

# Bestimmung der Korrespondenz zwischen historischen Gebäudeaufnahmen basierend auf der Zuordnung geometrischer Merkmale

FERDINAND MAIWALD<sup>1</sup>, DANILO SCHNEIDER<sup>1</sup> & FRANK HENZE<sup>1</sup>

*Zusammenfassung: Dieser Beitrag befasst sich mit der Korrespondenzbestimmung historischer Gebäudeaufnahmen aus der Fotothek der Sächsischen Landesbibliothek – Staats- und Universitätsbibliothek Dresden (SLUB). Im Gegensatz zu aktuellem Bildmaterial sind bei historischen Fotografien verschiedene Faktoren ausschlaggebend, die eine automatische Bildanalyse (Merkmalssuche, Merkmalszuordnung und relative Orientierung) deutlich erschweren. Aufgrund dessen werden in dem hier präsentierten Ansatz ausschließlich geometrische Merkmale verwendet, um die Verknüpfung zwischen zwei oder mehreren historischen Aufnahmen zu ermöglichen. Als robuste Merkmale haben sich vor allem Rechtecke im Objektraum herausgestellt. Diese können zum einen relativ leicht und schnell detektiert werden und zum anderen zeigen Rechtecke bei historischen Gebäuden oft eindeutige Formationen (Lage und Anordnung von Fenstern, Steinen, Verkleidungen) und können so in anderen Bildern wiedererkannt werden.*

## 1 Einleitung

Im Rahmen des interdisziplinären Forschungsprojektes „UrbanHistory4D“ soll der Zugang zu digitalen Bildrepositorien für die Erstellung räumlich-zeitlicher Modelle (4D) sowie deren Nutzung in der stadthistorischen Forschung und Wissensvermittlung untersucht werden. Ein wichtiges Teilprojekt in diesem Forschungsvorhaben ist die automatisierte Erschließung größerer digitaler Bildbestände für eine photogrammetrische 3D-Auswertung.

Dieser Beitrag befasst sich mit der Verknüpfung und anschließender relativen Orientierung historischer Aufnahmen im Kontext eines historischen dreidimensionalen Stadtmodells der Stadt Dresden. Das Ziel ist eine Darstellung von historischen Bildern und Plänen in einem dreidimensionalen Modell. Notwendig für die Überlagerung von Bild und Modell ist die Bestimmung der Kameraposition und -orientierung. Grundlage dafür bildet das Bildarchiv der Fotothek der Sächsischen Landesbibliothek – Staats- und Universitätsbibliothek Dresden (SLUB), das zum jetzigen Zeitpunkt ca. 2 Millionen Bilder aus 88 Institutionen enthält. Die Aufnahmen wurden größtenteils zwischen 1940 und 1990 gemacht (deutschefotothek.de). Die Untersuchungen konzentrieren sich zunächst auf Gebäudeaufnahmen des historischen Stadtzentrums von Dresden. Im Gegensatz zu aktuellem Bildmaterial sind bei historischen Fotografien verschiedene Faktoren ausschlaggebend, die eine automatische Bildanalyse (Merkmalssuche, Merkmalszuordnung und relative Orientierung) deutlich erschweren:

---

<sup>1</sup> Technische Universität Dresden, Institut für Photogrammetrie und Fernerkundung, Helmholtzstraße 10, D-01062 Dresden, E-Mail: [ferdinand.maiwald, danilo.schneider, frank.henze]@tu-dresden.de

- Durch Digitalisierung der Aufnahmen in geringer Auflösung gehen Bildinformationen verloren.
- Über die Digitalisierung liegen keine Informationen vor. Beim Scan von Bildausschnitten kann der Bildhauptpunkt am Rand oder sogar außerhalb des digitalen Bildes liegen.
- Für die verwendeten Kameras liegen keine Kameraparameter und Kamerastandpunkte vor. Innere und äußere Orientierung der Kameras sind somit teilweise oder gänzlich unbekannt.
- Aufgrund verschiedener Ursachen (Filmkörnigkeit, Staubpartikel, Digitalisierungsprozess...) weisen die digitalisierten Aufnahmen ein Bildrauschen (Noise) auf, das die Objekttextur (Signal) überlagern kann. In einigen Bildern entsteht ein ungünstiges Signal-Rausch-Verhältnis (signal-to-noise ratio, SNR), das eine automatisierte Merkmalsuche erschwert.
- Die radiometrischen Differenzen zwischen zwei Bildern derselben Epoche sind in den meisten Fällen sehr groß (Abbildung 1).

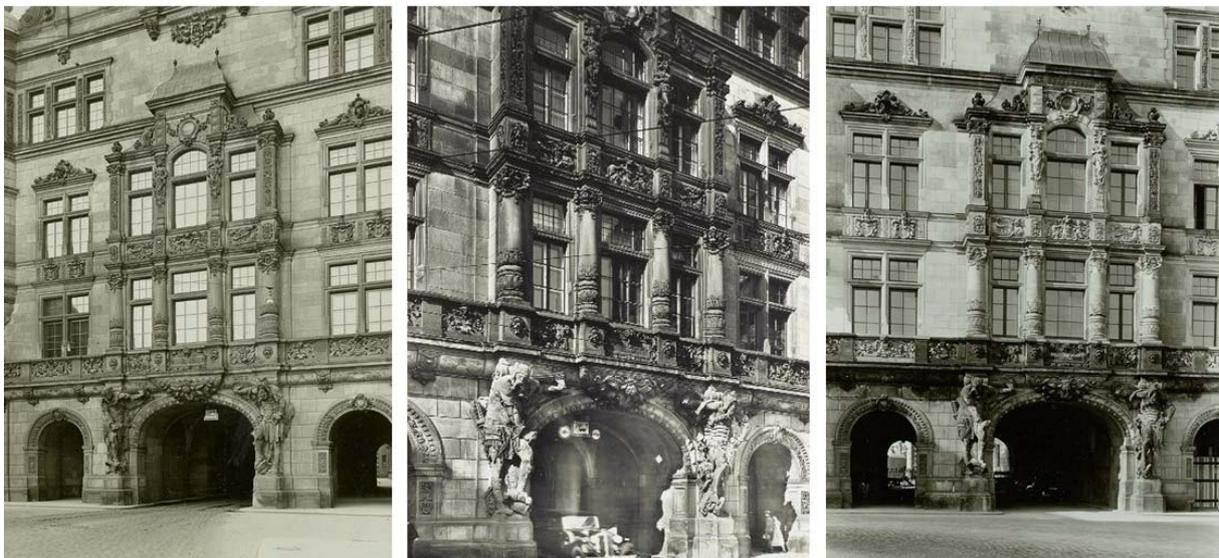


Abb. 1: Beispiel für radiometrische Differenzen zwischen drei historischen Aufnahmen des Georgentors in Dresden

Herkömmliche Matching-Verfahren wie Brute-Force-Matching und Flann-based-Matching (MURAJA & LOWE 2009) führten mit Deskriptoren wie SIFT (LOWE 2004), SURF (BAY 2006) und ORB (RUBLEE et al. 2011) zu keinen bzw. wenigen korrekten Zuordnungen bei Anwendung auf zwei historische Aufnahmen.

In dem hier präsentierten merkmalsbasierten Ansatz werden daher ausschließlich *geometrische* Merkmale verwendet, um die Verknüpfung zwischen zwei oder mehreren historischen Aufnahmen zu ermöglichen. Als robuste Merkmale haben sich dabei zunächst rechteckige Strukturen im Objektraum herausgestellt. Diese können zum einen relativ leicht und schnell als allgemeine Vierecke im Bildraum detektiert werden und zum anderen zeigen Rechtecke bei historischen Gebäuden oft eindeutige Formationen (Lage und Anordnung von Fenstern, Steinen, Verkleidungen) und können so in anderen Bildern über semantisch-topologische Beziehungen wiedererkannt werden.

## 2 Stand der Forschung

Die Arbeit mit historischen Fotografien wird durch die zunehmende Erschließung und Digitalisierung von Datensätzen ein immer größer werdendes Forschungsfeld. Sowohl Forschungsplattformen wie Europeana (europeana.eu), prometheus (prometheus-bildarchiv.de) oder die deutsche Fotothek (deutschefotothek.de) als auch Seiten auf denen private Nutzer historische Bilder verwalten können wie rephotos (re.photos) oder Mapillary (mapillary.com) vergrößern jährlich ihren Bildbestand. Diese Bilddaten werden meist in einer klassischen Bilddatenbank abgelegt, in der mittels Stichwortsuche gesucht werden kann. Um diese Datenmengen besser zugänglich zu machen, kann zusätzlich eine räumliche Darstellung hilfreich sein. Für diese Art der Darstellung ist die Position des Bildes im Modell und damit eine Bestimmung der Kameraposition und Aufnahmerichtung (äußere Orientierung) notwendig. SCHINDLER & DELLAERT (2012) entwickelten mit „4D Cities“ bereits eine Plattform die mittels Structure-from-Motion große Bilddatenbestände in einem vierdimensionalen Modell darstellt und lokalisiert. Allerdings erreichte die automatisierte Variante mittels SIFT-Merkmalen deutlich weniger korrekte Zuordnungen als die manuelle Punktzuordnung in verschiedenen historischen Bildern durch einen Operateur (SCHINDLER & DELLAERT 2012).

In weiteren Publikationen wurden zusätzliche Deskriptoren und Matching-Verfahren auf historische Bilder angewendet. Eine Kombination aus SIFT/SURF und ORB erreichte die größte Anzahl an korrekten Zuordnungen, wobei der Digitalisierungsprozess stets als Grund für große Unsicherheiten in Merkmalsuche und -zuordnung angegeben wurde (WOLFE 2015), (ALI & WHITEHEAD 2015). Eine Weiterverarbeitung der homologen Punkte zur Bestimmung der relativen Orientierung von zwei Bildern ist daher kritisch zu sehen. Vor allem periodische Strukturen und ein ungünstiges Verhältnis zwischen Bildrauschen und Objekttextur sorgen bei der Arbeit mit historischen Gebäudeaufnahmen für wenige robuste Merkmalspunkte und fehlerhafte Punktzuordnungen. Weitere Fehlerquellen sind Informationsverlust, Bilddeformation und Artefaktbildung hervorgerufen durch den Digitalisierungsvorgang (ALI & WHITEHEAD 2015).

Deshalb sind für die Arbeit mit historischen Bildern intensitäts- bzw. flächenbasierte Deskriptoren weitestgehend ungeeignet (STYLIANOU et al. 2015). Geometriebasierte Matching-Verfahren wie Linien-Matching basierend auf speziellen Signaturen und Semantik (WANG et al. 2009), oder Rechteck-Matching basierend auf einzelnen Liniensegmenten (MIČUŠÍK et al. 2008), können bei historischen Aufnahmen zu besseren Ergebnissen führen. Im Folgenden soll ein auf diesen Anwendungsfall entwickelter Viereck-Deskriptor sowie die damit erreichten Zuordnungsergebnisse vorgestellt werden.

## 3 Arbeitsablauf

Der Arbeitsablauf wird in den folgenden Unterkapiteln in verschiedene Schritte aufgeteilt. Es wird dabei zunächst von der Bearbeitung eines Bildpaares ausgegangen. Zuerst werden die zu untersuchenden Bilder mit unterschiedlichen Filtern bearbeitet, um die Kantendetektion zu erleichtern. Anschließend werden allgemeine Vierecke in den verschiedenen Bildern erkannt und abgespeichert. Die Schwerpunkte der Vierecke werden berechnet und diesen wird ein Deskriptor zugeordnet, der verschiedene Merkmale des Vierecks sowie der umliegenden Vierecke be-

schreibt. Abschließend erfolgt eine Zuordnung der Merkmalspunkte über die berechneten Deskriptoren. Die folgenden Schritte wurden in C++ unter Verwendung von Funktionen aus der OpenCV Bibliothek (opencv 3.2.0, opencv.org) programmiert.

### 3.1 Angleichung und Filterung der Bilder

Um die Aufnahmen zu vergleichen und die Vierecke besser detektieren zu können, werden die Bilder vorbehandelt. Ziel ist es die Kanten zu verstärken, Bildrauschen abzuschwächen und die Bilder radiometrisch auf ein Niveau zu bringen, um möglichst viele Vierecke in den historischen Aufnahmen zu detektieren. Der Deskriptor (Kapitel 3.3) wird robuster wenn eine große Anzahl an Vierecken erkannt wird. Um dies zu erreichen, wurden verschiedene Filter an unterschiedlichen Bildern empirisch getestet. Eine Automatisierung der Filterung ohne manuelle Schwellwertsetzung stand dabei im Vordergrund. Die Variante, bei der die meisten Strukturen erkannt wurden, wird im Folgenden erläutert.

Zuerst wird jeweils ein Histogrammausgleich auf beide Bilder angewandt, um den Kontrast in den meist kontrastarmen Fotografien zu verstärken. Nach diesem Schritt wird das Bild bilateral gefiltert, um Kanten zu erhalten aber Texturen (z. B. Vorhänge in Fenstern, Steinkörnung) zu vermindern (TOMASI & MANDUCHI 1998). Anschließend werden in den Aufnahmen mit Hilfe des Canny-Algorithmus' (CANNY 1986) und automatisch bestimmten Parametern mittels Otsu-Schwellwertbildung (OTSU 1979) Kanten detektiert. Diese Kanten werden abschließend mittels Closing verstärkt und geschlossen (Abbildung 2).

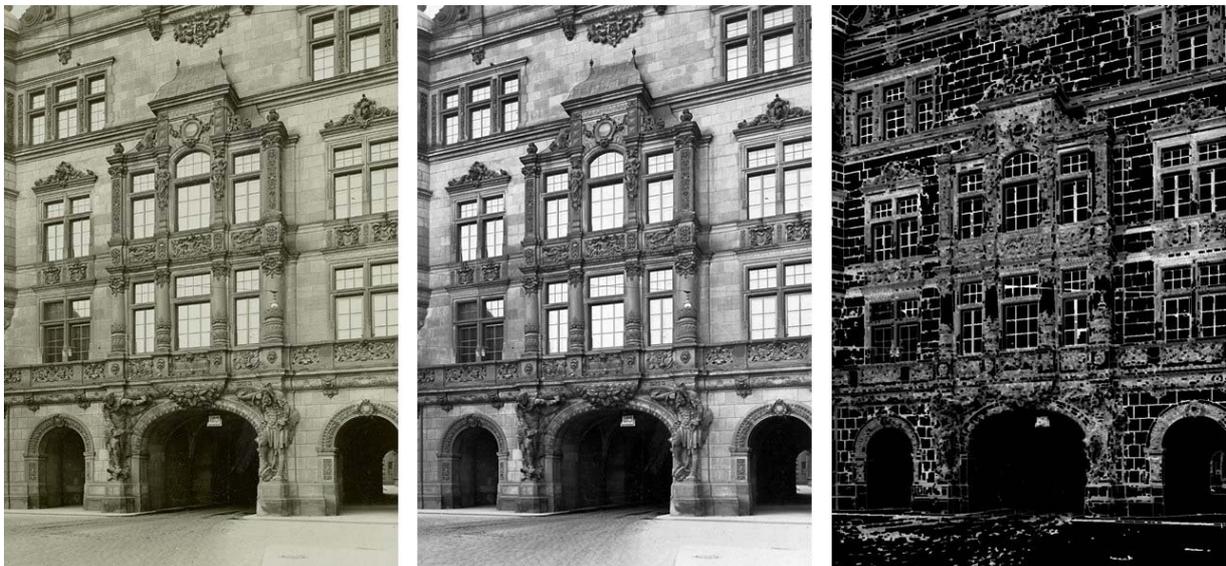


Abb. 2: Originalbild (links), mit Histogrammausgleich und bilateralem Filter bearbeitetes Bild (Mitte) und mit Canny-Algorithmus und Closing bearbeitetes Bild (rechts)

### 3.2 Detektion der Rechtecke

Als besonders robuste Geometrien für die Verknüpfung der Bilder haben sich Rechtecke im Objektraum herausgestellt (häufiges Vorkommen in Gebäudeaufnahmen, charakteristische Eigenschaften). Aber auch Liniengeometrien werden häufig für merkmalsbasierte Zuordnungen verwendet. Problematisch ist hierbei allerdings die Wahl definierter Punkte, da detektierte Geraden

erst in Liniensegmente geteilt werden müssen. Die Liniensegmente können in zwei Bildern durch fehlerhafte Segmentierung unterschiedlich fragmentiert sein. Eine Wahl von Fixpunkten zur Beschreibung der Linien (z. B. Endpunkte der Linie) ist erschwert (ZHANG & KOCH 2013). Zusätzlich ist die eindeutige Beschreibung von Liniensegmenten für die Zuordnung in einem zweiten Bild deutlich komplexer (WANG et al. 2009). Rechtecke bzw. Vierecke können hingegen bei eindeutiger Detektion auf ihren Schwerpunkt oder ihre vier Eckpunkte reduziert werden.

Bei ersten Versuchen wurden die Linien bzw. Geraden und anschließend die Vierecke mittels Hough-Transformation detektiert. Bei diesem Ansatz sind allerdings die lange Berechnungszeit und die zu definierenden Parameter problematisch. Diese lassen sich schlecht automatisch bestimmen und sind besonders für sich stark unterscheidende Bilder sehr unterschiedlich. Für die Liniendetektion wird in Zukunft noch der sogenannte Line Segment Detektor (LSD) getestet, der etwas schneller als das jetzige Verfahren ist und die Rechtecke möglicherweise robuster detektiert (VON GIOI et al. 2012).

Zum jetzigen Zeitpunkt werden die Vierecke im Canny-Bild über ein Template-Matching gefunden. Zuerst werden Konturen mittels „border following“ (SUZUKI 1985) erkannt und hierarchisch abgespeichert. Anschließend werden Konturen, deren Figur grob mit dem Template eines Rechtecks übereinstimmt, detektiert. Alle anderen vorhandenen Konturen werden verworfen. Die verbliebenen Konturen sind zu diesem Zeitpunkt jeweils als Vektor von vielen zweidimensionalen Punkten gespeichert. Eine Vereinfachung der Konturen erfolgt rekursiv durch den Douglas-Peucker-Algorithmus (DOUGLAS & PEUCKER 1973). Alle Konturen die anschließend aus genau vier Punkten bestehen, werden abgespeichert (Abbildung 3).



Abb. 3: Canny-Bild (links), detektierte Konturen nach Template-Matching (Mitte) und Vereinfachung der Konturen auf allgemeine Vierecke (rechts)

In der Abbildung kann man erkennen, dass besonders fehlerhafte Strukturen entfernt werden und die Konturen effizient als allgemeine Vierecke durch vier Eckpunkte beschrieben werden können.

In weiteren Untersuchungen soll aus den ermittelten Vierecken die projektive Verzerrung des Bildes ermittelt werden. Eine entzerrte Darstellung würde die im Objektraum rechteckigen

Strukturen auch im Bildraum unter Berücksichtigung der Verzeichnung als rechteckige Strukturen darstellen. Für den Deskriptor könnten weitere Annahmen (z. B. Seitenverhältnis der Rechtecke in zwei Bildern sollte gleich sein) getroffen werden, wodurch die Zuordnung von Punkten zuverlässiger werden sollte.

### 3.3 Beschreibung der Vierecke durch einen Deskriptor

Um die allgemeinen Vierecke eindeutig zu beschreiben, wird für jedes einzelne ein Deskriptor definiert. Andere Verfahren verwenden auch Deskriptoren, welche sich allerdings ähnlich wie bei SIFT auf Grauwertumgebungen um besondere Merkmalspunkte beschränken (MIČUŠÍK et al. 2008, WANG\* et al. 2009). Der im Folgenden vorgestellte Deskriptor hat eine ähnliche Datenstruktur wie der SIFT-Deskriptor, ist jedoch statt mit vielen Gradientenrichtungen mit Werten gefüllt, die die Vierecke sowohl geometrisch als auch semantisch-topologisch beschreiben. Die geometrischen Eigenschaften und Nachbarschaftsbeziehungen auf die der Deskriptor beruht, sind weitestgehend invariant zu projektiven Transformationen (HARTLEY & ZISSERMAN 2003).

Acht verschiedene Werte beschreiben jedes gefundene Viereck  $V_i$  (wobei  $i \in 1..n$  Anzahl gefundener Vierecke). Zunächst wird für jeden Schwerpunkt  $S_i$  des Vierecks  $V_i$  (Ursprungsviereck) die Distanz zu allen Schwerpunkten  $S_j$  benachbarter Vierecke  $V_j$  berechnet (wobei  $j \in 1..n - 1$ ). Diese Distanzen werden aufsteigend sortiert und die jeweils  $k$  ganzzahligen nächsten Nachbarn eines einzelnen Vierecks können betrachtet werden.

Die ersten vier Werte des Deskriptors beschreiben die Lage benachbarter Vierecke zum Ursprungsviereck. Es wird davon ausgegangen, dass bei erfolgreicher Viereckdetektion gleiche Vierecke in zwei Bildern ähnlich viele Nachbarn in vier verschiedenen Richtungen (Quadranten) haben sollten. Für die Berechnung wird der Winkel  $\theta_j$  zwischen der negativen y-Achse und der Strecke zwischen den Schwerpunkten  $S_i$ ,  $i \in 1..n$  und  $S_j$ ,  $j \in 1..k$  bestimmt und einer von vier verschiedenen Richtungen (A, B, C, D) zugewiesen. Die Summe der Treffer in den verschiedenen Richtungen wird anschließend mit  $k$  normiert (Abbildung 4):

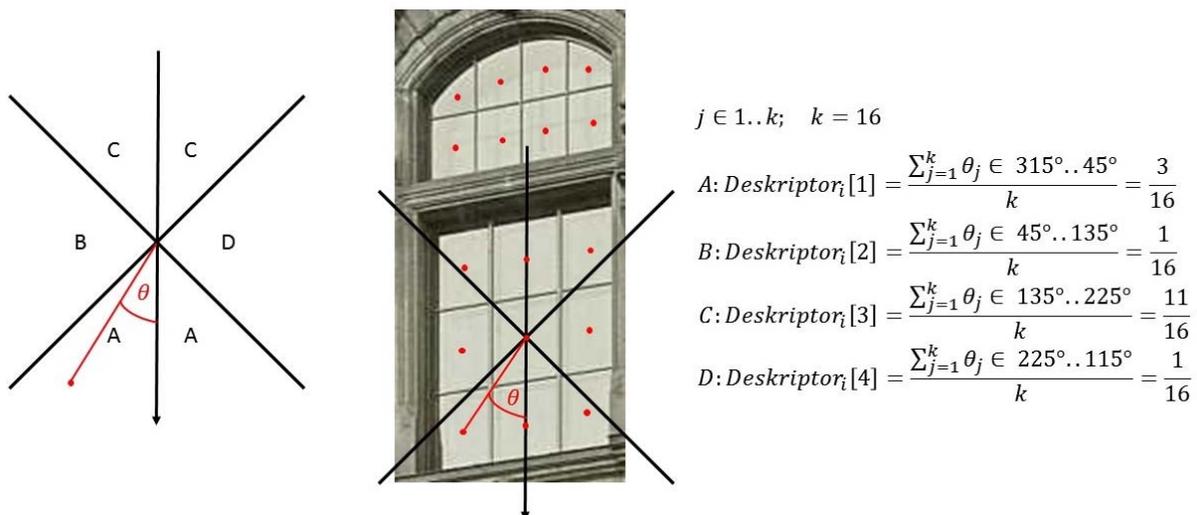


Abb. 4: Beschreibung der ersten vier Werte des Deskriptors für jedes Viereck. Benachbarte Vierecke werden je nach Winkel den verschiedenen Regionen A, B, C und D zugeordnet und anschließend mittels der Gesamtzahl der betrachteten Vierecke normiert.

Der Deskriptor ist somit nur bedingt rotationsinvariant. Da historische Aufnahmen nach der Digitalisierung in der Regel „aufrecht“ orientiert sind, fallen die umliegenden Vierecke meist in den gleichen Quadranten. Um vollständige Rotationsinvarianz zu erreichen, ist bereits geplant die Hauptrichtung der Vierecke zu berechnen und die Winkel zwischen den Hauptachsen (in Abbildung 4 etwa bei 0° und 85°) und der Strecke zwischen den Schwerpunkten zu berechnen. Die nächsten vier Werte des Deskriptors beschreiben Eigenschaften des Ursprungsvierecks, die mit den  $k$  benachbarten Vierecken verglichen werden. Dazu werden verschiedene Verhältnisse miteinander verglichen, die nach einer beliebigen projektiven Transformation von einem Bild in ein zweites Bild erhalten bleiben. Die Werte im Deskriptor sind dann bei der Detektion aller (oder fast aller) Rechtecke in der näheren Umgebung gleich (bzw. fast gleich). Die für das Brute-Force-Matching berechnete L2-Norm wird annähernd 0 und zwei homologe Vierecke in zwei Bildern werden einander zugeordnet (Abbildung 5).

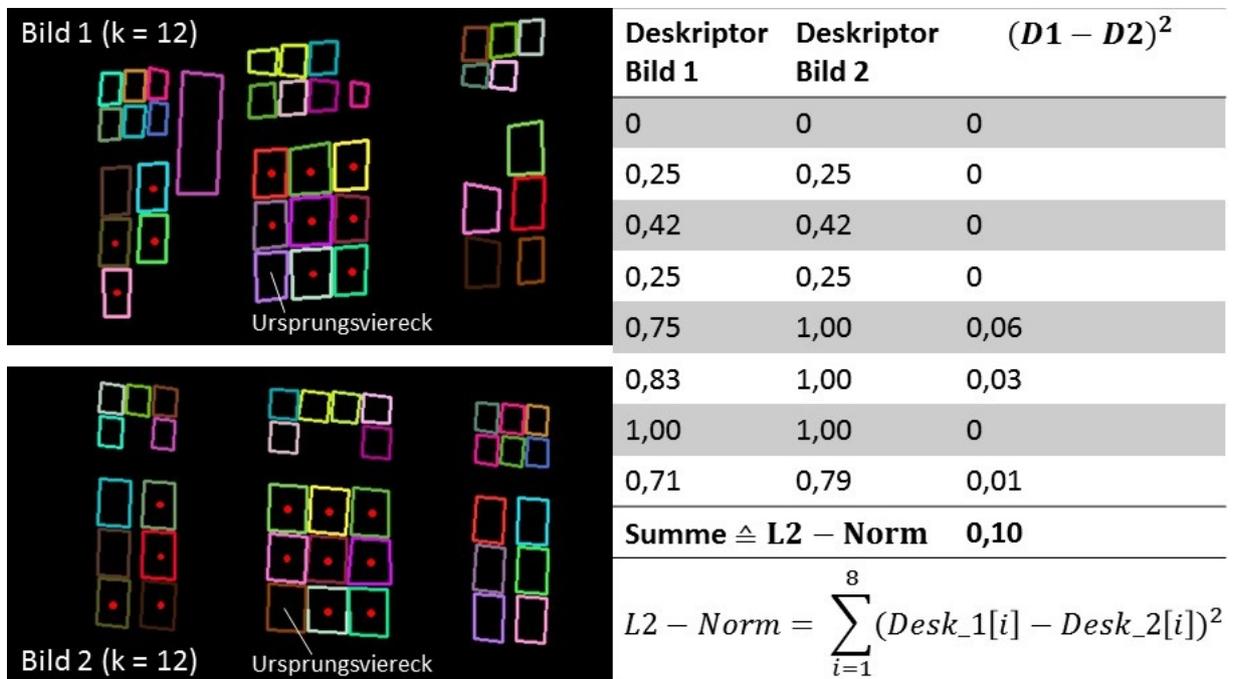


Abb. 5: Zwei homologe Vierecke in zwei verschiedenen Bildausschnitten und deren Nachbarschaft für  $k=12$  (markiert mit rotem Punkt). Für jedes Viereck werden die acht Deskriptorwerte berechnet, die in der rechten Tabelle zu sehen sind. Deskriptoren werden über Brute-Force-Matching einander zugeordnet. Dazu wird die L2-Norm zwischen verschiedenen Kombinationen zweier Deskriptorvektoren berechnet. Bei geringen Werten werden die Deskriptoren und somit die Vierecke einander zugeordnet

Für den fünften Wert des Deskriptors werden die Flächen der benachbarten Vierecke  $A_j$  mit der Fläche des Ursprungsvierecks  $A_i$  verglichen. Bei annähernd gleich großen Flächen wird der Deskriptor um 1 inkrementiert und nach Prüfung aller benachbarten Flächen wieder mit  $k$  normiert, um zur besseren Vergleichbarkeit im gleichen Wertebereich der ersten vier Werte zu bleiben.

$$\text{Flächenverhältnis}_i[j] = \begin{cases} \forall (A_i \approx A_j) = 1 \\ \forall (A_i \not\approx A_j) = 0 \end{cases}; j \in 1..k; i \in 1..n$$

$$\text{Deskriptor}_i[5] = \frac{\sum_{j=1}^k \text{Flächenverhältnis}[j]}{k}; i \in 1..n$$

Der sechste Wert beschreibt die Parallelität der Seiten benachbarter Vierecke. Dafür wird der Winkel der Seiten  $a_i, b_i$  des Ursprungsvierecks mit den Winkeln der Seiten  $a_j, b_j$  des benachbarten Vierecks verglichen. Bei Parallelität wird der Wert des Deskriptors um 1 inkrementiert und schließlich wieder mit  $k$  normiert.

$$\text{Parallelität}_i[j] = \begin{cases} \forall (a_i \parallel a_j) \cap (b_i \parallel b_j) = 1 \\ \forall (a_i \parallel a_j) \cap (b_i \parallel b_j) = 0 \end{cases}; j \in 1..k; i \in 1..n$$

$$\text{Deskriptor}_i[6] = \frac{\sum_{j=1}^k \text{Parallelität}[j]}{k}; i \in 1..n$$

Der siebte Wert des Deskriptors zeigt das Verhältnis der Seitenverhältnisse benachbarter Vierecke. Es wird das Seitenverhältnis  $\frac{a_i}{b_i}$  des Ursprungsvierecks mit dem des benachbarten Vierecks  $\frac{a_j}{b_j}$  verglichen und bei annähernd gleichen Verhältnissen wird der Deskriptorwert um 1 inkrementiert.

$$\text{Seitenverhältnis}_i[j] = \begin{cases} \forall \left( \frac{a_i}{b_i} \approx \frac{a_j}{b_j} \right) = 1 \\ \forall \left( \frac{a_i}{b_i} \not\approx \frac{a_j}{b_j} \right) = 0 \end{cases}; j \in 1..k; i \in 1..n$$

$$\text{Deskriptor}_i[7] = \frac{\sum_{j=1}^k \text{Seitenverhältnis}[j]}{k}; i \in 1..n$$

Wert 8 zeigt die Hausdorff-Distanz zwischen Ursprungsviereck und benachbartem Viereck und wird zuerst mit  $k$  und dann mit 100 normiert um den Wertebereich der anderen Zahlen des Deskriptors zu erhalten. Die Hausdorff-Distanz kann als Maß für die Ähnlichkeit zweier Punkt-mengen gesehen werden.

$$\text{Deskriptor}_i[8] = \frac{\sum_{j=1}^k \text{Hausdorff}(R_i, R_j)}{k} \cdot \frac{1}{100}; i \in 1..n$$

Der vollständige Deskriptor besteht damit aus 8 verschiedenen Werten, die sowohl maßstabs- als auch translationsinvariant sind. Eine gewisse Rotationsinvarianz ist bereits gegeben. Diese soll aber noch verbessert werden um einen affininvarianten Deskriptor zu erhalten. Die Werte des Deskriptors wurden alle auf Fließkommazahlen zwischen 0 und 1 normiert, um vergleichbare Ergebnisse bei den Abweichungen zwischen zwei Deskriptoren zu erhalten. Eine Korrelation der letzten vier Werte des Deskriptors muss noch geprüft werden.

### 3.4 Zuordnung der Merkmalspunkte

Die Schwerpunkte der Vierecke, die über den bereits erläuterten Deskriptor beschrieben werden, werden mittels Brute-Force-Matching einander zugeordnet. Das bedeutet, dass jeder Deskriptor in Bild 1 mit jedem Deskriptor in Bild 2 verglichen und das beste Ergebnis (geringster Wert der L2-Norm zwischen zwei Deskriptoren) abgespeichert wird. Die Ergebnisse werden weiter mit Hilfe eines Symmetrietests gefiltert. Das bedeutet, dass Zuordnungen von Bild 1 nach Bild 2 nur akzeptiert werden, wenn auch eine Zuordnung von Bild 2 nach Bild 1 der gleichen Punkte existiert (Abbildung 6).

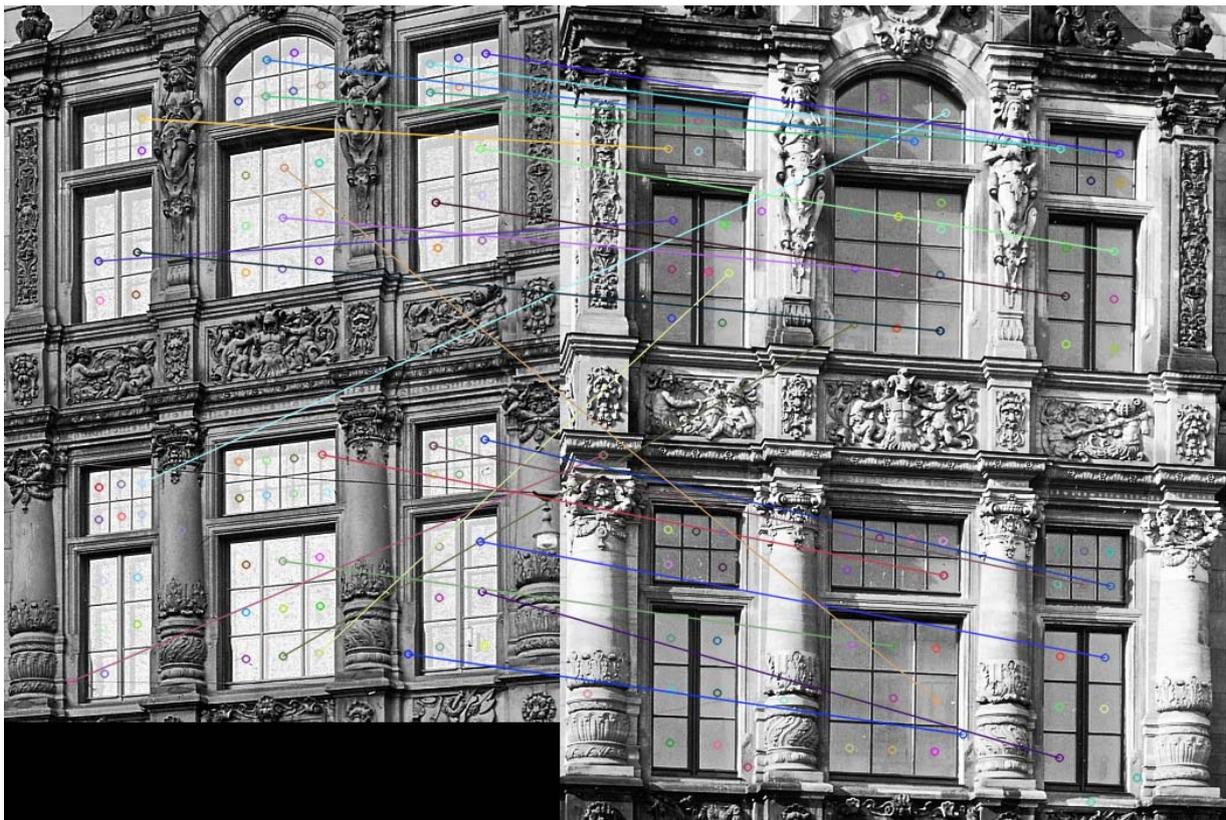


Abb. 6: Zuordnung von Punkten zwischen zwei unterschiedlichen Bildausschnitten nach Brute-Force-Matching und Symmetrietest für  $k = 20$  (Korrekte Matches: 8, Inkorrekte Matches: 14)

Die verbliebenen Zuordnungen werden für die Berechnung einer Fundamentalmatrix verwendet. Fehlerhafte Matches werden mittels RANSAC (FISCHLER & BOLLES 1981) eliminiert. Für die untersuchten Aufnahmen bleiben je nach Größe des Bildes und Anzahl der gefundenen Vierecke 7-15 korrekte Zuordnungen übrig. Diese Ergebnisse werden vor allem bei Bildern von Fassaden mit vielen Fenstern oder besonderen Steininformationen erreicht. Bei Bildern mit weniger als sieben detektierten Vierecken (Kapitel 3.2) ist eine Berechnung der Fundamentalmatrix nicht möglich. Fehlerhafte Zuordnungen die zusätzlich bestehen bleiben, sind meist Vierecke, die nur ein Viereck vom korrekten Match entfernt sind (Abbildung 7). Diese entstehen, wenn bereits in Schritt 3.2 das korrekte Viereck nicht detektiert wurde oder mehrere Vierecke in der Nachbarschaft nicht zu sehen sind. Auch bei starken perspektivischen Unterschieden oder langer Basisli-

nie sind die detektierten Vierecke bzw. deren Nachbarschaften meist zu unterschiedlich, um sie einander zuordnen zu können.



Abb. 7: Zuordnung von Punkten nach Berechnung der Fundamentalmatrix und Filterung mittels RANSAC jeweils für  $k \in 7..70$  (Korrekte Matches: 10, um ein Rechteck verschobene Matches: 2)

## 4 Ergebnisse

Wie in Abbildung 6 und 7 zu sehen ist, sind die Ergebnisse vor allem dort gut, wo viele Vierecke detektiert werden können. Die Zuordnung über den bereits beschriebenen Deskriptor ist robust und ausschließlich über geometrische Merkmale bestimmbar. Radiometrische Differenzen wie Helligkeitsunterschiede oder Kontraständerungen können so vollständig eliminiert werden (außer wenn sie sich bereits stark auf die Viereckdetektion auswirken). Problematisch ist die Zuordnung, wenn insgesamt nur wenige Vierecke detektiert werden können. Dies tritt beispielsweise bei Verdeckungen, kleinen Bilddateien, Gebäudeteilen ohne Fenster/Steine auf. In Tabelle 1 sind verschiedene Bildbeispiele und die korrekte Anzahl an Zuordnungen aufgeführt.

Tab. 1: Anzahl korrekter und fehlerhafter Zuordnungen durch den entwickelten Deskriptor bei verschiedenen historischen Bildpaaren. Zusätzlich werden Ergebnisse bei Bildpaaren des Benchmark-Datensatzes aus LI ET AL. (2016) gezeigt.

Bild			
Beschreibung	Georgentor 1964-69	Dinglinger Haus vor 1945	Zwinger 1952
Zuordnungen	Korrekt: 15 Falsch: 2	Korrekt: 7 Falsch: 2	Korrekt: 5 Falsch: 2
			
Beschreibung	Benchmarkbild building	Benchmarkbild outdoor	Benchmarkbild bikes
Zuordnungen	Korrekt: 7 Falsch: 4	Korrekt: 7 Falsch: 3	Korrekt: 3 Falsch: 1

In der Tabelle sieht man, dass die Anzahl korrekter Zuordnungen in den meisten Bildern noch zu gering ist. Die gefundenen Matches sind zwar sehr robust aber meist nicht über das gesamte Bild verteilt und somit nicht optimal für die Berechnung einer Fundamentalmatrix. Da jeweils der Schwerpunkt homologer Vierecke nicht zwingend exakt den gleichen Objektpunkt beschreibt, könnte die Reduzierung der Vierecke auf deren vier Ecken zu einer Verbesserung der Fundamentalmatrix führen. Denkbar wäre die Einführung weiterer Geometrien (Kreise, Linien), um den Deskriptor lokal zu verbessern. Über einen iterativen Ansatz könnten über die bereits gefundenen korrekten Zuordnungen weitere Matches ermittelt werden, die in das Modell der berechneten Fundamentalmatrix passen.

Des Weiteren zeigen die Ergebnisse, dass die Auswahl eines fixen Parameters  $k$  für die Anzahl betrachteter benachbarter Vierecke schwierig ist. Während für den Parameter  $k = 14$  acht korrekte Zuordnungen in den zwei Bildern des Georgentors vorhanden sind, so sind für den Parameter  $k = 9$  nur 2 korrekte Zuordnungen vorhanden (Abbildung 8).

Die korrekten Zuordnungen sind zudem meist von unterschiedlichen Vierecken an verschiedenen Positionen im Bild. Tendenziell steigt die Anzahl korrekter Zuordnungen leicht mit steigendem  $k$ . Eine Auswahl eines möglichst großen  $k$  ist sinnvoll, wenn die Mehrzahl der Vierecke auch im zweiten Bild zu sehen sind.

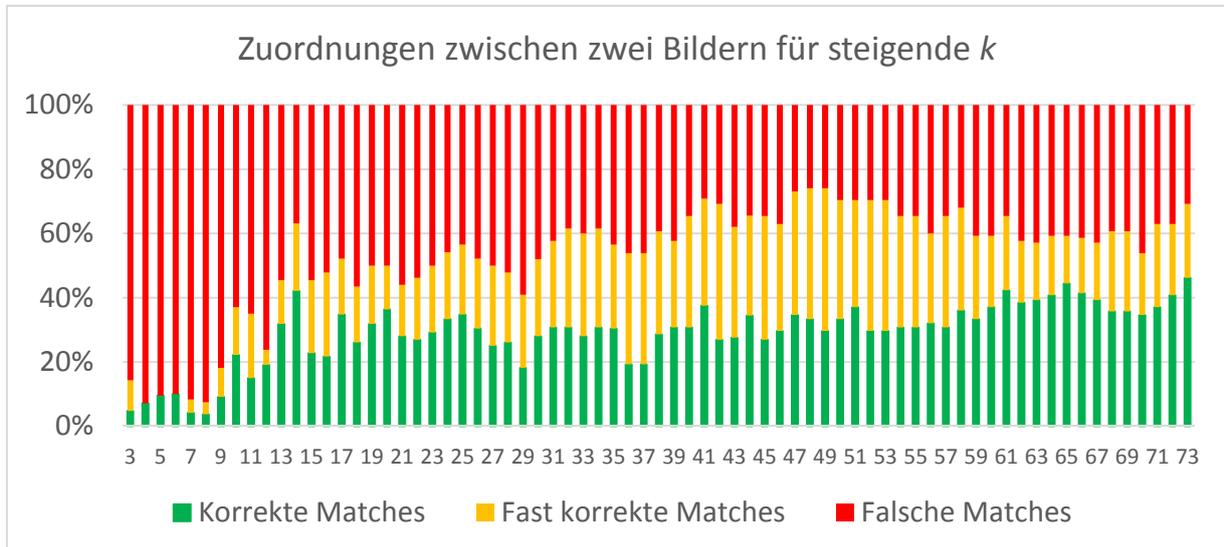


Abb. 8: Anzahl korrekter, fast korrekter (um ein Viereck verschoben) und falscher Zuordnungen von Punkten zwischen den zwei Bildausschnitten aus Abbildung 5 für  $k \in 3..73$

Stabiler ist jedoch eine Berechnung über mehrere  $k$  und anschließend die Auswahl der am meisten vorkommenden Zuordnungen. Dies hat sich als robuste aber durchaus rechenintensive Lösung herausgestellt. Die Aussagen zum Parameter  $k$  sind an dem dargestellten Bildpaar (Abbildung 7) gut zu sehen, zeigen sich aber in anderen Bildpaaren ähnlich.

## 5 Fazit & Ausblick

Das präsentierte Verfahren zeigt die Zuordnung von Merkmalspunkten über geometrische und semantisch-topologische Eigenschaften. Dabei werden in historischen Bildern zuerst Vierecke gesucht und über einen eigens entwickelten Deskriptor, der auch Nachbarschaftsverhältnisse von Vierecken darstellt, beschrieben. Die Größe der Nachbarschaft kann über einen Parameter frei gewählt werden. Die Deskriptoren werden anschließend zwischen den zwei Bildern mittels Brute-Force-Matching zugeordnet. Schließlich werden die am häufigsten vorkommenden Zuordnungen und deren Fundamentalmatrix als korrekt angenommen.

Die verbliebenen Zuordnungen sind besonders robust und können unabhängig von einer schlechten SNR und periodisch vorkommenden Objektstrukturen berechnet werden.

Da besonders eine große Anzahl an Vierecken hilfreich ist, soll die Viereckdetektion in weiteren Arbeitsschritten verbessert werden. Denkbar ist die Verwendung eines Liniensegment-Detektors, bei dem einzelne Linien zu Vierecken zusammengefügt werden. Zusätzlich ist eine rechentechnische Optimierung geplant, um die Zuordnung für eine große Anzahl an Bildpaaren zu ermöglichen. An einer rotationsinvarianten Form des Deskriptors wird bereits gearbeitet.

## 6 Danksagung

Dieser Beitrag wurde durch die Förderung des Bundesministeriums für Bildung und Forschung unter der Projektnummer 01UG1630 im Rahmen der Nachwuchsforschergruppe UrbanHistory4D ermöglicht.

## 7 Literaturverzeichnis

- ALI, H. K. & WHITEHEAD, A., 2014: Feature Matching for Aligning Historical and Modern Images. *International Journal of Computer Applications*, **21**(3), 188-201.
- BAY, H., TUYTELAARS, T. & VAN GOOL, L., 2006: Surf: Speeded up robust features. *Computer vision – ECCV 2006*, 404-417.
- CANNY, J., 1986: A computational approach to edge detection. *IEEE Transactions on pattern analysis and machine intelligence*, **6**, 679-698.
- DOUGLAS, D. H., & PEUCKER, T. K., 1973: Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. *Cartographica: The International Journal for Geographic Information and Geovisualization*, **10**(2), 112-122.
- FISCHLER, M. A. & BOLLES, R. C. 1981: Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, **24**(6), 381-395.
- VON GIOI, R. G., JAKUBOWICZ, J., MOREL, J. M. & RANDALL, G., 2012: LSD: a line segment detector. *Image Processing On Line*, **2**, 35-55.
- HARTLEY, R. & ZISSERMAN, A., 2003: *Multiple view geometry in computer vision*. Cambridge university press.
- LI, K., YAO, J., LU, M., HENG, Y., WU, T. & LI, Y., 2016: Line segment matching: A benchmark. *IEEE Winter Conference on Applications of Computer Vision (WACV)*, 1-9.
- LOWE, D. G., 2004: Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, **60**(2), 91-110.
- MIČUŠÍK, B., WILDENAUER, H. & KOSECKA, J., 2008: Detection and matching of rectilinear structures. *IEEE Conference on Computer Vision and Pattern Recognition*, 1-7.
- MUJA, M. & LOWE, D.G., 2012: Fast approximate nearest neighbors with automatic algorithm configuration. *International Conference on Computer Vision Theory and Applications (VISAPP'09)*, 331-340.
- OTSU, N., 1979: A Threshold Selection Method from Gray-Level Histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, **9**(1), 62-66.
- RUBLEE, E., RABAUDE, V., KONOLIGE, K. & BRADSKI, G., 2011: ORB: An efficient alternative to SIFT or SURF. *IEEE international conference on Computer Vision (ICCV)*, 2564-2571.
- SCHINDLER, G. & DELLAERT, F., 2012: 4D cities: analyzing, visualizing, and interacting with historical urban photo collections. *Journal of Multimedia*, **7**(2), 124-131.
- STYLIANOU, A., ABRAMS, A. & PLESS, R., 2015: Characterizing Feature Matching Performance over Long Time Periods. *IEEE Winter Conference on Applications of Computer Vision*, 892-898.

- SUZUKI, S. 1985: Topological structural analysis of digitized binary images by border following. *Computer vision, graphics, and image processing*, **30**(1), 32-46.
- WANG, L., NEUMANN, U., & YOU, S., 2009: Wide-baseline image matching using line signatures. *IEEE 12th International Conference on Computer Vision*, 1311-1318.
- WANG\*, Z., WU, F., & HU, Z., 2009: MSLD: A robust descriptor for line matching. *Pattern Recognition*, **42**(5), 941-953.
- WOLFE, R., 2015: Modern to historical image feature matching. <http://robbiewolfe.ca/programming/honoursproject/report.pdf>.
- ZHANG, L., & KOCH, R., 2013: An efficient and robust line segment matching approach based on LBD descriptor and pairwise geometric consistency. *Journal of Visual Communication and Image Representation*, **24**(7), 794-805.