Analyse durch Synthese Modellierung von 3D Objekten in Stereobildfolgen

Ralf Tönjes, Oliver Grau, Reinhard Koch

Universität Hannover,
Institut für Theoretische Nachrichtentechnik und Informationsverarbeitung,
Abteilung Automatische Bildinterpretation, Prof. Dr.−Ing. Liedtke,
Appelstr. 9A, D−30167 Hannover,
E−mail: toenjes@tnt.uni−hannover.de (grau@,koch@)

Der Beitrag beschreibt eine Modellierung von 3D Objekten durch Analyse von Stereobildfolgen. Basierend auf der stereoskopischen Korrespondenz−analyse wird ein 3D Modell bestehend aus Oberflächengeometrie und Oberflächentextur generiert. Computergraphik wird zum einen während der Bildanalyse zur Verifikation des Modellierungszustandes und zum anderen während der Visualisierung zur realitätsgetreuen Synthetisierung der modellierten Szene verwendet. Zur Verbesserung der datengestützten Modellierung wird eine wissensbasierte Bildinterpretation vorgeschlagen.

1 Einleitung

Die Leistungsfähigkeit moderner Computerhardware eröffnet immer neue Anwendungsbereiche für die Computergraphik und −visualisierung. Die Generierung von künstlichen Welten hat bereits breiten Einzug in der Werbung und Filmindustrie und in Simulationssystemen unterschiedlichster Art (Flug− und Fahrsimulatoren) gehalten. Neue Anwendungsbereiche, liegen beispielsweise im Architekturbereich: Ein Architekt könnte ein neues Bauprojekt als CAD−Modell in eine Baulücke einfügen und die dadurch entstandene Veränderung als Simulation visualisieren. Bei derartigen Anwendungen entsteht ein großer Bedarf an Modellen, die zum derzeitigen Stand der Technik meist manuell erzeugt werden, eine zeit− und kostenintensive Vorgehensweise. Daher wird eine möglichst automatische Gewinnung der Modelle durch Analyse von Bildfolgen angestrebt.

Mittels CAD−Techniken erstellte Modelle wirken häufig stilisiert und unnatürlich. Die Modellgeneration durch Bildanalyse hat dort Vorteile, wo es auf die realitätsgetreue Darstellung bereits existierender Objekte ankommt. So sind mittels Bildanalyse erfolgreich Landschaften [1] und Häuser [2] für Simulatoren sowie Personen für die Bildtelefonie [3] modelliert worden.

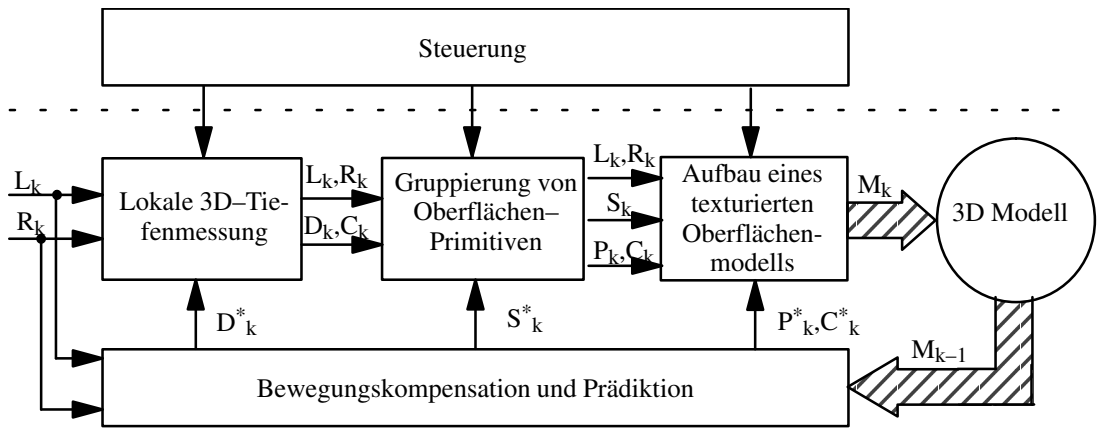
Der Artikel beschreibt ein automatisches Verfahren zur 3D−Modellierung, das auf

der Analyse von Stereobildfolgen basiert. Computergraphik wird zur Verifikation des Modellierungszustandes während der Bildanalyse und zur realitätsgetreuen Visualisierung der modellierten Szene verwendet.

Im folgenden Abschnitt wird ein kurzer Überblick über die Arbeitsweise des verwendeten datengetriebenen Modellierers gegeben. Die Bedeutung der Bildsynthese für die Analyse der Modelle wird im 3. Abschnitt herausgestellt. Abschnitt 4 beschreibt einen Ansatz zur Modellverbesserung durch Einbringung von Wissen. Der Artikel schließt mit einer Darstellung der Modellierungsergebnisse.

2 Modellierung durch Bildanalyse

In diesem Kapitel soll kurz die Struktur und Arbeitsweise des datengetriebenen Modellierers (Bild 2.1) skizziert werden, der als Sensor über ein bewegtes binokulares Kamerasystem (Sensor L_k, R_k) verfügt. Das Kernstück der Modellierung ist die Bildanalyse−Pipeline [2], deren Verarbeitung sich prinzipiell in drei Module gliedert. Das erste Modul extrahiert aus den Sensordaten lokale 3D−Primitive. Die 3D−Tiefenmessung basiert auf der Korrespondenzanalyse, die die tiefenabhängige Disparität der homologen Punkte im Stereobildpaar mißt. Die gemessenen 3D−Primitive werden in 3D−Tiefenkarten abgelegt, in der für jeden Bildpunkt der Abstand der Szene zum Sensor D_k sowie die Meßsicherheit C_k gespeichert ist.



L_k: Linkes Bild D_k: 3D–Tiefenmessung S_k: Oberflächenmaske M_k: 3D Modell
R_k: Rechtes Bild C_k: Meßgenauigkeit P_k: kont. Oberflächenkarte
k: Index der aktuellen Bildes $*_k$: = bewegungskonmpensierte Prädiktionswerte aus M_{k-1}

Bild 2.1 Datengetriebene geometrische Modellierung für einen stereoskopischen Sensor

Die gemessenen 3D−Primitive sind verrauschte, diskrete Abtastwerte (D_k, C_k) der gesuchten, kontinuierlichen Szenenoberflächen. Im zweiten Modul erfolgt daher eine Gruppierung sowie eine Ausgleichung und Interpolation der lokalen 3D−Meßpunkte zu 3D−Oberflächenprimitiven. In der Tiefenkarte werden Bereiche von 3D−Meßpunkten zusammengefaßt, die durch eine kontinuierliche Objektoberfläche

beschreibbar sind. Die Lage der Oberflächen in der Tiefenkarte ist in einer Maskierungskarte S_k gespeichert. Vorwissen und interaktive Zuweisungen über die Lage der Oberflächen lassen sich leicht über die Steuerung in die Analyse einspeisen.

Aus allen Tiefenmessungen einer Oberfläche läßt sich ein parametrisches Oberflächenmodell berechnen. Je nach vorgegebener Ordnung des Oberflächenmodells führt die Ausgleichsrechnung auf Ebenen oder gekrümmte Flächen, welche stetig zusammenhängende Oberflächen $\mathbf{P_k}$ interpolieren [4].

Das dritte Modul transformiert die bisher in Sensorkoordinaten beschriebenen Oberflächen in das Szenenkoordinatensystem und fügt die Oberflächen zu einem konsistenten 3D−Oberflächenmodell M_k zusammen. Eine effiziente Oberflächenrepräsentation wird durch eine parametrische Beschreibung der Objekte mit texturierten Oberflächennetzen erreicht. Die Knoten des Netzes approximieren die Objektgeometrie, während die Facetten des Netzes die photometrische Oberflächentextur speichern. Die Textur gestattet eine photorealistische Rekonstruktion der realen Szene [5].

Zur vollständigen Beschreibung einer Szene mit Objektverdeckungen muß der Sensor durch die Szene bewegt und die jeweils sichtbaren Objektteile im Szenenkoordinatensystem bewegungskompensiert eingefügt werden. Ebenso ist eine Positions− und Abbildungskompensation bei multisensoriellen Messungen erforderlich. Eine notwendige Komponente des geometrischen Modellierers besteht daher in der Messung und Kompensation von Sensorpositionen, um die im Modell gespeicherte Szeneninformation bei der Analyse auszunutzen. Die bereits gespeicherte Szeneninformation M_{k-1} wird in die aktuelle Sensorposition k transformiert und durch Prädiktion den Analysemodulen als Startwerte zugeführt. Die Fusion der prädizierten mit den aktuell gemessenen Werten verbessert so ständig das 3D−Modell.

3 Unterstützung der Analyse durch Bildsynthese

Die automatische Modellierung erfolgt durch die Analyse von Bildfolgen, die einer dreidimensionalen Welt entstammen. Die prinzipielle Architektur zeigt Bild 3.1. Die Modellierung erfolgt in einem iterativen Prozeß durch Vergleich des Modells mit den Bilddaten. Die Oberflächentextur der 3D−Objekte, die während der Bildfolge auf den Kamerasensor projiziert werden, ist die einzige Information, die dem Analysesystem zur Verfügung steht. Die Strategie zur Analyse durch Synthese basiert auf dem Vergleich von Bildern. Aus dem bereits bekannten texturierten 3D−Modell wird eine synthetische Modellansicht des Objektes aus der neuen Sensorposition berechnet und mit dem realen Bild des Sensors verglichen. Abweichungen werden als Modellfehler erkannt und das Modell iterativ modifiziert bis seine Abbildung dem realen Bild möglichst ähnlich wird.

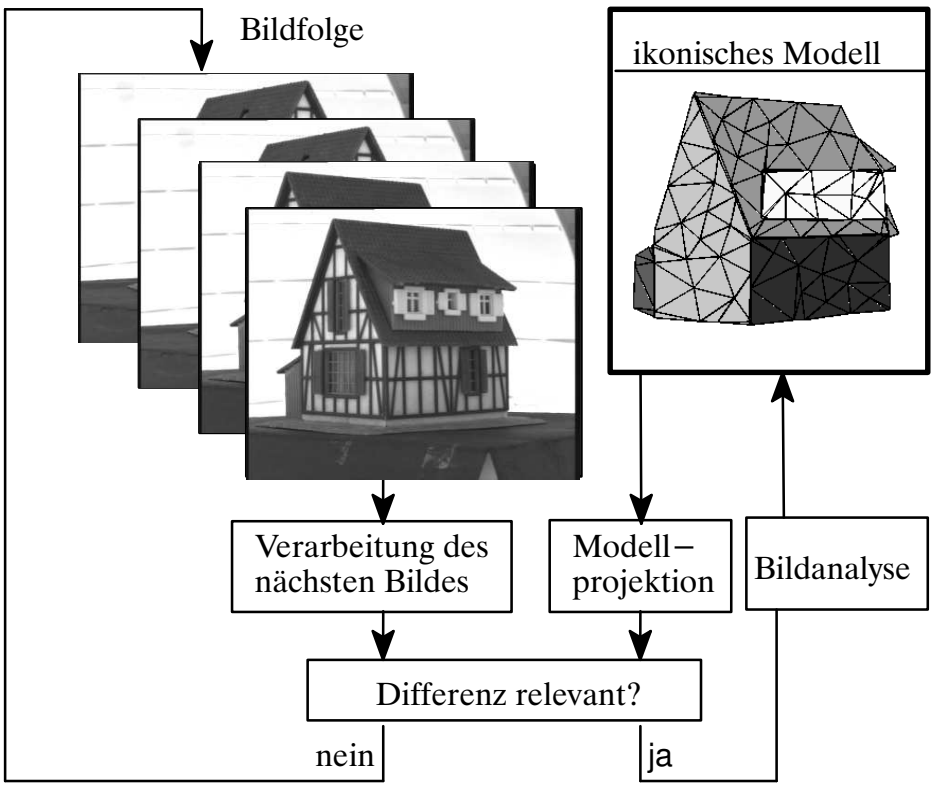


Bild 3.1: Analyse–durch–Synthese–Prinzip

Um mehrere Bilder einer Folge in den Modellierungsprozeß zu integrieren, muß die Relativbewegung zwischen Kamera und Objekt bestimmt werden. Zur Bestimmung der Bewegungsparameter werden Meßpunkte auf der Objektoberfläche ausgewählt und ihr Versatz, der optische Fluß, zwischen zwei aufeinanderfolgenden Bildern gemessen. Die Kamera wird entsprechend der gemessenen Bewegung nachgeführt und eine Modellansicht des Objektes aus der neuen Sensorposition synthetisiert. Dieser Vorgang wird wiederholt bis die Änderung der Bewegungsparameter gegen Null konvergiert [6].

Die Schätzung von Bewegung und Form bei der Analyse durch Synthese basiert auf einem Vergleich im Bildpunktbereich. In diesem Bereich, in dem eine möglichst naturgetreue Bildrekonstruktion gefordert ist, kann die Computergraphik, die aus Modellen Bilder unter Berücksichtigung der texturellen Oberflächeneigenschaften synthetisiert, die Bildanalyse unterstützen.

4 Steuerung der Analyse durch Wissenseinbringung

Die rein datengetriebene Modellierung durch Bildanalyse basiert auf lokalen Bildvergleichsoperationen. Die einzige Information sind die Bildfolgen mit den Projektionen der texturierten 3D−Objekte. Die generierten Modelle liefert zwar eine hinreichend genaue Erklärung der Bilddaten in meßtechnischer Hinsicht, wirken aber

dennoch auf den Betrachter unrealistisch in Bereichen, in denen der Mensch über umfangreiches Zusatzwissen verfügt, wie beispielsweise über die Geometrien künstlicher Objekte. Unstetigkeiten innerhalb von Hauswänden sind ebenso störend wie fehlerhaft plazierte Ecken zwischen Hauswänden. Um Wissen über die Szenenzusammenhänge zu nutzen, wird im folgenden ein wissensbasierter Ansatz zur Steuerung der Analyse vorgeschlagen.

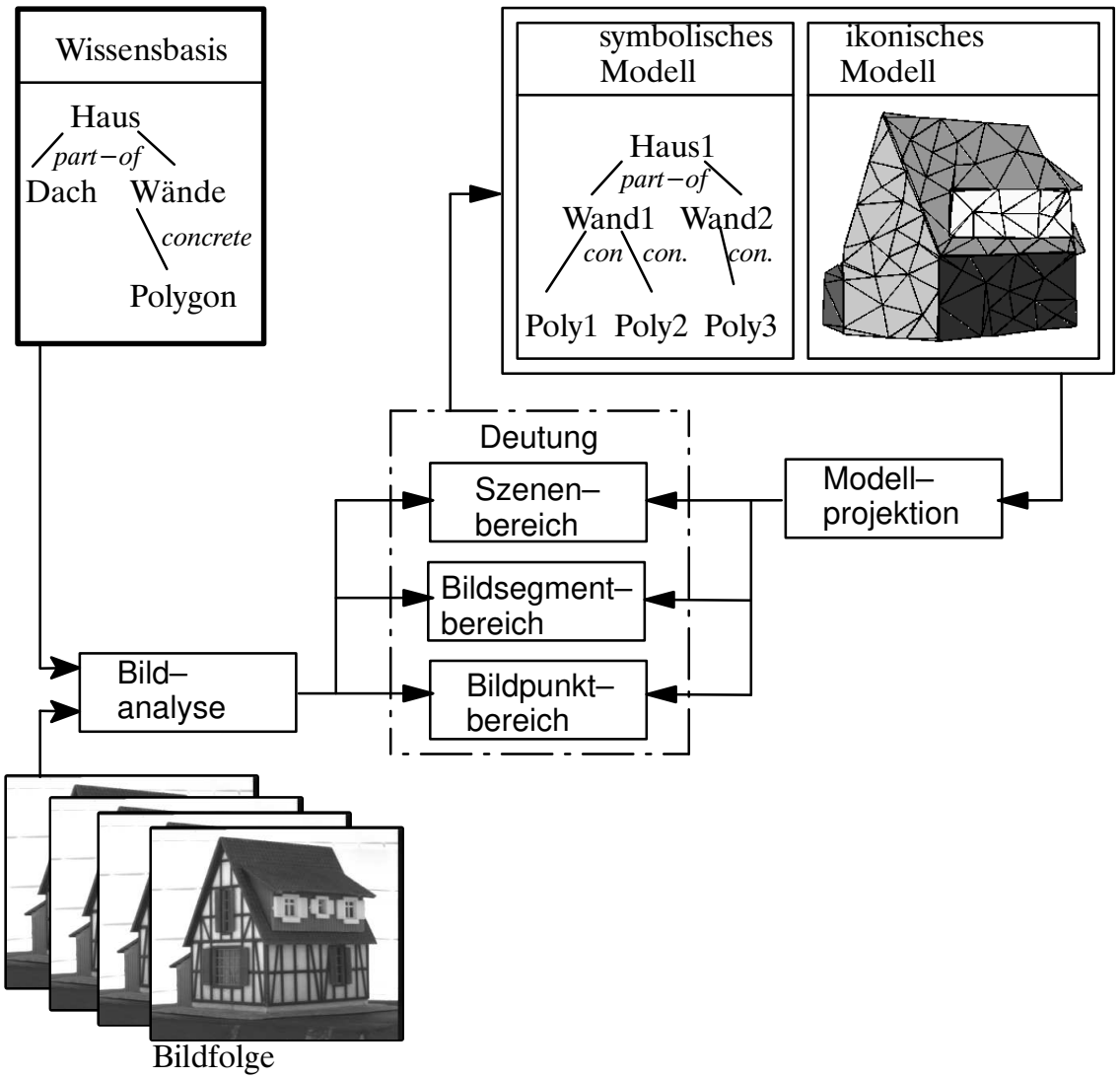


Bild 4.1 Wissensbasierte Analyse duch Synthese Modellierung

Das System muß hierzu um eine Wissensbasis erweitert werden, in der explizit das für die jeweilige Analyse benötigte Zusatzwissen bereitgestellt wird. Für die Modellierung von Häusern können die wichtigen Elemente eines Hauses sowie ihre Relationen zueinander vorgegeben werden. Ein Haus gliedert sich in Dach und Wände, die wiederum in geometrischer Hinsicht Polygone darstellen. Das Wissen über die Objekte der Szene und die Abbildung vom Szenenbereich in den Bildbereich wird als semantisches Netz repräsentiert [7][8]. Bild 4.1 zeigt das semantische Netz für das Ob-

jekt Haus, das sich aus Dach und Wänden zusammensetzt (part−of Relation), die ihre bildliche Realisierung in Polygonen finden (concrete Relation). Die part−of Relation zerlegt ein Objekt in zu detektierende Einzelobjekte. Die concrete Relation transformiert das Erkennungsproblem auf eine andere Ebene. Jede Ebene betrachtet unterschiedliche Aspekte des Objektes, wie beispielsweise szenenspezifische, geometrische und photometrische. Dieses a priori Wissen kann nun zur Verbesserung der Modellierung verwendet werden, da es die Anzahl der zu schätzenden Modellparameter einschränkt und so die Analyse robuster macht.

Um die Modelle durch Wissenseinbringung zu verbessern, ist eine Bildinterpretation notwendig. Das Ziel ist ein symbolisches Modell, das auf seine ikonische Entsprechung in Geometrie und Textur verweist. Die symbolische Beschreibung dient zur Steuerung der Modellierung während der Bildanalyse. Die während des Bildanalyseprozesses generierten typischen Darstellungsformen sind ikonische Bilder, segmentierte Bilder und symbolische Beschreibungen der Szene unter Verwendung von Primitiven. Bild 4.1 zeigt die Abstraktionsniveaus Bildbereich, Bildsegmentbereich und Szenenbereich zur Überführung der ikonischen in die symbolische Darstellung. Die Modellierung durch Interpretation erfolgt in einem iterativen Prozeß.

Im Szenenbereich wird eine Analyse dadurch initiiert, daß z.B. ein bestimmtes Objekt als Haus gedeutet wurde und nun eine Suche nach den noch unbekannten Komponenten 'eines Hauses' (im Sinne eines Prototypen), wie Wände, Dach und so weiter, eingeleitet wird. Für jede Komponente wird eine Hypothese generiert, die zunächst lokal, d.h. mit den Daten der jeweiligen Ebene zu prüfen ist. Fehlen die notwendigen Daten, so wird die Suchanfrage an die untergeordnete Ebene weitergeleitet. Besitzt eine Ebene die zur Verifikation notwendigen Daten oder kennt Verfahren zur Generierung der Daten, so wird die Antwort an die übergeordnete Ebene zurückgereicht. Wichtig ist, daß erkannte Objekte neue Hypothesen generieren können. Die zahlreichen Hypothesen können zum einen durch lokales Wissen oder Einschränkungen von der übergeordneten Ebene auf konsistente und zum anderen durch Kenntnisse über die untergeordnete Ebene (Steuerungswissen) auf erfolgversprechende reduziert werden.

Für die Modellierung bedeutet das, daß die Hypothese des Szenenbereiches 'ebene Wand' verifiziert wird, indem eine Anfrage für die Region 'Wand' an den Bildsegmentbereich weitergeleitet wird. Der Modellierer führt eine Interpolation über die lokalen 3D−Meßpunkte eines Wandsegments durch und liefert eine verbesserte Schätzung für das Oberflächenprimitiv 'ebene Wand' zurück.

Die Bildinterpretation erfolgt parallel zum Aufbau des ikonischen Modells in einem Analyse durch Synthese Prozeß und generiert eine symbolische Beschreibung, durch die den geometrischen Primitiven (Polygon) eine Bedeutung (Wand) zugewiesen wird.

5 Modellierungsergebnisse

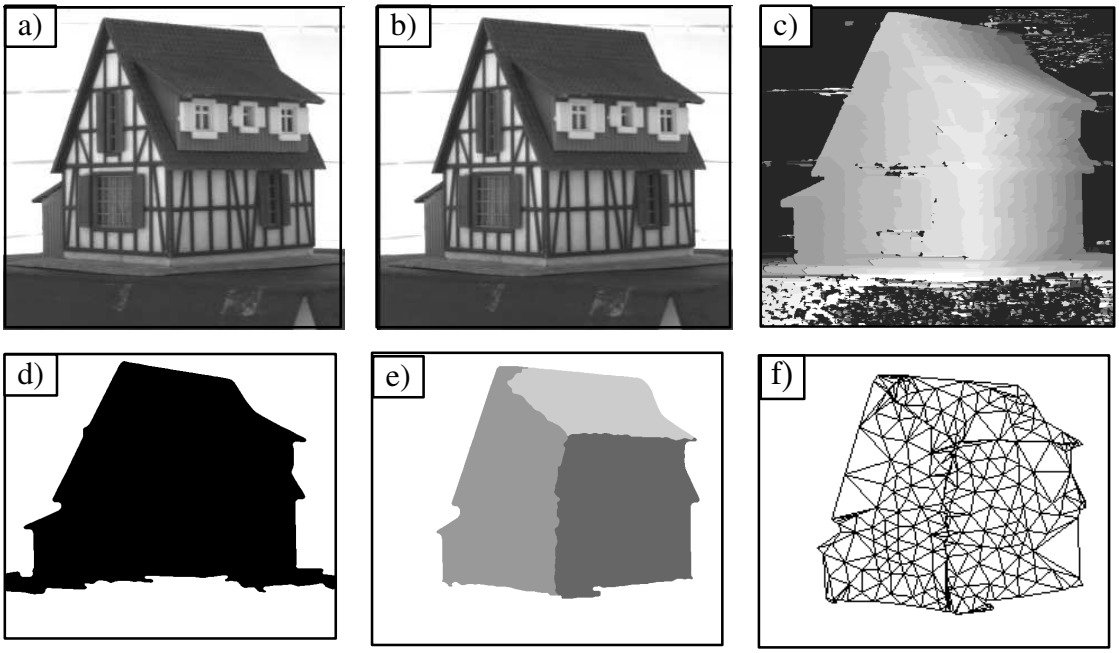


Bild 5.1 a) linkes und b) rechtes Originalbild c) Tiefenkarte d) Segmentiertes Haus e) Randbedingung (Ebenen) f) Geometrie (trianguliertes Oberflächenmodell)

In Bild 5.1 sind Zwischenergebnisse der datengetriebenen Modellierung dargestellt, die durch zusätzliche Randbedingungen gesteuert wurde. Als Sensor wurde ein Stereobildpaar verwendet, dessen linkes Bild in Bild 5.1a dargestellt ist. Zusammen mit dem rechten Stereobild in Bild 5.1b folgt aus der Stereokorrespondenzanalyse die Tiefenkarte in Bild 5.1c und daraus das segmentierte Objekt in Bild 5.1d. In der Tiefenkarte sind nah an der Kamera liegende Oberflächenpunkte hell, entfernt liegende Punkte dunkel dargestellt.

Das Objekt wird als aus Ebenen zusammengesetzt angenommen. Aus der Tiefenkarte lassen sich die lokalen Orientierungen der 3D−Primitive berechnen. 3D−Primitive mit ähnlicher Orientierung werden zu Ebenen zusammengefaßt (Bild 5.1e). Die Ausgleichung der Tiefenkarte über die Ebenen berechnet 3D−Oberflächen, die durch Dreiecksnetze (Bild 5.1f) approximiert werden.

Aus dem linken Originalbild extrahiert der Modellierer die Oberflächentextur, die als Texturkarte zusammen mit der 3D−Geometrie als Modell des Objektes gespeichert wird. Aus dem 3D−Modell können neue Ansichten rekonstruiert werden (Bild 5.2).

a) b)

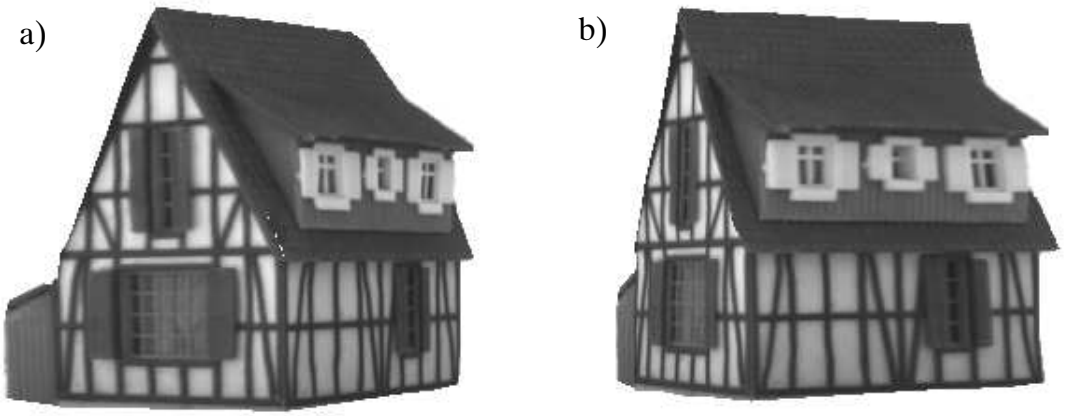


Bild 5.2 Rekonstruktion des Objektes aus Bild 5.1 a) rotiert 0° b) rotiert 30°

Literatur

[1] Ackermann, F., Krzystek, P. , "MATCH–T: Automatic Mensuration of Digital Elevation Models", *Proceedings of Technical Seminar of the Sociedad Espanola de Cartografia Fotogrammetria y Teledeteccion*, pp.67–73, Barcelona, 12th April 1991.

[2] Koch, R.,1993, "Automatic Reconstruction of Buildings from Stereoscopic Image Sequences", *Eurographics '93*, Barcelona,Spain.

[3] Kappei, F., "Modellierung und Rekonstruktion bewegter dreidimensionaler Objekte in einer Fernsehbildfolge", *Dissertation, Institut für Theoretische Nachrichtentechnik*, Universität Hannover 1989.

[4] Terzopoulos, D., "The computation of visible–surface representations", *IEEE Trans. Patt. Anal. Mach. Intell.*, Vol 10, pp.417–438, 1988.

[5] Koch, R.,"Automatic Modelling of Natural Scenes for Generating Synthetic Movies", *Eurographics '90*, Montreux, Switzerland, 1990.

[6] Koch, R., "Dynamic 3–D Scene Analysis through Synthesis Feedback Control, *IEEE Trans. Patt. Anal. Mach. Intell.*, Vol 15 (6), pp.556–568, 1993.

[7] Kummert, F., Niemann, H., Prechtel, R., Sagerer, G., "Control and Explanation in a Signal Understanding Environment" Sprachanalyse", *Signal Processing*, Vol. 32, pp. 111–145, 1993.

[8] Niemann, H., "Pattern Analysis and Understanding", Springer Verlag, 1989.