# 3D-Deskriptoren für Aufgaben der Rekonstruktion und Objekterkennung

*Christian Lengert* — 17. Juni 2017

Sommersemester 2017 · Computervision Group · FSU Jena

**Zusammenfassung**

Im Rahmen des Seminars Rechnersehen der Friedrich-Schiller-Universität entsteht dieser Text als Ausarbeitung des Themas *3D-Deskriptoren für Aufgaben der Rekonstruktion und Objekterkennung* und soll zusammen mit dem Vortrag einen Einblick in den Aufbau und die Einsatzbereiche von 3D-Deskriptoren geben. Abschnitt 1 beleuchtet das Konzept hinter dem Begriff 3D-Desriptor, erklärt das für den Themenbereich notwendige Vokabular und liefert Einblick in die Einsatzbereiche. Dann folgt in 2 ein Überblick über den Entwicklungsprozess und die Bewertung von Deskriptoren. In 3 wird auf die in Tombari et. al. [1] entwickelte Methode (SHOT) eingegangen. Hier sollen die Unterschiede zu den vorherigen Methoden herausgearbeitet werden. Dann erfolgt in 4 die Evaluation der vorgestellten Methoden, und abschließend ein Fazit und das Literaturverzeichnis.

# 1 Einführung

**3D-Deskriptor** Ein 3D-Deskriptor ist eine Methode um Punkte $p$, welche zu einer Fläche im $\mathbb{R}^3$ gehören, anhand der Punkte in Ihrer Umgebung zu beschreiben. Diese Beschreibung soll dann der Identifikation dienen, sodass man die Beschreibung eines Punktes mit einer anderen Beschreibung vergleichen kann. So ist es möglich, aus beschriebenen Punkten zusammengesetzte Flächen miteinander zu vergleichen und den Grad der Ähnlichkeit von Objekten festzustellen. Das Problem des Oberflächenvergleichs ist allgemein als **Surface-Matching** bezeichnet.

Die Eingabe für einen Deskriptor ist ein Teil einer Punktwolke, zum einen der Punkt, welcher später als Merkmal dienen soll, sowie die Punkte in seinem Umfeld, ab jetzt als **Nachbarschaft** bezeichnet. Um mehrere Features aus einem Bild zu extrahieren werden die Deskriptoren für viele Punkte im Bild berechnet. Diese können zum Beispiel zufällig gewählt sein [1] oder durch die Berechnung und Auswertung von Gradientenbildern an interessante Punkte im Bild, wie zum Beispiel Ecken, gelegt werden [2]. Der Auswahlprozess wird als **Feature Selection** bezeichnet.

## 1.1 Einsatzbereiche

Die Einsatzbereiche sind sehr vielseitig, Beispiele sind die Lokalisation von Robotern anhand der Auswertung von Daten aus den an Ihnen angebrachten optischen Sensoren [3] , sowie die Konstruktion von 3D-Modellen aus zweidimensionalen Bildern [4]. Desweiteren verbessert der Einsatz von dreidimensionalen Abbildern und Methoden im Bereich der Biometrie die Erkennungsrate signifikant, da Sie robuster gegenüber Variation in Objektposition und Beleuchtung sind [5]. Bereits vor dem Erkennen von Objekten1.1.1 wurden

Deskriptoren für die Detektion von Bewegungen durch das Erkennen von Punktverschiebungen im sogenannten Motion-Tracking eingesetzt.

### 1.1.1 Objekterkennung

Die Verwendung zur Erkennung von Objekten ist einer der Haupteinsatzbereiche von 3D-Deskriptoren. Aus der oben beschriebenen Möglichkeit, Flächen durch mehrere Punktdeskriptoren zu beschreiben, lässt sich eine Methode der Objekterkennung konstruieren.
Angenommen, es wird über ein Model eines Objektes verfügt, welches zum Beispiel mit einem Verfahren der 3D-Bildgewinnug, oder durch Modellierung gewonnen wurde. Dann kann ein das Objekt repräsentierende Struktur erzeugt werden, indem an hinreichend vielen Punkten ein Deskriptor berechnet wird. Diese lokalen Punktdeskriptoren werden dann zusammengenommen als Repräsentation für das Objekt verwendet und zum Beispiel in einer Datenbank hinterlegt. Liegt nun eine Szene vor, in der sich das Objekt, in anderer räumlicher Lage, vielleicht teilweise verdeckt von anderen Szeneobjekten, befindet, dann besteht die Aufgabe darin, das Objekt zu suchen und zu lokalisieren. Dieses Problem wird im Verlauf des Textes noch ausgiebig behandelt.

### 1.1.2 Rekonstruktion

Punktdeskriptoren können auch zur Rekonstruktion von Oberflächen verwendet werden. Durch optische Sensoren, wie zum Beispiel LIDAR, Stereokamerasysteme, oder photogrammetrische Methoden, erzeugte Punktwolken beschreiben Oberflächen oftmals nicht ausreichend gut für bestimmte Einsatzbereiche. Im Zusammenhang mit der Vollautomatisierung von Baumaschinen [6], der automatischen Landung von Raumsonden oder im Bereich der Archeologie [7] sind, werden sehr detailierte Informationen über die Oberflächen der jeweiligen Umgebungen benötigt. Diese müssen durch das Schätzen von Punktposition durch Betrachtung der bekannten Punkte interpoliert werden.

## 1.2 Struktur

Nach [1] kann man die Art der Betrachtung der Nachbarschaft in zwei Gruppen einteilen. Der Unterschied besteht in der Form des Nachbarschaftsbereichs und in der Berechnung, die auf ihn angewand wird.

### 1.2.1 Signatur

Eine Signatur zeichnet sich durch die Definition eines Referenzrahmens aus, welcher die die Nachbarschaft definiert. Auf jeden der benachbarten Punkte wird dann unter Berücksichtigung seiner relativen Position zum beschriebenen Punkt, zum Beispiel durch ein Distanzmaß, eine Berechnung ausgeführt. Die Kombination der Ergebnisse liefert dann den Deskriptor. Beispiele folgen später mit den Methoden der Point-Signatures und des Expontial Mappings.

### 1.2.2 Histogramm

Wird das Histogramm als Beschreibungsmethode gewählt, so wird eine diskrete Anzahl an Merkmalen benötigt, welche dann durch Auszählung der Ausprägungen als diskrete Wahrscheinlichkeitsdichtefunktion verwendet wird.

Bei der Reduktion durch Zählung können unter Umständen Informationen über die lokale Position der Punkte in der Nachbarschaft verloren gehen. Bei der Berechnung von Spin Images [4] zum Beispiel wird auf den Radius und die Höhe im Zylinder relativ zur

Punktnormalen reduziert, es ist also nur bekannt, auf welcher Rotationsbahn sich ein benachbarter Punkt befunden hat. Dieser Informarionsverlust reduziert die Beschreibungskraft des lokalen Umfelds, kann jedoch zu mehr Robustheit gegenüber dem Einfluss von Rauschen führen [1]

Histogrammbasierte Deskription erfordert die Definition einer Referenzachse, wie z.B. die eben genannte Punktnormale.

# 2 Bisherige Ansätze

Nun soll eine Auswahl der in [1] verglichenen und erprobten Ansätze kurz vorgestellt werden.

**Spin Images (SI) [4]** Die Methode der Spin-Images wurde 1999 von Andrew Johnson und Martial Herbert vorgestellt. Sie wird in die Klasse der histogrammbasierten Methoden eingeordnet, ihre Nachbarschaftsbeschreibung basiert auf einer Referenzachse. Johnson und Herbert berechnen für ausgewählte Punkte die Oberflächennormale, indem Sie eine Ebene an die Nachbarschaft des Punktes einpassen. Die Nachbarschaft definieren sie als alle Punkte, welche durch ein Oberflächennetz verbunden direkte Nachbarn des als Feature gewählten Punktes sind. Nun spannen sie einen Zylinder mit Radius $r$ und Höhe $h$ auf, dessen Rotationsachse die eben berechnete Normale des Punktes ist. Nun wird ein Fläche $r * h$ um die Achse rotiert und es werden Zähler für jeden Punkt mit den Koordinaten $(r, h)$ inkrementiert. Durch Zählung der Punkte an spezfischen Koordinaten relativ zur Rotationachse ergibt sich ein Histogramm, das Spin-Image. Darüber hinaus wird die Anzahl der betrachteten Punkte dadurch reduziert, dass nur Punkte betrachtet werden, deren Normalenwinkel bis zu einem Maximalmaß von Winkel des betrachteten Punktes abweichen.

**Point Signatures (PS) [8]** Die Beschreibungsform der Point-Signatures ist ein Vertreter der Signatur-basierten Ansätze für die Punktbeschreibung. Chua und Jarvis legen eine Sphäre um den zu beschreibenden Punkt, sodass dieser das Zentrum darstellt, und betrachten dann den Ausschnitt der Fläche des Objektes als Nachbarschaft, welcher durch den Schnitt mit der Sphäre begrenzt wird. Eine Ebene wird in die Region eingepasst, wessen Normale für gegen Null strebenden Radius der Sphäre die tangentiale Ebene im Punkt $p$ realisiert. Wenn diese Ebene dann auf den Punkt $p$ angehoben oder abgelassen wird, wird das durch den Sphärenschnitt erzeugte eindimensionale Höhenprofil auf die Ebene projeziert. Dadurch wird eine vorzeichenbehaftete Repräsentation des Umfeldes erzeugt. Für jeden Winkel der Sphäre ist die Distanz zwischen Profil und Ebene negativ wenn die Ebene über dem Profil liegt und positv, wenn sie sich darunter befindet.

Ein Vergleich zweier so erzeugter Signaturen $s_{1,2}$ erfolgt durch den Vergleich der Betragsdiffererenz mit einem Schwellenwert $\epsilon$ wie folgt:

$$|s_1(\Theta_i) - s_1(\Theta_i) > \epsilon|$$

wobei die Vektoren $\Theta_i$ die Profil-Ebenen-Distanzen an einer diskreten Anzahl an Winkeln sind.

**Exponential Mapping [EM** [9]] Der Deskriptor in der Veröffentlichung basiert auf der Idee, die Größe der zu betrachteten Nachbarschaft vom Umfeld des Betrachteten Punktes abhäning zu machen. Die Wahl der relevanten Punkte wird durch eine Eckendetektion auf der zu betrachtenden dreidimensionalen Struktur getroffen. Der

Ausgangspunkt dazu ist eine Normalmap, eine Sicht der Struktur in der ein Punkt durch seine Normale repräsentiert wird. Durch die Filterung mit Gauss-Kernen mit steigender Standardabweichung wird ein sogenanter Scale-Space erzeugt, welcher inherent die Skalenvarianz, also Schwankungen in der Dichte der Punktwolke verkörpert. Die Berechnung der Eckpunkte erfolgt dann durch die Berechnung der Gram-Matrix, der Matrix der partiellen ersten Ableitungen auf allen Skalen (Standardabweichungen).
Der eigentliche Deskriptor wird nun auf dem Scale-Space definiert, indem für alle Skalen jeweils ein skalenabhängiger Deskriptor berechnet wird und die Ergebnisse zu einem unabhängigen zusammengefasst werden. Der abhängige Deskriptor basiert auf dem geodätischen Distanzbegriff, welcher durch den kürzesten Weg zwischen zwei Punkte definiert ist, wenn der Weg der Oberfläche folgt, zu der die zwei Punkte gehören. Die geodätische Distanz zwischen zwei Punkten auf der Erdkugel wäre der Umfang des Kreisausschnittes den ein Reisender auf der Kugel zurücklegt. Da die Anzahl an Punkten im Modell endlich ist wird die geodätische Distanz zwischen Punkten $a, b$ durch die Summe der euklischen Distanzen zwischen den Knoten des kürzesten Weges zwischen $a$ und $b$ definiert.

Zur skalenabhängingen Deskription wird jeder Punkt $n$ in der Nachbarschaft des detektierten Eckpunktes $e$ auf ein Wertepaar $(d, \Theta)$ abgebildet, wobei $d$ die geodätische Distanz zwischen $e$ und $d$ ist, und $\Theta$ der polare Winkel einer Tangente an der Geodäte ist. Diese Tangente ist relativ zu einer für alle Punkte fixierten Basis.

Der skalenunabhängige Deskriptor wir dann aus der Kombination aus von abhängigen Deskriptoren einer Skalenreihe erzeugt.
Die Berechnung der Ähnlichkeit zwischen zwei Deskriptoren wird dann über die Kreuzkorrelation der entstandenen zweidimensionalen Repräsentaionen gewonnen, wobei nur die Punkte in der Schnittmenge der Nachbarschaften in die Berechnung einbezogen werden.

# 3 Fortschritt

Nun sollen der von Tombari et al. konstruierte Deskriptor vorgestellt werden.

## 3.1 SHOT

Der **SHOT** Deskriptor verwendet als Basis der Umgebungsbeschreibung nicht wie die vorher betrachteten Methoden, entweder ein Histogramm oder eine Signatur, sondern eine Kombination aus beidem.
Die durch die sukzessive Aufteilung der Nachbarschaft erreichte Wahrung der räumlichen Lageinformationen relativ zum betrachteten Punkt ähnelt der Bildung einer Signatur. Deshalb wird in der Veröffentlichung von einer Deskription durch Signaturen von Histogrammen gesprochen. Der Deskriptor soll so die Stärken der beiden Ansätze in sich vereinen.

### 3.1.1 Der Referenzrahmen

Zuerst wird ein adquater Referenzrahmen definiert. Dafür wird zunächst wie in anderen Methoden auch die Normale des Punktes geschätzt. Dies geschieht als Wei-

terentwicklung eines Verfahrens aus [10, 11], welche eine Least-Squares-Schätzung durch die Wahl des zum kleinsten Eigenwert gehörenden Eigenvektors der Kovarianzmatrix der $k$ nächsten Nachbarn verwendet. Um die Performanz in mehrere Objekte enthaltenden Punkträumen zu erhöhen, werden die Punkte aus der Nachbarschaft absteigend mit der Distanz gewichtet. Außerdem werden die alle Punkte in einer Sphäre um den zu betrachtenden Punkt ebenfalls in die Berechnung einbezogen. Dies ermöglicht die Belegeung der durch die drei größten Eigenvektoren repräsentierten Basis mit Vorzeichen.

### 3.1.2 Der Deskriptor

Der Deskriptor ist inspiriert durch eine im Bereich der zweidimensionalen Deskription sehr erfogreiche Methode, genannt *Scale Invariant Feature Transform (SIFT).*

**SIFT [12]** In dieser Methode wird, ähnlich wie in 2, durch kaskadierte Anwendungen von Gauß-Filtern ein Scale-Space erzeugt. Die gefilterten Bilder werden in sogenannte Oktaven eingeteilt. Nun wird zwischen je zwei Bildern einer Oktave die Differenz der gefilterten Bilder berechnet. Aus diesem Scale-Space sollen markante Punkte im Bild gewonnen werden, welche dann durch den Deskriptor beschrieben werden. Die Relevanz eines Punkte wird bewertet, indem lokale Minima und Maxima über die benachbarten Skalen und die Skala, in dem sich der Punkt befindet, gebildet werden. Nur wenn sein Wert größer oder kleiner als der Wert von acht benachbarten Punkten im selben Bild ist, und wenn sein Wert in den benachbarten Skalen sich zu neun benachbarten Punkten genau so verhält, wird er ausgewählt.

Die so gewählten Punkte garantieren eine skalierungsinvariante Beschreibung, da sie durch mehrere Skalen hinweg relevanten Einfluss auf die Extremwertbildung haben. Im Anschluss daran wird für die gewählten Punkte im Umfeld die Gradientenrichtung bestimmt und es werden mehrere Histogramme über eine diskrete Anzahl an möglichen Richtungen gebildet. Dafür wird das Umfeld in eine Anzahl an Quadraten eingeteilt (z.b. 4x4). Für jeden dieser Bereiche wird ein eigenes Histogramm berechnet und die Kombination in Form eines einzelnen Vektors dient als Deskriptor.
Abschließend wird der fertige Deskriptor normiert um Invarianz gegenüber Beleuchtung zu garantieren.

Der Vergleich mit anderen Deskriptor-Vektoren kann dann zum Beispiel über den Nächster-Nachbar-Klassifikator oder über einen Clustering-Ansatz erfolgen.

**Von 2D nach 3D** Für den *SHOT*-Deskriptor wird die Histogrammbildung über Teilbereiche der Nachbarschaft in den dreidimensionalen Raum übertragen. Dazu muss natürlich das Gitter angepasst werden, bezüglich dem die Histgramme berechnet werden. Dazu wird ein isotropisches sphärisches Koordinatensystem verwendet, im dem jeder Punkt durch Betrag des Ortsvektors und zwei Winkel, den Azimuth-Winkel in der $XY$-Ebene und der polaren Winkel in der $YZ$-Ebene, beschrieben wird. Durch eine Diskretisierung dieser drei Achsen wird eine sphärische Nachbarschaft in Bereiche eingteilt, über die dann jeweils im nächsten Schritt ein Histogramm gebildet werden kann. Tombari et al. schlagen eine Skalierung vor in der die Azimuth-Achse in acht Bereicht geteilt wird und die beiden verbleibenden Maße jeweils in zwei Stufen geteilt werden, wodurch sich dann 32 Bereiche ergeben.
Über diese Bereiche sollen nun Histogramm erzeugt werden, welche am Ende zum fertigen Deskriptor vereint werden. Die Einteilung erfolgt dabei über den Kosinus des Winkels zwischen Normale im Feature Punkte und der Normalen der Punkte

in der Nachbarschaft. Der Kosinus wird gewählt, da so die Auflösung des Histogramms variabel von der Größe des Winkels abhängig wird. Im Bereich der kleinen Abweichungen ist die Auflösung hoch, mit größer werdendem Winkel nimmt sie ab. Nebeneinander liegende Punkte mit großen Winkelunterschieden sind eher selten, die geringe Auflösung in diesem Bereich stärkt die Robustheit gegenüber durch Rauschen verursachten Ausreißern.

# 4 Evaluation

Im Experiment wurden die in 2 vorgestellten Deskriptoren miteinander verglichen. Dazu wurden drei unterschiedliche Szenarien konstruiert, in denen sich die Deskriptoren beweisen sollen. In den ersten beiden Experimenten wurden für jedes Modell 1000 Feature Points ausgewählt, im letzten dann 3000 Punkte pro Szene.
Die Parameter der Modelle, der Radius des die Nachbarschaft beschreibenden Körpers und die Länge des Deskriptors, wurden anhand einer Testszene eingestellt.
Die Performanz wurde durch die Precision- und Recallmaße gemessen, wobei die Precision die Anzahl der richtig klassifizierten $tp$ mit den fälschlicherweise als richtig klassifizierten Punkten $fp$ in Verbindung bringt $\frac{tp}{tp+fp}$. Der Recall gibt Aufschluss darüber, wie viele der als richtig klassifizierten Modelle im Verhältnis den fälschlicherweise als negativ klassifiziert Modellen gegenüberstehen: $\frac{tp}{tp+fn}$.

**Experiment 1** Zuerst wurden sechs Modelle aus dem *Stanford 3D Scanning Repository* gewählt. Mit diesen wurden dann 45 Szenen erzeugt, indem die Modelle zufällig ausgewählt, verschoben und rotiert wurden. Dann wurden die Szenen künstlich mit Gaussrauschen in 3 unterschiedlichen Intensitäten versetzt. Hier soll die Robustheit gegenüber Rauschen getestet werden.
Es zeigt sich, dass die Histogrammbasierte Methode der Spin-Images das schlechteste Ergebnis erzielt. Die Methoden EM und PS erzielen aufsteigend bessere Resultate, SHOT verliert mit steigendem Rauschanteil zwar auch an Genauigkeit, jedoch erzielt es immernoch bessere Ergebnisse, so, dass mit hoher Precission auch ein hoher Recall einhergeht.

**Experiment 2** Hier wurden wieder die Szenen aus dem ersten Experiment verwendet. Dieses Mal wurde aber die Punktedichte in den Szenen drastisch reduziert($\frac{1}{8}$). Hier wird klar, dass die Dichtereduktion einen sehr starken Einfluss auf die Güte der Deskriptoren hat.
Alle Testkandiaten verlieren stark an Güte, SHOT, PI und SI liegen dabei ungefähr gleich auf. Exponential-Mapping erkennt durchweg sehr wenige der Muster und hat somit durchweg sehr niedrigen Recall.

**Experiment 3** Im letztem Versuch wurden im Labor erzeugte Szenen verwendet, die verwendeten acht Modelle sind durch [13] erzeugt worden, eine Methode um dreidimensionale Objekte aus zwei Videostreams einer Stereokamera zu erzeugen. Jeweils zwei der acht Objekte wurden in 15 Szenen verteilt, welche zusätzlich andere zufällig verteilte Objekte enthielt. Die zwei Modelle sollten trotz Überlappung und Clutter zuverlässig erkannt werden. Dieser Versuch kommt dem tatsächlichen Einsatz vom Deskriptoren am nächsten.
Dabei liegt wieder SHOT vor den anderen Methoden, es folgen dann SI, PS und zuletzt, wieder EM mit vergleichsweise sehr niedrigem Recall in dieser Reihenfolge.

**Geschwindigkeit** Bei der Analyse der Zeit pro Feature Punkt liegen Spin-Images und SHOT ungefähr gleich auf, wobei SI immer etwas schneller ist. Größenordnungen

langsamer ist EM und noch wesentlich später folgt PS.

# 5 Fazit

Der in der Veröffentlichung proklamierte SHOT-Deskriptor wurde im Nachinein vielfach zitiert und verwendet [14, 15, 16], jedoch wurde auch Skepsis bezüglich der Leistung von konstruierten Deskriptoren im Vergleich mit dem Einsatz von tiefen künstlichen neuronalen Netzen geäußert. Ich persönlich denke, dass die Nachvollziehbarkeit des Berechnungsprozesses und somit der Gewinn von Wissen über die Modellierung von Problembereichen ein grundlegender Vorteil von konstruktiven Methoden gegenüber dem konnektionistischen Ansatz ist. Jedoch überzeugt in der Praxis die Leistung und nicht der Wissensgewinn.

Die Arbeit mit dreidimensionalen Deskriptoren war sehr aufschlussreich und hat viele Einblicke in sehr abstrakte geometrische Sachverhalte geboten. Die Anzahl der Veröffentlichungen zum Thema 3D-Deskription ist hoch und es existieren eine Vielzahl an teilweise grundsätzlich unterschiedlichen Methoden. Es ist interessant zu sehen, wie Methoden, die sich bewährt haben, immer wieder aufgegriffen und verbessert werden.

**Anmerkung** Für allgemeine Inforamtionen im Bereich der 3D-Bildverarbeitung wurden zusätzlich folgende Bücher verwendet [17, 18, 19]

# Literatur


[1] Luigi Di Stefano Federico Tombari, Samuele Salti. Unique signatures of histograms for local surface description. 2010.

[2] Mike Stephens Chris Harris. A combined corner and edge detector. 1988.

[3] Se et al. Mobile robot localization and mapping with uncertainty using scale-invariant visual landmarks. 2002.

[4] Martial Hebert Andrew E. Johnson. Using spin images for efficient object recognition in cluttered 3d scenes. 1999.

[5] Licesio J. Rodríguez-Aragón Cristina Conde and Enrique Cabello. Automatic 3d face feature points extraction with spin images.

[6] Lee Kim Sung, Kwon. Fast and robust 3d terrain surface reconstruction of construction site using stereo camera. 2016.

[7] Bauer Paar. Cavity surface measuring system using stereo reconstruction.

[8] RAY JARVIS CHIN SENG CHUA. Point signatures: A new representation for 3d object recognition.

[9] Ko Nishino John Novatnack. Scale-dependent/invariant local 3d shape descriptors for fully automatic registration of multiple sets of range images.

[10] Duchamp McDonald Stuetzle Hoppe, DeRose. Surface reconstruction from unorganized points. 1992.

[11] Guibas Mitra, Nguyen. Estimating surface normals in noisy point cloud data. 2004.

[12] David G. Lowe. Distinctive image features from scale-invariant keypoints. 2004.

[13] Seitz Zhang, Curless. Spacetime stereo: Shape recovery for dynamic scenes. 2003.

[14] Branson Maiolino, Woolley. Flexible robot sealant dispensing cell using rgb-d sensor and off-line programming. 2017.

[15] Zhou Pi Lin, Min. A human-robot-environment interactive reasoning mechanism for object sorting robot. 2017.

[16] Oprea et al.Classification of Aerial Photogrammetric 3D Point Cloud Garcia-Garcia, Orts. Multi-sensor 3d object dataset for object recognition with full pose estimation. 2017.

[17] Olivier Faugeras. *Three-dimensional computer vision : a geometric viewpoint*. MIT Press, Cambridge, Mass, 1993.

[18] Richard Hartley. *Multiple view geometry in computer vision*. Cambridge University Press, Cambridge, UK New York, 2004.

[19] Emanuele Trucco. *Introductory techniques for 3-D computer vision*. Prentice Hall, Upper Saddle River, NJ, 1998.