

# Orts-relationale SIFT-Hierarchien zur Ähnlichkeitsbestimmung mit Graph-Matching

Armin Fritsche, Benedikt Fischer, Thomas M. Deserno

Institut für Medizinische Informatik, RWTH Aachen  
afritsche@imib.rwth-aachen.de

**Kurzfassung.** Zur Bestimmung von Bildähnlichkeiten werden häufig charakteristische Punkte berechnet. Ein bewährtes Verfahren zur Extraktion solcher Punkte ist die Scale Invariant Feature Transform (SIFT). Gerade auf medizinischen Grauwertbildern liefert dieses Verfahren aber unbefriedigende Ergebnisse. In dieser Arbeit wird ein neuer Ansatz vorgestellt, mit dem auch strukturelle Informationen über die Verteilung der SIFT-Punkte im Bild gespeichert und verglichen werden können. Das Verfahren wurde zur Knochenaltersschätzung auf Handradiographien angewendet. Im Vergleich zum Standard SIFT wurde Recall um 4 % verbessert.

## 1 Einleitung

Zum Vergleich zweier Bilder existieren zahlreiche Ansätze. Die meisten Methoden basieren dabei auf der Segmentierung der Bilder und dem Vergleich der Regionen. Die Beschreibungen der Regionen sind aber häufig anfällig gegen Rotationen, Skalierungen oder Formänderungen. Daher werden anstelle der Regionen auch charakteristische Punkte, sog. Interesting Points (IPs), extrahiert. Eine weit verbreitete Technik zur Bestimmung und Beschreibung dieser Punkte ist das SIFT-Verfahren [1]. Die Deskriptoren dieser Punkte sind größtenteils invariant gegenüber Rotation, Skalierung und kleineren Änderungen des Betrachtungswinkels. Jeder SIFT-Punkt wird dabei durch seinen Deskriptor, seine Orientierung und seine Koordinaten im Skalenraum  $(x, y, \sigma)$  beschrieben, wobei  $\sigma$  seine Skalierung ist. Der Deskriptor ist ein reellwertiger Vektor. Zwei Bilder werden verglichen, indem man die SIFT-Punkte miteinander vergleicht. Der Vergleich wird dabei über die euklidische Distanz der Deskriptorvektoren vorgenommen. Ein Matching zwischen zwei Nearest Neighbour Punkten wird vorgenommen, wenn ihre Distanz unter einem Schwellwert liegt.

Im medizinischen Bereich ist dieses Verfahren zur Registrierung von MRT-Aufnahmen des Gehirns getestet [2]. Die Registrierung fand dabei allerdings nur auf künstlich deformierten Bildern statt. In [3] wird eine Klassifikation von medizinischen Bildern anhand von SIFT-Features vorgenommen. Trotz akzeptablen Ergebnis wird empfohlen, andere Deskriptoren, wie etwa Thumbnails der Bilder, zu benutzen. Ferner wird eine Aufteilung in grobe Bildklassen (z.B. Handradiographie, Schädel-CT) vorgenommen [4]. Hierbei werden einerseits künstlich

deformierte Fingerepiphysen miteinander verglichen und andererseits wird das Knochenalter durch einen Vergleich mit Epiphysen bekannten Alters bestimmt. Während im ersten Schritt noch ca. 87% der SIFT-Punkte korrekt zugeordnet werden konnten, wird bei der Altersbestimmung nur bei 2% der Bilder eine korrekte Schätzung erreicht.

Bei bisherigen Arbeiten in der medizinischen Bildverarbeitung wurde das SIFT-Verfahren also entweder nur zur groben Klassifikation oder auf künstlich veränderten Bildern erfolgreich angewendet. Für die meisten klinischen Anwendungen scheint das Verfahren nicht genügend Informationen bereitzustellen. Einer der Gründe ist die lokale Betrachtungsweise des Verfahrens, bei der die Verteilung der Punkte im Bild außer Acht gelassen wird. Diese Arbeit beschäftigt sich daher mit einer Methode, um Informationen über die Lage der IPs im Bild in einen Bildvergleich mit einbeziehen zu können.

## 2 Material und Methoden

In [5] werden Bilder über Bildregionen verglichen. Dabei werden nicht nur einzelne Bildregionen betrachtet, sondern auch Beziehungen zwischen den Regionen. Alle Regioneninformationen werden in einem gewichteten Graphen gespeichert. Der Vergleich wird über ein Graphmatching vorgenommen. Das selbe Prinzip wird hier verwendet, indem statt Informationen zu den Regionen und ihrer Inklusionsrelation eine entsprechende Baumstruktur der SIFT-Punkt-Hierarchie erzeugt wird, so dass der Bildvergleich durch Graphmatching erfolgen kann.

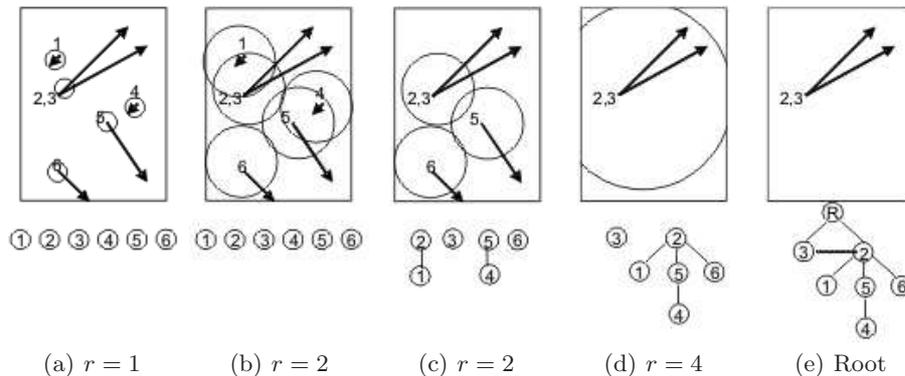
### 2.1 Grapherzeugung

Zunächst werden auf den Anfragebildern die SIFT-Punkte berechnet. Für jeden dieser Punkte wird dann ein Knoten im Graphen erzeugt. Die Knoten werden mit einem Deskriptor, seiner Orientierung und seinen Koordinaten im Skalerraum annotiert. Die Kantenstruktur wird aus der adaptiven nicht-maximalen Suppression (ANMS) [6] abgeleitet. Dort dient die ANMS einerseits zur Unterdrückung schlechter IPs und andererseits zur besseren Verteilung der IPs im Bild. Das Prinzip ist hierbei, dass IPs, die in ihrer lokalen Nachbarschaft nicht maximale Skalierung haben, aussortiert werden. Während bei Brown die Aussortierung über Löschung abläuft, wird in dieser Arbeit eine hierarchische Struktur eingeführt. Dafür wird um jeden SIFT-Punkt ein Kreis mit Radius  $r$  gelegt, der seine lokale Nachbarschaft definiert. Die Kreise um die SIFT-Punkte wachsen im Laufe der Grapherzeugung sukzessiv an. Ist ein Kreis soweit gewachsen, dass er nicht nur einen, sondern mehrere SIFT-Punkte enthält, wird der SIFT-Punkt mit maximalem  $\sigma$ -Wert gesucht und hierarchisch über alle anderen SIFT-Punkte des Kreises erhöht. Das heißt, im Graphen wird der Knoten des maximalen SIFT-Punktes zum Vaterknoten aller anderen Knoten in der selben Nachbarschaft. Alle Knoten bzw. SIFT-Punkte, die in ihrer Nachbarschaft nicht mehr maximal sind, entfallen aus der Betrachtung (Abb. 1).

Da es durch das SIFT-Verfahren zu mehreren IPs mit den selben Koordinaten, aber mit verschiedenen Orientierungen kommen kann, muss auch dieses in der Baumstruktur repräsentiert werden. Solche Knoten werden mit einem neuen Kantentyp, den Adjazenzkanten verbunden. Diese Punkte können beim Bildvergleich beliebig vertauscht werden. Um stets eine eindeutige Baumstruktur zu erhalten (abgesehen von den Adjazenzkanten), muss abschließend ein künstlicher Wurzel-Knoten eingefügt werden. Dies wird etwa dann nötig, wenn es mehrere maximale SIFT-Punkte im Bild gibt.

### 2.2 Relationale Merkmale und Bildvergleich

Durch die Struktur des Graphen sind schon erste Informationen über die Verteilung der IPs im Bild gespeichert. Diese geben aber zu wenig Aufschluss über die Beziehungen der Punkte untereinander. Daher werden beim Vergleich zweier Graphen zusätzlich relationale Merkmale berechnet und als Kantengewichtung in den Graphen aufgenommen. Relationale Merkmale beschreiben die Beziehung zweier IPs zueinander. Hierbei ist darauf zu achten, dass alle verwendeten Merkmale aus den Knotengewichtungen berechnet werden können und jeweils in Relation zu den SIFT-Merkmalen stehen. Durch Wahl anderer Relationen (z.B. Farbvergleich, Ortsvergleich) würde man riskieren, die Invarianzeigenschaften der SIFT-Punkte zu verlieren. Für diese Arbeit wird die Orientierungsdifferenz, die Skalierungsdifferenz, die Entfernung in Relation zu ihrer Skalierung, die Lage



**Abb. 1.** (a) Jeder SIFT-Punkt wird durch seine Orientierung und Skalierung dargestellt. Um jeden Punkt wird ein Kreis gelegt. (b) Die Kreise wachsen, bis mehrere SIFT-Punkte verschiedener Skalierung in einem Kreis liegen. (c) Der jeweils maximale SIFT-Punkt in einem Kreis erhöht sich in der Hierarchie. Ist der maximale Punkt nicht eindeutig, wird einer zufällig gewählt. Punkte, die in ihrer Nachbarschaft mit Radius  $r$  nicht mehr maximal sind, fallen aus der Betrachtung heraus. (d) Dieses Verfahren wird immer weiter fortgesetzt, bis nur noch maximale Punkte im Bild existieren. (e) Abschließend werden noch die adjazenten Knoten 2 und 3 mit Adjazenzkanten verbunden, und es wird ein Wurzel-Knoten eingefügt, der Vater der Knoten 2 und 3 ist.

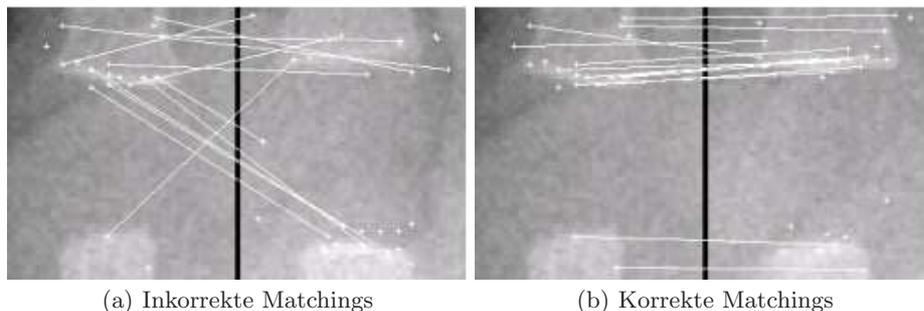
in Relation zur Orientierung und die euklidische Distanz zwischen den Deskriptoren verwendet.

Für zwei Bilder bekommt man demnach zwei knoten- und kantengewichtete Graphen, die sich mit einem Matchingverfahren vergleichen lassen. Hierbei ist zu beachten, dass Knoten, die über Adjazenzkanten verbunden sind, beliebig ausgetauscht werden können. In dieser Arbeit wurde ein Graphmatchingverfahren angewendet, welches auf evolutionärer Spieltheorie basiert [5]. Die Ausgabe des Graphmatchers sind schließlich Knoten- und somit auch SIFT-Punktzuordnungen.

Diese Punktzuordnungen werden aus Epiphysen-ROI von Handradiographien ausgewertet (Abb. 2). Auf Grundlage der Epiphysenvergleiche zu Epiphysen mit bekannten Alter lässt sich schließlich eine Schätzung des Knochenalters von Epiphysen unbekanntes Alters vornehmen. Die Testmenge besteht aus 98 Epiphysen unterschiedlichen Alters. Hierbei wird jede ROI mit jeder anderen verglichen. Zur Evaluation werden einerseits die Anzahl der gefundenen Matchings gezählt, andererseits werden korrekte Zuordnungen ermittelt und in Relation zu den Gesamtzuordnungen (Precision) und zu der Anzahl der extrahierten IPs (Recall) gesetzt. Zur Bewertung wird das Ergebnis mit dem Standard SIFT-Verfahren [1] verglichen.

### 3 Ergebnisse

Die Ergebnisse aus Tabelle 1 basieren auf den insgesamt 9506 Bildvergleichen. Die Bewertung, ob ein Matching korrekt ist, wurde in beiden Fällen anhand einer Ground Truth vorgenommen (Abb. 2). Während bei der normalen SIFT-Methode eine leicht bessere Precision erreicht wird, lässt sich der Recall beim Graphvergleich annähernd verdoppeln. Im Einzelnen wurden im Standard SIFT-Verfahren ca. 2,03 Punktmatchings pro Bildvergleich gefunden (0,75 korrekte). Im Graphmatchingansatz werden durchschnittlich 5,16 Punktmatchings pro Bildvergleich gefunden (1,53 korrekte).



**Abb. 2.** Beispiele für zwei Epiphysen-ROI Vergleiche mittels SIFT-Graphmatching. Zur besseren Ansicht wurde der Kontrast nachträglich verstärkt.

**Tabelle 1.** Ergebnisse.

Verfahren	Matchings		Precision	Recall
	gefunden	richtig	in %	in %
Standard-SIFT	19314	7205	23,53	5,20
SIFT-Graphen	49089	14583	22,20	9,44

## 4 Diskussion

Die schlechte Precision der verglichenen Verfahren kann durch den Aufbau der SIFT-Deskriptoren erklärt werden. Diese bestehen aus Gradientenhistogrammen. Sind sich mehrere Gradienten sehr ähnlich, etwa an den Knochenkanten, so kann keine eindeutige Zuordnung gewählt werden. Andererseits werden durch verschiedene Intensitätsänderungen auch viele Matchings fälschlich verworfen. Durch die Betrachtung der Lage der Punkte, werden einerseits ähnliche Punkte unterscheidbar, andererseits wird die harte Bedingung der korrekten Gradienten aufgeweicht. Daher können mehr Zuordnungen berechnet werden.

Insgesamt konnte ein Verfahren gezeigt werden, welches IPs um weitere Bildinformationen ergänzt und es daher auf neuen Anwendungsgebieten einsetzbar macht. Für weitere Arbeiten empfiehlt sich eine genauere Untersuchung der relationalen Merkmale, da hier zunächst recht einfache gewählt wurden, sowie ein Wechsel der IPs.

**Danksagung.** Diese Arbeit entstand im Rahmen des DFG Projektes „Strukturierte Prototypen in Radiologischer Routine“ (Le 1108/9)

## Literaturverzeichnis

1. Lowe DG. Distinctive image features from scale-invariant key points. *Int J Computer Vis.* 2004;60(2):91–110.
2. Moradi M, Abolmaesumi P, Mousavi P. Deformable registration using scale-space keypoints. *Proc SPIE.* 2006;6144:791–8.
3. Deselaers T, Keysers D, Ney H. Features for image retrieval: An experimental comparison. *Inf Retr Boston.* 2008;11(2):77–107.
4. Wittenhagen M. Verwendung von SIFT-Features für die Evaluierung von Bildähnlichkeiten [Studienarbeit]. Institut für Medizinische Informatik, RWTH Aachen; 2008.
5. Fischer B, Fritsche A, Thies C, et al. Evolutionäres Graphmatching zur Handknochen-Identifikation. *Proc BVM.* 2009; p. 26–30.
6. Brown M, Szeliski R, Winder S. Multi-image matching using multi-scale oriented patches. *Proc CVPR.* 2005;1:510–7.