Wissensbasierte Modellierung von Landschaften für die Visualisierung

RALF TÖNJES[1]

Der Beitrag beschreibt ein System zur automatisierten Erzeugung realistischer virtueller Landschaftmodelle aus überlappenden Luftbildern. Ausgehend von der stereoskopischen Korrespondenzanalyse wird ein digitales Geländemodell erzeugt, das zur effizienten Visualisierung durch ein Polygonnetz mit überlagerter Phototextur approximiert wird. Wissen über die Objektgeometrie wird verwendet, um stetige Regionen durch Teilnetze zu modellieren und Höhensprünge an Diskontinuitäten wie Waldkanten zu erzwingen. Zur objektspezifischen Modellierung wird eine Bildinterpretation durchgeführt.

1. Einleitung

Für die Visualisierung von Landschaften werden realistische 3D–Modelle benötigt, aus denen neue synthetische Ansichten berechnet werden können. Für die Fahr– und Flugsimulatoren, aber auch für den virtuellen Kulissenbau in der Film– und Werbeindustrie, besteht ein großer Bedarf an realistischen Modellen. Die sinkenden Kosten für leistungsfähige Graphikhardware eröffnen der Landschaftssimulation zunehmend neue Anwendungsbereiche. Auch in der Umweltüberwachung und Landschaftsplanung gewinnt die Visualisierung virtueller Landschaften mehr und mehr an Bedeutung. Ein Landschaftsarchitekt könnte ein neues Bauprojekt als CAD–Modell in die Landschaft einfügen und die Veränderung als Simulation visualisieren.

Zur 3D–Rekonstruktion von Landschaften hat sich die Korrespondenzanalyse als leistungsfähiges Verfahren zur Messung der Höhe aus Stereoluftbildern etabliert (ACKERMANN, 1991). Zur effizienten Visualisierung wird die Höhenkarte unterabgetastet und durch ein räumliches Polygonnetz approximiert. Die geometrische und photometrische Feinstruktur wird durch eine Phototextur modelliert, die aus den Luftbildern gewonnen und auf das Oberflächennetz projiziert wird.

Die 3D–Rekonstruktion einer Szene aus ihren 2D–Ansichten ist jedoch ein inverses und mathematisch unterbestimmtes Problem. Eine realistische Rekonstruktion muß folgende Einschränkungen berücksichtigen:

– Aus den Bildern läßt sich nur ein unvollständiges Modell, ein $2^1/_2$ D–Modell, gewinnen, da einige Ansichten der Objekte, wie z.B. Waldkanten in Luftbildern, aufgrund von Verdeckungen nicht sichtbar sind.
– Wegen der begrenzten Sensorauflösung können Details, wie Gebäude und Bäume in Luftbildern, nicht rekonstruiert werden. Insbesondere sind fehlende Höhensprünge in Verbindung mit Schattenwürfen kritisch, da ein Schatten ohne Höhensprung mit zugehörigen Verdeckungen optisch falsch erscheint.
– Die Approximation der Geometrie durch ein äquidistantes Raster ebener Polygone berücksichtigt nicht die Konturen der modellierten Teilobjekte und führt häufig zu fehlerhaften Bruchkanten.

[1] Institut für Theoretische Nachrichtentechnik und Informationsverarbeitung,
Universität Hannover, Appelstr. 9A, D–30167 Hannover,
E–mail: toenjes@tnt.uni–hannover.de, WWW: http://www.tnt.uni–hannover.de/~toenjes.html

Das hat zur Folge, daß die Modelle häufig nicht den Erwartungen des Betrachters entsprechen und unrealistisch wirken. Der Betrachter weiß, daß Straßen stetig verlaufen und an Waldrändern ein Höhensprung auftritt. Daher verwendet das vorgestellte System Vorwissen über die Landschaft zur Rekonstruktion, um den Realismus des Modells zu verbessern. Um das a priori Wissen über Objektklassen wie Straßen, Wälder, Wiesen etc. zu nutzen, ist eine Bildinterpretaion notwendig. Die Bildinterpretation hat zum Ziel, den Regionen im Bild Bedeutungen zuzuweisen und somit die verschiedenen Objekte zu unterscheiden. Die erkannte Objektbedeutung wird anschließend verwendet, um die geometrischen Parameter des Modells für die Rekonstruktion aus dem Bildausschnitt einzuschränken. Zur Bildinterpretation wird das im folgenden beschriebene Modellierungssystem AIDA (Automatic Image Data Analyzer) eingesetzt (GRAU, 1995; TÖNJES 1994).

Um die Wissensbasis leicht an neue Modellierungsaufgaben anpassen zu können, muß das Wissen in Hinblick auf Wiederverwendbarkeit und Erweiterbarkeit formuliert werden. Hierzu wird ein explizite Darstellung der Wissensinhalte mithilfe semantischer Netze verwendet.

Der folgende Abschnitt gibt einen Überblick über das Gesamtsystem. Der dritte Abschnitt stellt die Methoden der expliziten Wissensrepräsentation vor. Anschließend wird diese Wissensbasis in Abschnitt vier zur Interpretation und Rekonstruktion genutzt. Abschnitt fünf zeigt, wie das Landschaftsmodell für die Echtzeitvisualisierung aufbereitet wird. Der Artikel schließt mit einer Zusammenfassung der Ergebnisse.

2. Systemübersicht

Bild 1 gibt einen Überblick über das Gesamtkonzept, das die drei Module Registrierung, 3D–Modellierung und Visualisierung umfaßt.

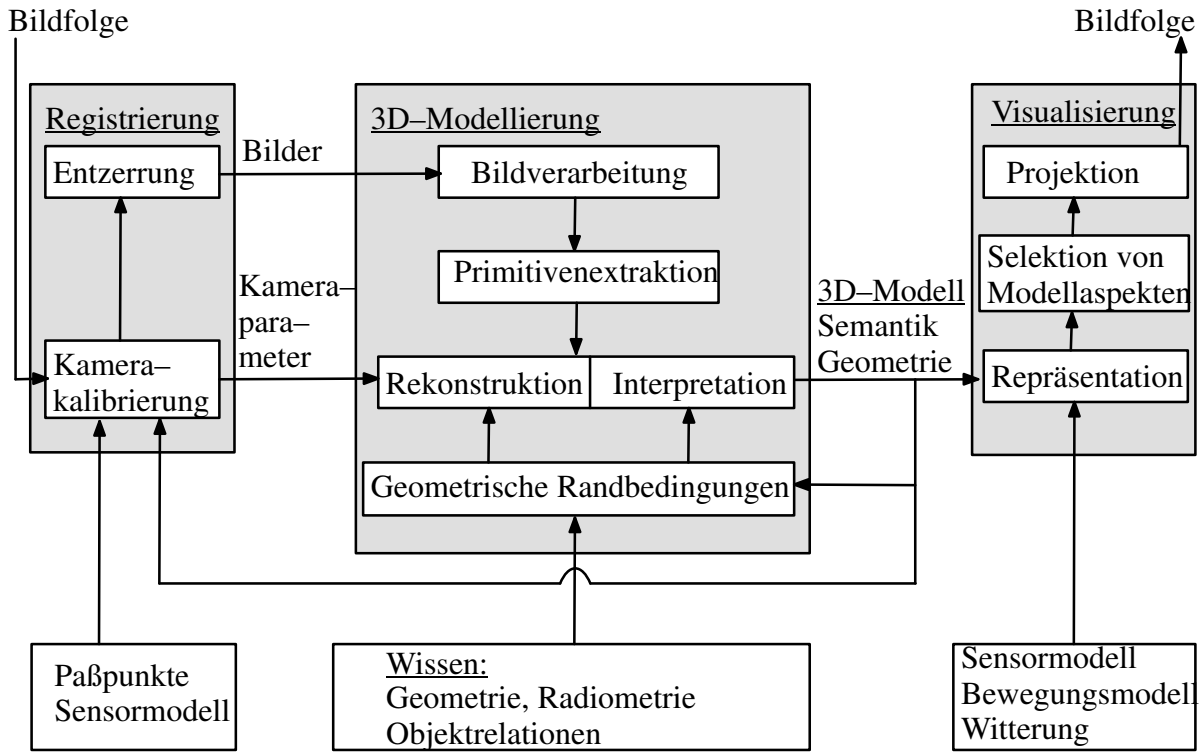

Abb. 1: Architektur des Systems zur 3D–Modellierung und Visualisierung

Eingangsdaten für die Modellierung sind überlappende Bildpaare. Um die Bilder im gewählten Koordinatensystem zu registrieren, ist die Bestimmung der inneren und äußeren Kameraorientierung notwendig. Die Kalibrierung der Kamera erfolgt mithilfe von Paßpunkten, deren räumliche Lage bekannt ist. Anschließend werden die Bilder geometrisch und radiometrisch entzerrt bzw. auf Epipolargeometrie transformiert, um die nachfolgende Stereobildanalyse zu erleichtern.

Die Modellierung nutzt Wissen über die in der Szene erwarteten Objekte, wie Wissen über deren geometrische und radiometrische Eigenschaften und den Relationen zwischen den Objekten.

Das Bildverarbeitungsmodul berechnet aus dem stereoskopischen Bildpaar eine Höhenkarte. Basierend auf der Höhenkarte und den photometrischen Merkmalen werden die Bilder segmentiert. Die Primitivenextraktion liefert eine erste symbolische Beschreibung durch geometrische Primitive. Die Interpretation gruppiert die geometrischen Primitive unter Verwendung der in der Wissensbasis formulierten Prototypen und ordnet ihnen eine szenenspezifische Bedeutung zu. Aus der Objektbedeutung werden geometrische Randbedingungen abgeleitet, die die freien Modellparameter einschränken und die Oberflächenrekonstruktion verbessern.

Für die Visualisierung wird die Szene durch hierarchisch strukturierte Polygonnetze mit Phototextur für die Einzelobjekte approximiert. Um einen räumlichen Eindruck von den Modellen zu vermitteln, wird ein Stereoprojektionssystem mit polarisiertem Licht verwendet. Der 3D–Charakter des Modells kann zudem durch Beleuchtung, Dunsteffekte und Bewegung durch die Szene hervorgehoben werden.

3. Wissensrepräsentation

Wissen kann allgemein durch formale Logik, Fuzzy Logik, Frames (CLEMENT, 1993; FORESTI , 1993) und semantisches Netze (NIEMANN, 1990), Produktionssysteme, regelbasierte Systeme (MATSUYAMA, 1990; MC KEOWN, 1985) und neuronale Netze repräsentiert werden (SHAPIRO, 1992). Für die Repräsentation von strukturellen Beziehungen eignen sich sehr gut semantische Netze. Daher wird hier, inspiriert durch die Arbeiten von Niemann (1990), ein semantisches Netz zur Wissensrepräsentation gewählt.

Ein semantische Netz besteht aus Knoten und Kanten zwischen diesen. Die Knoten sind als Frames realisiert, welche eine Beschreibung der Attribute, Relationen und Methoden enthalten. Die Kanten oder Relationen des semantischen Netzes beschreiben die Beziehungen zwischen den Knoten. Die verwendeten Relationen beschreiben die Dekomposition des Modells in seine Bestandteile (part–of), deren Spezialisierung (is–a) und konkrete Realisierung in den Daten (con–of).

Abbildung 2 zeigt ein vereinfachtes semantisches Netz für die Landschaftsmodellierung, das in drei Schichten organisiert ist: Die Szenenebene beschreibt Objekte und Objektteile auf einem hohen Niveau in Form von abstrakten Begriffen und ordnet den Teilen szenenspezifische Bedeutungen zu. Die mittlere Schicht enthält die Materialeigenschaften der Objekte sowie deren geometrische Ausprägungen im Dreidimensionalen und in absoluten Weltkoordinaten. Die unterste Schicht ist sensorbezogen und beschreibt die sensorspezifische, photometrische und geometrische Abbildung der Objekte. Jede Ebene verwendet ein eigenes geeignetes Vokabular. Die Eigenschaft Größe wird z.B. für die 3D–Geometrieebene in Metern und für die Sensorebene in Pixel gemessen.

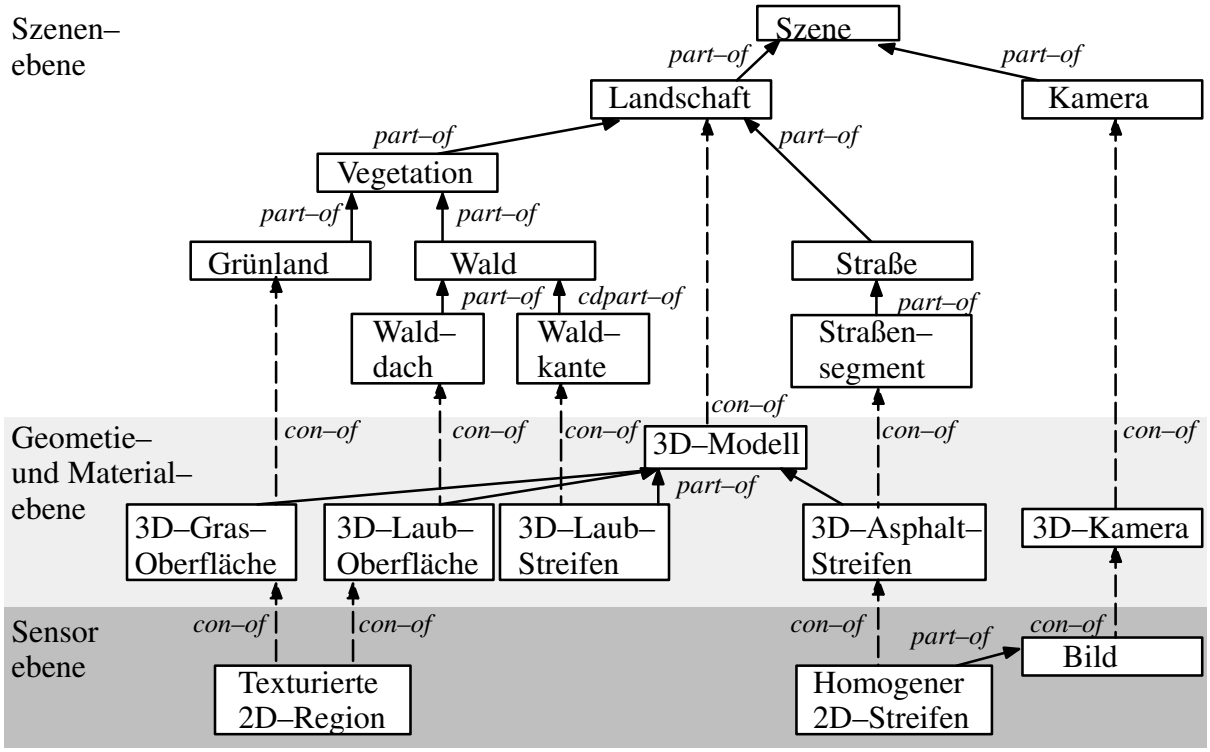

Abb. 2: Stark vereinfachter Ausschnitt aus dem semantischen Prototypen–Netz.

Für die Modellierung von Landschaften sollen hier Straßen, Wälder und Grünflächen unterschieden werden. Der Wald setzt sich aus Walddach und Waldkante zusammen, die separat zu modellieren sind. Hierbei muß jedoch nur das Walddach als texturierte Region in den Bilddaten sichtbar sein. Straßen werden im Luftbild als homogene Streifen abgebildet. Die initialen Konzepte 'texturierte 2D–Region' und 'homogener 2D–Streifen' besitzen Methoden zur Segmentierung texturierter Regionen bzw. homogener Streifen aus den Bilddaten.

4. Wissensnutzung

4.1 Interpretation

Die Interpretation hat zum Ziel, den Objekten der betrachteten Szene die zutreffenden Knoten des semantischen Prototypennetzes zuzuordnen. Die Bildinterpretation nutzt die Wissensbasis, um Hypothesen über die in der Szene erwarteten Objekte zu formulieren, d.h. zu instanzieren. Die Hypothesen werden von der Szenenebene nach unten in die Sensorebene propagiert und in den Daten überprüft. Hierzu wird eine Methode zur Segmentierung texturierter Regionen bzw. homogener Streifen aufgerufen. Das Ergebnis der Verifikation wird dann an die übergeordneten Konzepte zurückgegeben, die dann weitere Hypothesen instanzieren.

Beim Auftreten von Mehrdeutigkeiten wird für jede Deutungsmöglichkeit eine Hypothese erzeugt und bewertet. Ein modifizierter A*–Algorithmus selektiert daraus die am besten zutreffende Hypothese. Die so gewonnene Szenenbeschreibung dient dazu, eine objektspezifische Modellierung durchzuführen und die Parameter für die Oberflächenrekonstruktion einzuschränken.

Bislang erfolgt die Interpretation primär durch Propagierung der Segmentierungsergebnisse in die Szenenebene. Weitere Analysestrategien, die andere strukturelle Zusammenhänge des semantischen Netzes auswerten, sind in der Entwicklung.

4.2 Segmentierung

4.2.1 Segmentierung linienförmiger Objekte

Die Extraktion der homogenen Streifen basiert auf Gradientenfiltern mit nachfolgender lokaler Rangfolge um schwache Konturen hervorzuheben. Nach der Schwellwertbildung ergibt die Suche nach langen parallelen Linien die Kandidaten für homogen Streifen. Abbildung 3a zeigt das Segmentierungsergebnis.

4.2.2 Segmentierung flächenförmiger Objekte

Die Wald– und Grünlandflächen natürlicher Landschaften werden durch unterschiedliche Texturen charakterisiert. Die Textur einer Klasse k, z.B. Wald, sei durch einen stationären stochastischen Prozeß entstanden. Bei statistischer Unabhängigkeit wird die Textur durch die Wahrscheinlichkeit $P_1(y|k)$ der Luminanzwerte y beschrieben, die aus dem Luminanzhistogramm der Lernregion berechnet werden kann. Da die Auftretenswahrscheinlichkeit eines Luminanzwertes im allgemeinen jedoch von seinen Nachbarn abhängt wird im folgenden ein Texturmodell zweiter Ordnung verwendet, das auf Gibbs–Zufallsfeldern basiert. Dieses Texturmodell berücksichtigt die statistischen Abhängigkeiten ausgewählter Pixelpaare, den Cliquen, in einer lokalen Nachbarschaft. Für jede Clique wird in einer Lernphase die

– Wahrscheinlichkeit P_2 gleicher Klassenzugehörigkeit,
– Wahrscheinlichkeit P_3 der Luminanzdifferenz innerhalb einer Region,
– Wahrscheinlichkeit P_4 der Luminanzdifferenz zwischen Regionen.

berechnet. Gibbs–Zufallsfelder beschreiben die Verbundwahrscheinlichkeitsverteilung:

$$P(\overline{k}|\overline{y}) = \frac{1}{Z} \quad e^{-\sum_c V_c(y|k)} \quad \text{mit } Z = \text{konstant und } V_c(y|k) = -\sum_{i=1}^{4} \lambda_i \ln P_i(y|k)$$

Die Segmentierung erfolgt iterativ unter Verwendung einer Maximum–a–posteriori Schätzung (GIMMELFARB & ZALESNY, 1991; TÖNJES, 1994). Abschließend werden die Konturen mithilfe eines Konturmodells basierend auf Gibbs–Zufallsfeldern, das glatte Konturen bevorzugt, geglättet (MESTER & FRANKE, 1988). Abbildung 3b zeigt die segmentierten Wälder.

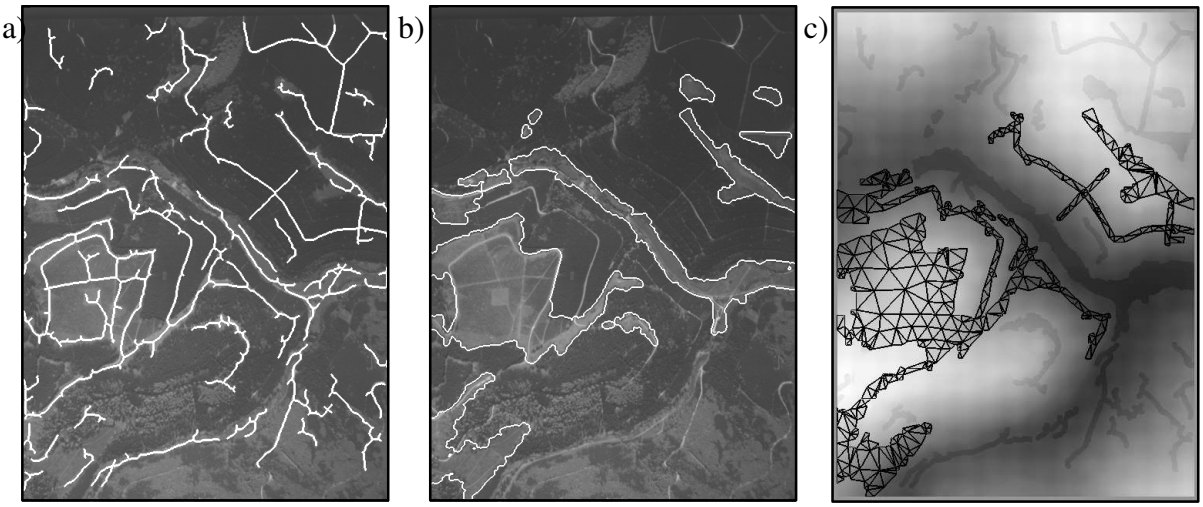

Abb. 3: a) Luftbild mit überlagerten Segmentierungsergebnissen für a) Straßen und b) Grenzen zwischen Wäldern und Grünland. c) Höhenkarte mit erhabenen Waldgebieten und überlagerten Teilnetzen für ausgewählte Objekte aus a) und b).

4.3 Rekonstruktion

Die wissensbasierte Modellierung nutzt Vorwissen über Objekte für deren geometrische Rekonstruktion. Während datengetriebene Modellierungsansätze nur wenige und allgemeine geometrischen Randbedingungen (Constraints) nutzen, wie beispielsweise die Annahme einer stückweise kontinuierlich verlaufenden Objektoberfläche, ermöglicht eine Interpretation die Verwendung sehr spezifischer Objekteigenschaften. Die Interpretation liefert die Segmentierung der Luftbilder in verschiedene Regionen, wie Wälder, Wiesen und Straßen. Die Szene kann durch Objektmasken beschrieben werden. Nachfolgend nutzt das Modul zur Oberflächenrekonstruktion diese Information, um mithilfe der durch Stereokorrespondenzanalyse gewonnenen Tiefenkarten die Geometrie nachzubilden. An Übergängen von Wäldern auf Wiesen wird ein Höhensprung im Modell erzwungen und die nicht sichtbare Waldkante eingefügt. Straßen werden als separate Teilnetze mit einem Kontinuitätskriterium modelliert, was bewirkt, daß sie nicht zu starke Knicke enthalten und damit für den Betrachter natürlicher wirken.

Die Semantik der einzelnen Modellteile ermöglicht zudem eine objektspezifische Nachbearbeitung des Modells. Objekte können um künstliche Details aus Computer Graphik Bibliotheken ergänzt werden, die in den Sensordaten nicht sichtbar sind. So können beispielsweise für Nahansichten Baumreihen mit detaillierter Blattstruktur vor Waldkanten positioniert werden (Abb. 4).

5. Visualisierung auf einem Simulator

Für eine effiziente Visualisierung des 3D–Modells auf einem Simulator wird die gewonnene Tiefeninformation des Modells abgetastet und durch ein räumliches Dreiecksnetz approximiert. Für die Darstellung der 3D–Modelle auf einem Simulator müssen dessen spezielle Hardware–Eigenschaften genutzt werden, um ein Maximum an Performance zu erzielen. Da die meisten Graphik Pipelines zur Visualisierung von Polygonen eine Triangulierung durchführen, eignen sich insbesondere Dreiecksnetze. Größere zusammenhängende Flächenstücke sind nach außen hin zwar Polygone, setzen sich aber ebenfalls aus Dreiecken zusammen. Die Feinstruktur von Objekten wird nicht mehr als Geometrie modelliert, sondern als Phototextur gespeichert. Zu diesem Zweck werden die Eingangsbilder auf das Netz projiziert und in einer Texturemap als Phototextur gespeichert.

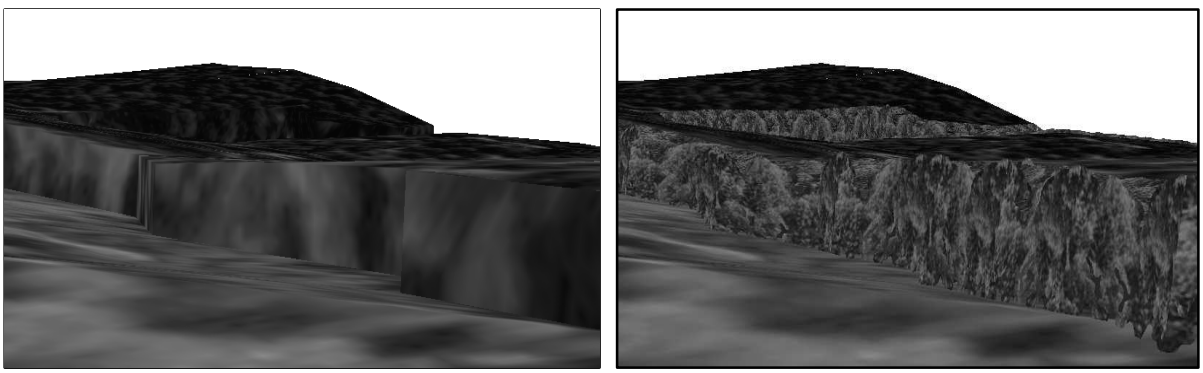

Abb. 4: Waldkante a) ohne und b) mit synthetisierten Baumstrukturen des höherem Level of Detail

Die während der Interpretation durchgeführte Zerlegung der Szene in einzelne Objekte und Objektteile wird beibehalten und an das Visualisierungssystem übergeben. Dies ermöglicht eine ef-

fiziente Vorselektion von Objektteilen, die nicht sichtbar sein können, weil sie beispielsweise nicht im Blickwinkel der (virtuellen) Kamera liegen (Culling). Das Culling ist notwendig, um die Graphikhardware nicht zu überfordern und eine möglichst hohe Framerate zu erreichen.

Eine weitere Methode zur Verringerung der Szenenkomplexität während der Bildsynthese ist der Einsatz von Objekten mit abgestufter Komplexität (Levels of Detail). So kann beispielsweise ein Waldrand, der in der virtuellen Welt (im Rechner) weit entfernt ist und somit in der Abbildung sehr klein erscheint, mit nur sehr wenigen Polygonen dargestellt werden, ohne unnatürlich zu wirken. Erst wenn der Abstand zur virtuellen Kamera kleiner wird, muß das Modell detaillierter dargestellt werden, einen höheren 'Level of Detail' erhalten. Bild 4 zeigt zwei Detailierungsgrade einer Waldkante.

Als sehr effiziente Möglichkeit zur Modellierung von Feinstrukturen können den Bildelementen Transparenzwerte zugeordnet werden (Alpha–Kanal). So kann beispielsweise Blattwerk von Bäumen nicht als eine Vielzahl von Einzelblättern modelliert werden, sondern muß gruppenweise in einem Dreieck zusammengefaßt werden. Ein zugeordneter Alpha–Kanal sorgt für eine Durchlässigkeit an Stellen, die keinem Blatt entsprechen, so daß mit geringem Aufwand an Geometrie viel Struktur (Blätter) nachgeahmt wird.

Für die Archivierung der Modelle stehen verschiedene Standarddateiformate zur Verfügung. Bei dem hier verwendeten Simulator, einer Silicon Graphics (Onyx RealityEngine^2), wird das Inventor–Dateiformat verwendet. Beim Laden der Modelle konvertiert ein Parser die Modelle in eine Darstellung zur Echtzeitvisualisierung. Die Software ist mit der Virtual–Reality–Software Performer erstellt. Dieses Paket sieht eine Reihe von Möglichkeiten zum Aufbau eines Simulators vor, wie Culling, Level of Detail, verschiedene Bewegungsmodelle, Kollisionsüberprüfung und so weiter.

6. Ergebnisse

Das vorgestellte System nutzt a priori Wissen über die Szene, um zu realistischeren Modellen zu gelangen. Der Vorteil des Systems ist, daß durch das Vorwissen die Modellparameter eingeschränkt und objektspezifische Oberflächenprimitive ausgewählt werden können. Verdeckte Modellteile und aufgrund der endlichen Sensorauflösung nicht erfaßte Details, sind a posteriori ergänzbar, um das Modell zu vervollständigen. Die Modellierung berücksichtigt insbesondere das, was für einen realistischen Eindruck auf einen menschlichen Betrachter wichtig ist, wie ebene Straßen und Höhensprünge an Waldkanten.

Bild 5 zeigt die synthetisierte Ansicht des aus einem Luftbildpaar gewonnenen Landschaftsmodells. Das Modell wurde unter Berücksichtigung der Objektsemantik modelliert: Die Straßen werden durch separate Teilnetze repräsentiert und weisen einen kontinuierlichen Verlauf auf. An den Waldrändern wurde ein Höhensprung eingefügt. Das für die Simulation verwendete Modell des Siebertales im Harz (ca. 2x2 km^2) umfaßt ca. 13000 Polygone und eine Texture Map der Größe 2048x2048 Pixel.

Abb. 5: Synthetisierte Ansicht des Landschaftsmodells.

7. Literatur

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.

Clement, V., Giraudon, G., Houzelle, S. , Sadakly, 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.

Foresti, G.L., Murino, V., Regazzoni, C.S., Vernazza, G., "Distributed spatial reasoning for multisensory image interpretation", *Signal Processing*, 32, pp. 217–255, 1993.

Gimmelfarb, G. L., Zalesny, A. V., "Low–Level Bayesian Segmentation of Piecewise–Homogenous Noisy and Textured Images", *International Journal of Imaging Systems and technology*, Vol. 3, pp. 227–243, 1991.

Grau, O., "Ein Szeneninterpretationssystem zur Modellierung dreidimensionaler Körper", *17. DAGM–Symposium Mustererkennung 1995*, 13.–15. Sep. 1995, Bielefeld.

Matsuyama, T., Hwang, V.S.–S., "SIGMA : A Knowledge–Based Aerial Image Understanding System", *Plenum Press*, New York 1990

McKeown, et al., "Rule–Based Interpretation of Aerial Imagery", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. PAMI–7, No. 5, pp. 570–585, Sept. 1985.

Mester, R., Franke, U., "Statistical model based image segmentation using region growing, contour relaxation and classification", *Proc. of SPIE Conf. on Visual Communication and Image Processing*, Cambridge Mass., Vol. 1001, November 1988.

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.

Shapiro, S. C. (Ed.), "Encyclopedia of Artificial Intelligence", *John Wiley & Sons*, New York, 1992.

Tönjes, R., "Realistic Landscape Modelling with High Level of Detail", *IEEE International Conference on Image Processing*, Austin, Texas, 13–16 November 1994.