Ein Szeneninterpretationssystem zur Modellierung dreidimensionaler Körper

Oliver Grau

Institut für Theoretische Nachrichtentechnik und Informationsverarbeitung.
Abteilung "Automatische Bildinterpretation", Prof. Dr.–Ing. C.–E. Liedtke
Universität Hannover, Appelstr. 9A, 30167 Hannover,
E–Mail: grau@tnt.uni–hannover.de

Zusammenfassung. Der Beitrag beschreibt das wissensbasierte 3D–Szenenanalysesystem AIDA. Die automatische Gewinnung dreidimensionaler Modelle für Computergrafik und –animation aus digitalisierten Bildern erfolgt derzeit meist datengetrieben. Fehler bei der Rückgewinnung der Tiefeninformation treten in den dabei gewonnenen Modellen vor allem dort unangenehm in Erscheinung, wo der menschliche Betrachter über Vorwissen verfügt. Das Ziel des hier vorgestellten Ansatzes ist es, die in einer expliziten Wissensbasis abgelegten Objektbeschreibungen für die Modellierung zu nutzen. Dies wird durch eine Interpretation der Szene und Aufstellung von geometrischen Constraints erreicht. Bei der Rekonstruktion der Objektoberfläche werden alle selektierten Constraints und Meßwerte in einer globalen Optimierung berücksichtigt und führen zu visuell ansprechenden Modellen.

Schlüsselwörter: Bildverarbeitung, Szenenanalyse, wissensbasiertes System, 3D–Modellierung, Bildverstehen

1 Einleitung

In der Computergrafik und –animation besteht ein wachsender Bedarf an realistisch aussehenden 3–D Modellen. Typische Anwendungsfelder sind die Film– und Werbebranche sowie Flug– und Fahrsimulationen und die Landschaftsplanung. Besonders bei der letztgenannten Anwendung kommt es oftmals auf eine wirklichkeitsgetreue Rekonstruktion bestehender, bekannter Objekte an. Eine manuelle Rekonstruktion mit CAD–Methoden ist sehr zeit– und kostenaufwendig. Eine automatisierte Modellierung dreidimensionaler Körper, insbesondere von Gebäuden, ist daher Gegenstand zahlreicher Forschungen [1, 2, 3].

Bei der Rekonstruktion der dreidimensionalen Objektgeometrie durch Bildanalyse finden in einem ersten Schritt Verfahren Anwendung, die die Tiefeninformation zurückgewinnen (Shape from X). Ein weites Anwendungsfeld und ein robustes Verhalten weist hier die Verwendung von binokularem Stereo auf. In einem datengetriebenen Ansatz [3] werden mit Hilfe einer Korrespondenzanalyse Tiefenkarten gewonnen, deren Elemente eine Punktwolke im dreidimensionalen Raum bilden. Um zu Oberflächenmodellen zu gelangen, wird die Tiefenkarte durch Dreiecksnetze approximiert. Die geometrische Feinstruktur sowie die photometrischen

Diese Arbeit wurde durch die Deutsche Forschungsgemeinschaft gefördert.

Accepted paper for: *17.DAGM–Symposium Mustererkennung 1995*
13.–15.Sept. 1995 Bielefeld

Oberflächeneigenschaften werden anschließend durch die Rückprojektion der Originalbilder auf das Dreiecksnetz als Fototextur gewonnen. Das entstandene Modell wird abgespeichert und mit Hilfe von Computergrafikmethoden visualisiert.

Das Problem bei der rein datengetriebenen Art der geometrischen Modellierung ist, daß die Approximation der Oberfläche durch das Dreiecksnetz relativ willkürlich ist. Durch Ungenauigkeiten und Fehler in den Tiefenkarten werden Modellierungsfehler besonders dort sichtbar, wo der menschliche Betrachter in Besitz von Vorwissen ist. Dies ist vor allem bei künstlichen Objekten wie Häusern oder ähnlichem der Fall. So wirken beispielsweise gewölbte Wände, geknickte Wand– oder Dachkanten oder nicht rechtwinklige Fenster– und Türrahmen sehr unnatürlich. Eine Erhöhung der Genauigkeit der Tiefenmessungen und Oberflächenapproximation, beispielsweise durch Integration einer Sequenz von Stereobildpaaren, ist nur begrenzt möglich.

Der vorliegende Beitrag beschreibt daher einen alternativen Ansatz. Ausgehend von den Tiefeninformationen aus der Stereokorrespondenzanalyse und Kontur– sowie weiterer Merkmale aus den Eingangsbildern wird das zu modellierende Objekt zuerst interpretiert. Mögliche Deutungen liegen hierzu als explizites generisches Modell in Form eines semantischen Netzes in einer Wissenbasis vor [5]. Während der Interpretation werden den aus den Eingangsbildern gewonnenen Merkmalen Bedeutungen zugewiesen, die die Szene als Szenengraphen beschreiben. Mit Hilfe der Szeneninterpretation können aus der Wissensbasis zusätzliche geometrische Constraints selektiert werden, die zusammen mit den berechneten Bildmerkmalen (Konturen, Regionen, Tiefenwerte) zur Rekonstruktion der dreidimensionalen Objektgeometrie genutzt werden.

Das realisierte System AIDA [5,4] dient in der hier beschriebenen Anwendung primär der Rückgewinnung der Objektoberfläche (Geometrie und Textur). Die durch die Interpretation gewonnene symbolische und strukturelle Beschreibung kann jedoch für verschiedene andere Anwendungen genutzt werden, so daß AIDA als Szenenanalysesystem charakterisiert werden kann. Zu anderen Szenenanalsysesystemen [6,7,8,9] unterscheidet es sich im wesentlichen dadurch, daß es eine dreidimensionale Analyse der Objekte vornimmt. Das heißt, es ist in der Lage, mehrere Ansichten eines Objektes auszuwerten und Verdeckungen zu berücksichtigen. Die Leistungsfähigkeit des Systems wird an der Modellierung von Häusern demonstriert.

Der Abschnitt 2 beschreibt das Gesamtkonzept des realisierten wissensbasierten Szenenanalysesystems AIDA. In Abschnitt 3 wird kurz die Wissensrepräsentation und Szeneninterpretation vorgestellt. Der Abschnitt 4 behandelt die Oberflächenrekonstruktion. Erste Ergebnisse und eine Diskussion schließen den Beitrag ab.

2 Überblick über die Systemkomponenten

Die Abbildung 1 zeigt die wesentlichen funktionalen Bestandteile von AIDA. Als Eingabe stehen ungeordnete Stereobildpaare, Stereobildsequenzen oder auch Einzelbilder (bei Verzicht auf die Stereoanalyse). Vor der Aufnahme der Stereobilder wird eine Kalibrierung der Stereokameraanordnung durchgeführt. Die Bilddaten durchlaufen anschliessend die Bildverarbeitungs–Pipeline. Mit Hilfe der aus der Kamerakalibrierung gewonnenen Kameraparameter werden Radialverzerrungen der Kameras ausgeglichen und die Stereoaufnahmen auf

Accepted paper for: *17.DAGM–Symposium Mustererkennung 1995*
13.–15.Sept. 1995 Bielefeld

Orthogonalgeometrie entzerrt. Durch diese Entzerrung vereinfacht sich die Korrespondenzanalyse auf eine zeilenweise Suche. Suchkriterium ist die maximale normierte Kreuzkorrelation innerhalb eines definierten Suchausschnittes. Aufgrund der in der Kamerakalibrierung gemessenen relativen Kameraposition können die in der Korrespondenzanalyse gemessenen Disparitäten in eine absolute Szenentiefe umgerechnet werden.

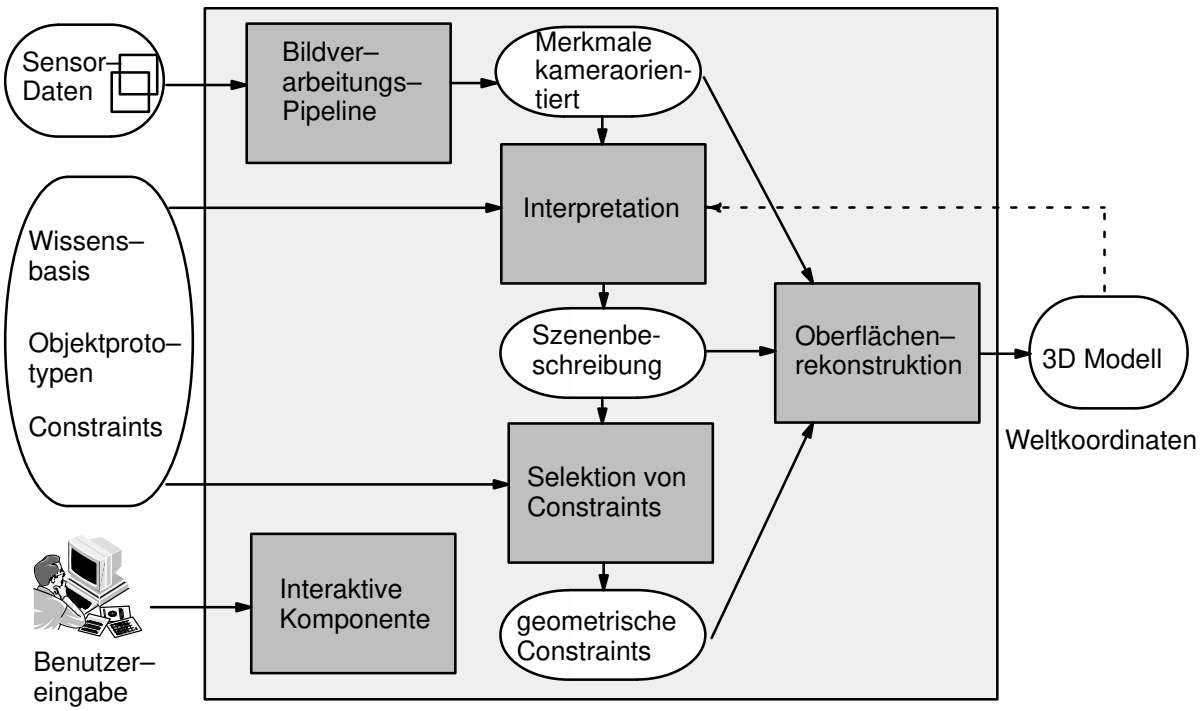


Abb. 1: Die funktionalen Komponenten von AIDA

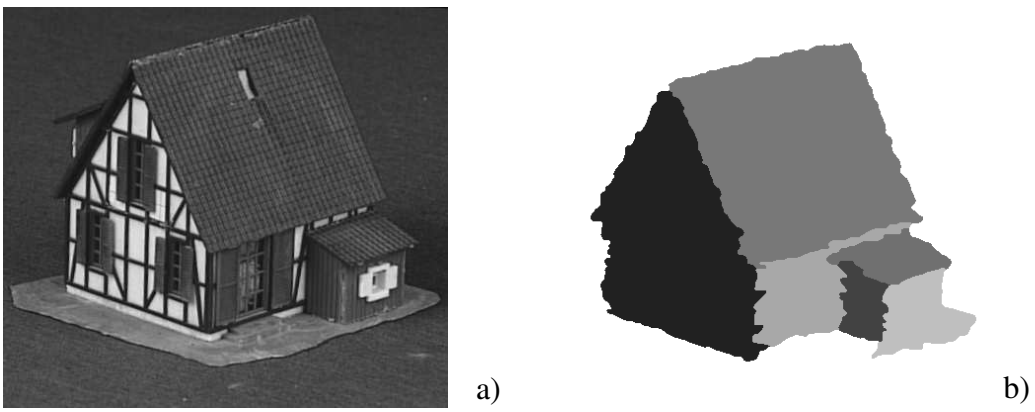


Abb. 2: a) Eingangsbild b) Segmentierungsergebnis aufgrund der Oberflächen–orientierung (Verfahren aus [3])

Die Tiefenwerte werden in Form einer Tiefenkarte bereitgestellt und für die spätere Oberflächenrekonstruktion benötigt. Eine nachfolgende Segmentierung nutzt die Tiefeninformation, um die Objektgrenzen zu bestimmen und planare Strukturen zu detektieren [3]. Dazu faßt ein Clusteralgorithmus Bereiche in der Tiefenkarte zusammen, die die gleiche Normalenrichtung (Orientierung) aufweisen. Die Abbildung 2 zeigt die Anwendung der Segmentierung für eine Laborszene. Zusätzlich werden Konturen in den Bildern extrahiert, die

die genaue Lage der Segmentgrenzen charakterisieren.

Die Wissensbasis enthält eine generische Prototypenbeschreibung der zu modellierenden Objekte in Form eines semantischen Netzes. Die Interpretation weist den Merkmalen und Bildprimitiven, die in der Bildverarbeitungs–Pipeline extrahiert wurden, eine Bedeutung zu. Dazu wird ein Szenengraph ebenfalls in Form eines semantischen Netzes aufgebaut.

Mit der Szenenbeschreibung werden zusätzliche geometrische Constraints aus der Wissenbasis selektiert, die zusammen mit den Bildmerkmalen in der Oberflächenrekonstruktion ausgewertet werden. Das gewonnene 3D–(Teil–)Modell dient der Interpretation bei der Auswertung neuer Ansichten zur Einschränkung des Suchraums (gestrichelter Pfeil in Abb. 1). Die Benutzerschnittstelle dient der Systemüberwachung und der Wissensakquisition.

3 Die Szeneninterpretation

Das Ziel der Interpretation ist es, eine Beschreibung der Szene in Form eines semantischen Netzes aufzubauen. Die Interpretationskomponente greift dazu auf eine generische Prototypenbeschreibung in der Wissensbasis zurück. Die Abbildung 3 zeigt einen stark vereinfachten Ausschnitt einer solchen Beschreibung. Die Darstellungsform ist stark angelehnt an die Wissensrepräsentation von ERNEST [9]. Die Knoten oder Konzepte, als Rechteck dargestellt, repräsentieren eine frame–artige Beschreibung, die eine Reihe von Attributen oder Slot–Einträgen besitzen können.

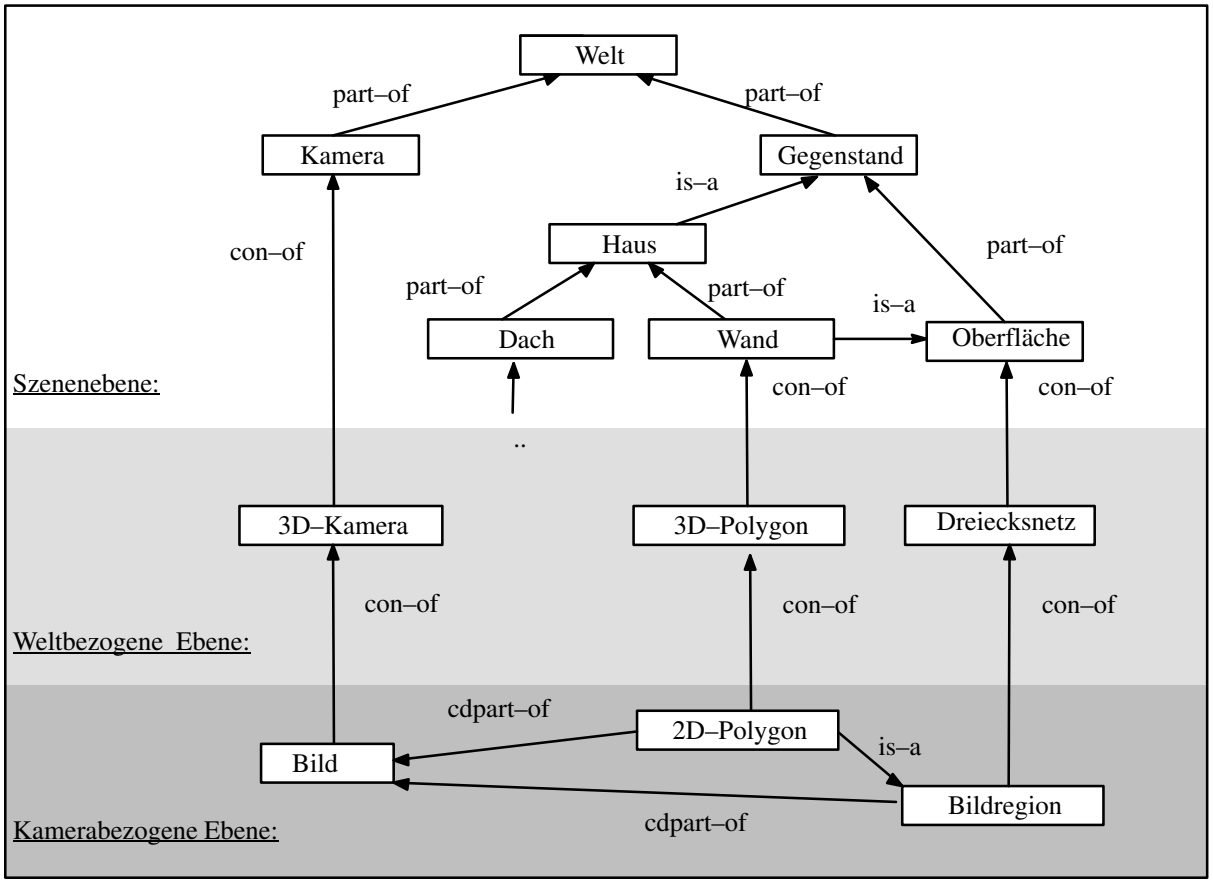


Abb. 3 Stark vereinfachter Ausschnitt aus dem semantischen Prototypen–Netz

Mit Hilfe der Kanten, auch Links genannt ist es möglich, eine kompositionelle Zerlegung der Szenen und der Objekte zu beschreiben (part–of, cdpart–of) und eine Vererbungshierarchie aufzubauen (is–a). Der Kantentyp Cdpart–of (context dependend part) ist eine spezielle Form des Part–of–Links. Mit seiner Hilfe werden Beziehungen modelliert, die nur im Kontext existieren. So kann eine Bildregion beispielsweise nur zusammen mit einem (kontext–) zugehörigem Bild existieren. Eine weitere spezielle Form des Part–of–Links ist der Concrete–of–Link. Damit werden Konzepte unterschiedlichen Konkretisierungsgrades oder Abstraktionsniveaus verbunden. Die Wissensbasis in AIDA ist derzeit in drei Abstraktionsniveaus unterteilt: die Szenenebene, die weltbezogene Ebene und die kamera– oder sensorbezogene Ebene.

Die Interpretationskomponente verwendet als Eingangsdaten die Bilder, Bildprimitive und Merkmale (in Abb. 3 nicht dargestellt) aus der Bildverarbeitungs–Pipline und baut nun Schritt für Schritt eine Beschreibung auf. Dabei werden Hypothesen erzeugt und durch Bewertungen verifiziert. Die Hypothese, ob eine Szene ein Haus zeigt, wird bestätigt, wenn die im Signal (den Bildern) gefundenen Primitive oder Merkmale zu der Prototypenbeschreibung passen. Die Auswahl und Verfolgung konkurrierender Hypothesen erfolgt mit einem modifizierten A*–Algorithmus [9]. Die Interpretation verwendet dreidimensionales Wissen (z.B. daß Wände meist senkrecht stehen und rechtwinklig verbunden sind), um den Suchraum einzuschränken, und ist in der Lage, Verdeckungen zu berücksichtigen. Eine detailliertere Beschreibung der Interpretation und Wissensrepräsentation in AIDA gibt [5].

Die Interpretation ist beendet, wenn alle Hypothesen verifiziert und damit alle offenen Deutungsmöglichkeiten abgearbeitet sind. Die Szenenbeschreibung dient dann der nachfolgenden Oberflächenrekonstruktion als strukturelle Beschreibung. Zusätzlich werden geometrische Constraints aus der Wissensbasis selektiert, so z.B. die Forderung bei Polygonen, daß alle Stützpunkte in einer Ebene liegen oder Winkelvorgaben zwischen Polygonen bei Hauswänden.

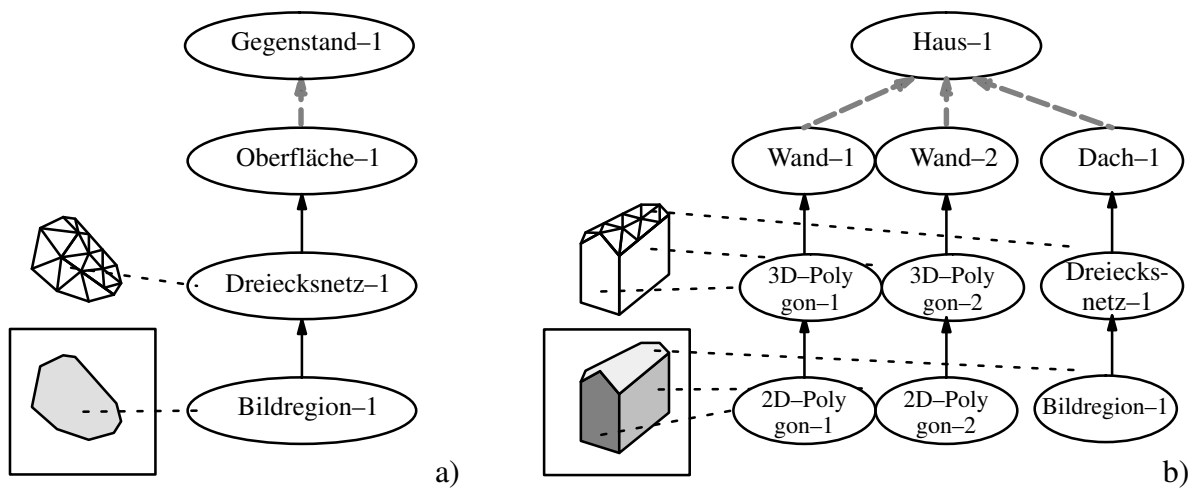


Abb. 4: Szenenbeschreibung mit semantischem Netz a) unbestimmtes Objekt b) Haus. Kanten: part–of, concrete–of

Im allgemeinen gilt, daß umso mehr geometrische Constraints für die Oberflächenrekonstruktion aufgestellt werden können, desto spezieller die Interpretation ist. Kann einem Objekt wie in Abbildung 4a keine spezielle Bedeutung zugewiesen werden, so lassen sich keine zusätzlichen Constraints nutzen, und die Objektoberfläche wird allein aus den gewonnen Tiefenmessungen (als Dreiecksnetz) modelliert. Wird das Objekt wie in Abbildung 4b als Haus und dessen Bestand-

teile erkannt, so lassen sich Constraints wie die Ebenheit und Winkelbeziehung der Wände nutzen. Können Bestandteile,wie hier das Dach nicht als spezielles Flach– oder Steildach erkannt werden, so erfolgt die Modellierung allein aus den Tiefenmessungen. Am Ende der Interpretation sind alle Daten (Bilder, Primitive, Merkmale, Deutungen) Bestandteil der Szenenbeschreibung und liegen als semantisches Netz vor.

Nur kurz erwähnt werden soll der Aspekt der Vervollständigung [4]: Durch das in der Interpretation gewonnene Wissen über die zu modellierenden Objekte können auch Teile modelliert werden, die nicht durch das Signal (Bild) unterstützt werden. So können beispielsweise bei Modellierung aus einer Ansicht und dem Wissen, daß Häuser aus mindestens vier Wänden bestehen, die Rückwände eines Hauses eingesetzt werden.

4 Die Oberflächenrekonstruktion

Die Oberflächenrekonstruktion ist eine integrative Komponente. Sie fügt die Tiefenmessungen und die aus der Interpretation gewonnenen Constraints zu einem dreidimensionalen Oberflächenmodell zusammen. Dies geschieht inkrementell, das heißt es wird im ersten Schritt ein Teilmodell aus der ersten Kameraansicht gewonnen. Danach werden vormals verdeckte Objektteile durch die Daten aus hinzukommenden Ansichten erweitert, beziehungsweise bereits modellierte Objektteile durch neue Messungen verfeinert.

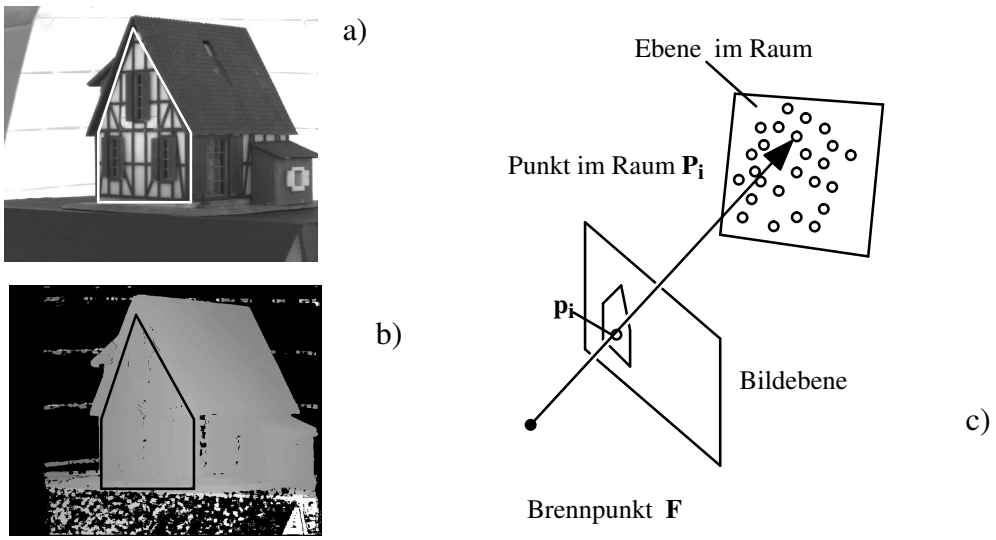


Abb. 5: Rekonstruktion eines planaren Polygons a) Eingangsbild, b) Disparitätskarte, c) Regression über alle Punkte der Bildregion

Die Oberflächenrekonstruktionskomponente von AIDA verarbeitet derzeit als Eingangsdaten Tiefenmessungen aus mehreren Ansichten und die begrenzenden Konturen von Teilflächen. Als Constraints können verarbeitet werden: Bedingung der Ebenheit (Planarität) einer Teilfläche und Winkel zwischen Teilflächen. Ebenheit wird erreicht, indem, wie in Abbildung 5 gezeigt, eine Teilfläche als (planares) Polygon modelliert wird. Zur Rekonstruktion des Polygons wird eine lineare Regression über alle Tiefenmessungen berechnet, die innerhalb der zugehörigen Bildregion liegen. Dies gleicht Ungenauigkeiten in den Einzelmessungen der Tiefenkarte aus.

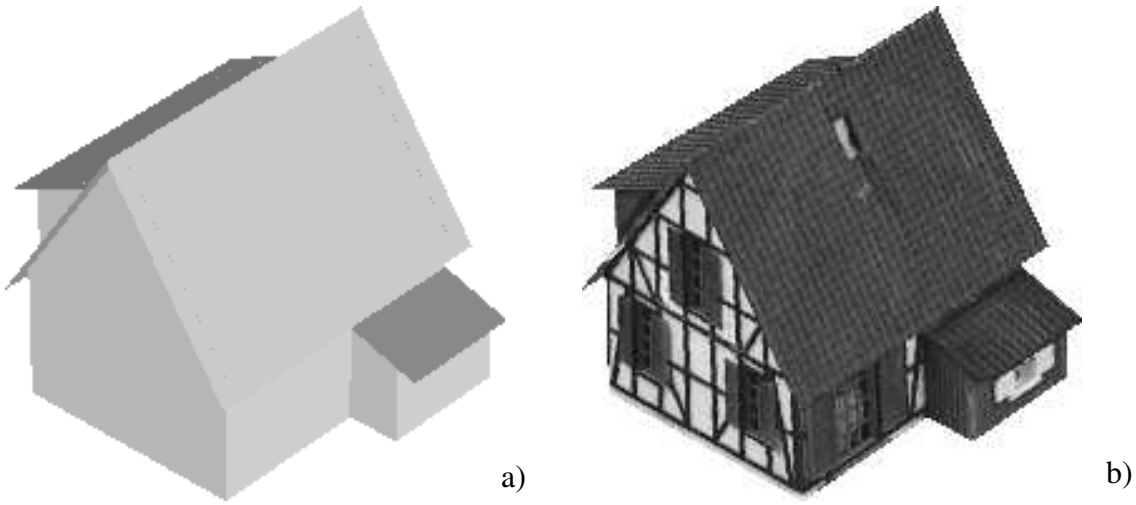
Die Begrenzung der Raumpolygone ergibt sich aus den Schnittgeraden der benachbarten Ebenen. Die Rückprojektion in die Bildebene muß mit der im Originalbild extrahierten Kante über-

einstimmen. Bei einer lokalen Adaption, also beispielsweise der Verschiebung zweier Eckpunkte eines Polygons, wird aber unter Umständen die Ebenheitsbedingung oder ein anderes Constraint der benachbarten Teilflächen verletzt. Daher wird die Einhaltung aller geometrischen Constraints unter Verwendung der gewonnenen Tiefenmessungen durch eine globale Ausgleichsrechnung gelöst. Dazu werden Kostenfunktionen für jedes Constraint aufgestellt und iterativ minimiert.

Eine weitere wichtige Aufgabe der Oberflächenrekonstruktions–Komponente ist die Auto–Kalibrierung neuer Kameraansichten: Die Tiefenmessungen aus der Stereokorrespondenzanalyse sind kamerabezogen. Eine Integration erfordert Umrechnung in weltbezogene Koordinaten. Dazu müssen die absoluten Kamerapositionen in Weltkoordinaten bekannt sein. Die Auto–Kalibrierung liefert die Kamerapositionen durch eine Schätzung der relativen Bewegung zwischen einem (bereits modelliertem) Teilmodell und der neuen Kameraansicht [3].

5 Ergebnisse

Die Abbildung 6 zeigt eine synthetisierte Ansicht eines modellierten Spielzeughauses in schattierter Flächendarstellung und mit aus den Originalbildern gewonnener Fototextur. Das Modell wurde aus zwei Stereobildpaaren rekonstruiert. Die Interpretation verwendet ein semantisches Netz mit 41 Knoten und 64 Kanten, das eine generische Beschreibung für ein Haus enthält. Der Kontrollalgorithmus startet die Interpretation Top–Down. Die Zuordnung der Bedeutungen zu den Bildregionen wurde manuell überwacht und arbeitet beim derzeitigen Stand der Systemimplementation nicht vollautomatisch.



a) b)

Abb. 6: Synthetisierte Ansicht des Modells in a) schattierter und b) texturierter Darstellung

Aus der gewonnenen Szenenbeschreibung extrahiert die Oberflächenrekonstruktion die benötigten Daten. Verwendet werden Konturbeschreibungen der Objektteile und die Tiefenmessungen aus der Stereokorrespondenzanalyse. Das resultierende Modell besteht aus 14 Polygonen.

6 Diskussion

Der Beitrag beschreibt das wissensbasierte 3D–Szenenanalysesystem AIDA. Das Ziel, geometrische Constraints, die für bestimmte Objekte gelten, für die Modellierung zu nutzen, wird durch

eine Interpretation der Szene erreicht. Mit Hilfe der so gewonnenen Szenen– bzw. Objektbeschreibung können die für die Rekonstruktion der Objektoberfläche notwendigen Constraints aus der Wissensbasis selektiert werden. Die Qualität des resultierenden 3D–Modells hängt unmittelbar von der Güte der Interpretation ab, der somit eine zentrale Rolle in dem Gesamtkonzept zukommt.

Die Formulierung von Zusammenhängen in einem expliziten semantischen Netz ermöglicht die Einführung sehr spezieller Constraints, wie der Winkel zwischen Hauswänden. Der Einsatz des Systems für eine neue Aufgabenstellung erfordert lediglich den Austausch oder die Erweiterung des semantischen Netzes.

Die Einhaltung aller selektierten Constraints und der Meßwerte (Konturbeschreibungen und Tiefenmessungen) bei der Rekonstruktion der Objektoberfläche wird durch eine globale Optimierung erreicht. Sie führt zu einer subjektiv besseren Erscheinung der 3D–Modelle in der Visualisierung. Der Gewinn ist umso größer, je spezieller die Interpretation der Objekte ist. Dies kann insbesondere bei künstlichen Objekten wie Häusern gezeigt werden.

Die Implementierung des Systems ist nicht abgeschlossen. Wurde in der ersten Phase die Interpretation manuell unterstützt, so ist es Gegenstand laufender Arbeiten, diese weiter zu automatisieren. Ständig erweitert und verallgemeinert wird auch die Wissensbasis.

Literatur

1. Leberl, F., Gruber, M., Uray, P., Madritsch, F., "Trade–Offs in the Reconstruction and Rendering of 3–D Objects", Mustererkennung 1994, 16. DAGM Symposium und 18. Workshop der ÖAGM, Wien 1994.
2. Braun, C., Kolbe, Th., Lang, F., Schickler, Steinhage, Cremers, Förstner, W., Plumer, L., "On the Models for Photogrammetric Building Reconstruction". Eingereicht zur Veröffentlichung in Computer & Graphics, 1995.
3. Koch, R., "3–D Surface Reconstruction from Stereoscopic Image Sequences", 5th International Conference on Computer Vision ICCV'95, Cambridge MA, 20.–23. Juni 1995.
4. Grau, O., Tönjes, R., "Knowledge Based modelling of Natural Scenes". European Workshop on Combined Real and Synthetic Image Processing for Broadcast and Video Production. 23.–24. Nov. 1994 Hamburg.
5. Liedtke, C.–E., Grau, O., Growe, S., "Use of Explicit Knowledge for the Reconstruction of 3–D Object Geometry". 6th International Conference Computer Analysis of Images and Patterns CAIP'95. 6.–8. Sept. 1995 Prag.
6. McKeown, D. M. Jr., Harvey, W. A. Jr., McDermott, J., "Rule–Based Interpretation of Aerial Imagery", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. PAMI–7, No. 5, pp. 570–585, Sept. 1985.
7. Matsuyama, T., Hwang, V.S.–S., "SIGMA : A Knowledge–Based Aerial Image Understanding System", *Plenum Press*, New York 1990.
8. Clément, V., Giraudon, G., Houzelle, S., Sandakly, F., "Interpretation of Remotely Sensed Images in a Context of Multisensor Fusion Using a Multispecialist Architecture", IEEE Trans. on Geoscience and Remote Sensing, Vol. 31, No. 4, July 1993.
9. Niemann, H., Sagerer, G., Schröder, S., Kummert, F., "ERNEST: A Semantic Network System for Pattern Understanding", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 12, No. 9, pp. 883–905, Sept. 1990.

Accepted paper for: *17.DAGM–Symposium Mustererkennung 1995*
13.–15.Sept. 1995 Bielefeld