassung

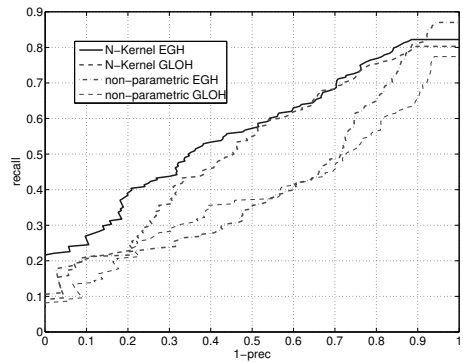
Das Hauptresultat dieser Arbeit ist, dass sich mit dem Prinzip der Gruppenintegration Equivarianzbedingungen optimal in Regressionsaufgaben integrieren lassen. Im speziellen wurde dieses Resultat in der Kern-Theorie ausformuliert, was zum Begriff des GIM-Kerns führte. Als weiterführende Arbeiten lassen sich die Entwicklung von 'margin-based' Lernalgorithmen nennen, die zusätzlich die Equivarianz nicht verletzen. Weiter steht eine Erweiterung der Theorie auf nicht kompakte Gruppen aus. Ist diese Erweiterung möglich, würde sich die sogenannte Volterra-Theorie [BCD84] als Spezialfall der vorgestellten Theorie erweisen.



Abbildung 3: **Erkennungsbeispiele.** Die gelben Rechtecke markieren die Erkennungen. Die gelben Punkte geben Positionen lokaler Maxima der Abstimmungskarten an. Die weißen Pfeile zeigen Probleme bei der Erkennung an.



(a) Verschiedene equivariante Kerne.



(b) Nicht-parametrisch versus parametrischer Ansatz

Abbildung 4: **PR-Graph zum Vergleich mit Referenzmethoden.**

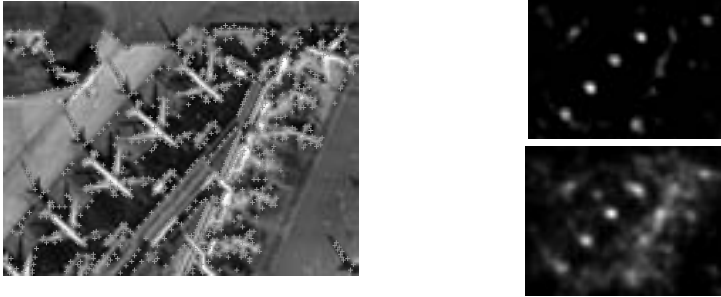


Abbildung 5: **Abstimmungskarten**. Links: Originalbild mit den verwendeten Interest Points. Rechts: parametrischer Ansatz (oben) und nicht-parametrischen Ansatz (unten).

Literatur

- [BCD84] S. Boyd, L.O. Chua und C.A. Desoer. Analytical Foundations of Volterra Series. *IMA Journal of Mathematical Control and Information*, 1:243–282, 1984.
- [Lin98] T. Lindeberg. Feature detection with automatic scale selection. *International Journal of Computer Vision*, 30(2):77–116, 1998.
- [LLS04] B. Leibe, A. Leonardis und B. Schiele. Combined Object Categorization and Segmentation with an Implicit Shape Model. In T. Pajdla und J. Matas, Hrsg., *Proceedings of the ECCV'04 Workshop on Statistical Learning in Computer Vision, Prague*. LNCS, Springer, 2004.
- [Low04] D.G. Lowe. Distinct Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, 60:91–110, 2004.
- [MS05] K. Mikolajczyk und C. Schmid. A Performance Evaluation of Local Descriptors. 27(10):1615–1630, 2005.
- [RB07] M. Reisert und H. Burkhardt. Learning Equivariant Functions with Matrix Valued Kernels. *J. Mach. Learn. Res.*, 8:385–408, 2007.
- [Rei08] M. Reisert. *Group Integration Techniques in Pattern Analysis - A Kernel View*. Dissertation, Albert-Ludwigs Universität Freiburg, Juli 2008.
- [SS02] B. Schölkopf und A. J. Smola. *Learning with Kernels*. The MIT Press, 2002.



Marco Reisert wurde geboren am 21. Dezember 1976 in Freiburg. Nach dem Studium der Informatik erhielt er 2004 den Titel des Diplom-Informatikers. Danach arbeitete er als Doktorand am Lehrstuhl für Mustererkennung und Bildverarbeitung in Freiburg, wo er 2008 als Dr.rer.nat abgeschlossen hat. Inzwischen ist er in der medizinischen Physik des Universitätsklinikums Freiburg beschäftigt. Seine Forschungsinteressen liegen in der mehrdimensionalen Signalverarbeitung, Invarianz in der Mustererkennung und Diffusionsgewichteten-MR Bildgebung.