IDENTIFIKATION UND VERFOLGEN VON OBJEKTEN ANHAND NICHT-PERFEKTER KONTUREN

Bernd Neumann
Universitaet Hamburg

ZUSAMMENFASSUNG

Es wird ueber ein Szenenanalyse-System berichtet, das Position und Identitaet von Objekten in Fernsehkamera-Aufnahmen ermitteln kann. Das System wurde im Hinblick auf moegliche Anwendungen im industriellen Bereich entworfen. Es erkennt Objekte auch bei gestoerten Verhaeltnissen (wechselnder Beleuchtung, teilweiser Verdeckung). Die Identifikation erfolgt mithilfe von Objektmodellen, in denen die Kantenverlaeufe von Objektansichten abgelegt sind. Neue Objektmodelle koennen durch interaktives Auswerten einer typischen Aufnahme auf einfache Weise erzeugt werden. Die zugrundeliegende Systemstruktur und die verwendeten Verfahren werden im Hinblick auf die beabsichtigten Anwendungsmoeglichkeiten diskutiert. Es werden Ergebnisse vorgefuehrt von Szenen mit ungeordnet uebereinanderliegenden Kleinteilen und einer Szenensequenz, in der ein Objekt auf einem simulierten Fliessband verfolgt wird.

1. EINLEITUNG

Auf dem Gebiet der Szenenanalyse sind in den letzten Jahren viele Probleme in Angriff genommen worden, die weit ueber einfache Erkennungsaufgaben hinausgehen. Das Hauptziel ist die Analyse "natuerlicher Szenen", mit Konfigurationen komplexer Objekte und unter Bedingungen, wie sie ein menschlicher Beobachter normalerweise vorfindet. Als Beispiele seien das VISION-Projekt von Hanson/Riseman [3] und Nagels Ansaetze zur Analyse von Szenenfolgen mit Bewegung genannt [6]. Vergleicht man solche Zielsetzungen mit einem Mustererkennungsproblem der ersten Generation, etwa der Druckzeichenerkennung, so sieht man, dass an drei grosse Problemkreise erhoehte Anforderungen gestellt werden muessen: Segmentierung (Aufgliedern eines Rohbildes, etwa in Kanten oder Bereiche), Repraesentation von Wissen (Organisation und Strukturierung des zur Analyse erforderlichen Vorwissens und des Analyseergebnisses) und Interpretation (Bedeutungszuweisung aufgrund von Segmentation und Vorwissen). Viele Arbeiten der Szenenanalyse lassen sich als Beitraege zu einem dieser drei

Aus: "Bildverarbeitung u. Mustererkennung", Proceedings des 1. DAGM Symposium, Oktober 1978,

Gebiete verstehen, und einige Fortschritte sind erzielt worden, besonders bei Wissensstrukturierung und Interpretationsstrategien.

In der vorliegenden Arbeit wird versucht, diese Fortschritte fuer ein bescheideneres Ziel zu verwerten, naemlich fuer die Objekterkennung im industriellen Bereich. Es werden Szenen betrachtet, bei denen es auf das Erkennen, Lokalisieren oder Verfolgen einer beschraenkten Anzahl von einfachen Objekten unter definierten Sichtbedingungen ankommt. Verschiedene Gesichtspunkte machen dieses Problem interessant. Zum einen handelt es sich um eine Aufgabe, fuer die es sofort verschiedene Anwendungen geben koennte, etwa die Steuerung von Manipulatoren. Zum Zweiten koennen neue Konzepte und Verfahren der Szenenanalyse in einem ueberschaubaren System zusammengeschlossen und erprobt werden. Schliesslich gilt es, die fuer eine solche Aufgabe spezifischen Probleme aufzudecken und zu loesen.

Es wird ein implementiertes System beschrieben, das Objekte anhand ihrer Konturen erkennt, also anhand der Kanten, die sich in einer Objektansicht abzeichnen. Objekte koennen teilweise verdeckt sein, z.B. ungeordnet uebereinander liegen. Ferner koennen verschieden Ansichten eines 3-dimensionalen Koerpers erkannt und unterschieden werden, etwa zur Steuerung eines Fliessbandzugriffs. Der Analyseprozess kann durch Kontextwissen gesteuert und beschleunigt werden, z.B. bei der Verfolgung bewegter Objekte.

Beim Aufbau des Systems stand die Einsicht Pate, dass Operationen auf der Pixelebene auch bei groesster Sorgfalt nicht zu einer perfekten Zerlegung einer Szene in Objekte fuehren [5]. Deshalb ist das System darauf angelegt, fehlerhafte Segmentierungen interpretieren zu koennen. Dies hat einige Vorteile. Zum einen wird das System bei Segmentierungsfehlern keine Zufallsergebnisse liefern, da es auf Fehler eingestellt ist ("graceful degradation"). Zum Zweiten kann man bei einem System mit einfacher Segmentierung aber dafuer aufwendiger Interpretation einen insgesamt reduzierten Rechenaufwand erwarten, da die Datenvolumina in hoeheren Verarbeitungsstufen meist geringer sind. Zum Dritten kann das System auch mit einem gewissen Mass von systematischen "Segmentierungsstoerungen" fertig werden, etwa mit teilweiser Verdeckung oder perspektivischer Verzerrung.

Das hier verwendete Segmentierungsverfahren extrahiert aus der Szene gerade Kanten als Naeherung von Objektkonturen. Der Algorithmus wird in Kapitel 2 beschrieben. Objekte sind dem System durch eine relationale Modelldatenbasis bekannt. Relationalstrukturen haben sich bereits bei anderen Szenenanalyseprojekten als Wissensspeicher und fuer Vergleichsoperationen bewaehrt [1,2]. Hier werden Kanten- und Winkelrelationen zur Beschreibung von Objektprototypen

verwendet. Struktur und Aufbau der Modelldatenbasis sind Gegenstand von Kapitel 3.

Der Interpretationsalgorithmus ist in Kapitel 4 beschrieben. Zunaechst werden Hypothesen durch einen 2-stufigen Relationalvergleich generiert und mit einem Konfidenzwert versehen. Zwei feste Schwellwerte entscheiden darueber, ob eine Hypothese sofort akzeptiert, zurueckgewiesen oder als unsicherer Kandidat aufbewahrt wird. Von unsicheren Hypothesen werden nur diejenigen akzeptiert, die einen abschliessenden Filterprozess ueberleben.

In Kapitel 5 werden einige Ergebnisse vorgestellt. Die Szenen sind mit einer Fernsehkamera aufgenommen und schliessen zwei Anwendungsfaelle ein, den "Griff-in-die-Kiste" und das Verfolgen eines Objektes auf einem Fliessband. Die Diskussion in Kapitel 6 befasst sich mit der Moeglichkeit, ein solches System praktisch einzusetzen.

2. SEGMENTIERUNG

Das hier verwendete Segmentierungsverfahren ermittelt Kanten als Grundlage fuer den nachfolgenden Interpretationsprozess. Da nicht erwartet wird, dass die Objekte vollstaendig und genau beschrieben werden, erzeugt der Algorithmus von vornherein nur geradlinige Annaeherungen und versucht auch nicht, Kantenzuege zu geschlossenen Konturen zu ergaenzen.

Zunaechst wird die urspruengliche Grauwertmatrix (574x512x8 Bit) in eine Gradientenmatrix (191x256x5 Bit) umgeformt. Die Gradientenrichtung ist mit 4 Bit und der Betrag (nach einem Schwellwertvergleich) mit 1 Bit kodiert. Im zweiten Schritt werden kollineare Kantenelemente zu "Strichen" verbunden, wenn ihre Richtungen sich um hoechstens eine Quantisierungseinheit (22.5 Grad) unterscheiden. Dies geschieht in 8 Abtastgaengen, bei denen die Gradientenmatrix jeweils vollstaendig unter einem bestimmten Winkel abgetastet wird. Da Kanten auch gekruemmt sind und in der Regel nicht genau mit einer der 8 Abtastrichtungen zusammenfallen, entstehen Striche, die sich seitlich versetzt fortsetzen. Sie werden verschmolzen, wenn sie eine genuegend gerade kante ergeben; andernfalls bleiben sie isoliert. Fig. 1 zeigt die Gradientenrepraesentation einer dunklen Ecke (die Richtungen sind hexagonal kodiert). Die Striche (durchgezogene Linien) wurden durch Abtasten in der 22.5- bzw. 90-Grad-Richtung erzeugt. Der nachfolgende Verschmelzungsprozess verband die vertikalen Striche zu einer Kante (gestrichelt) und eliminierte den kuerzeren der beiden schraegen Striche.

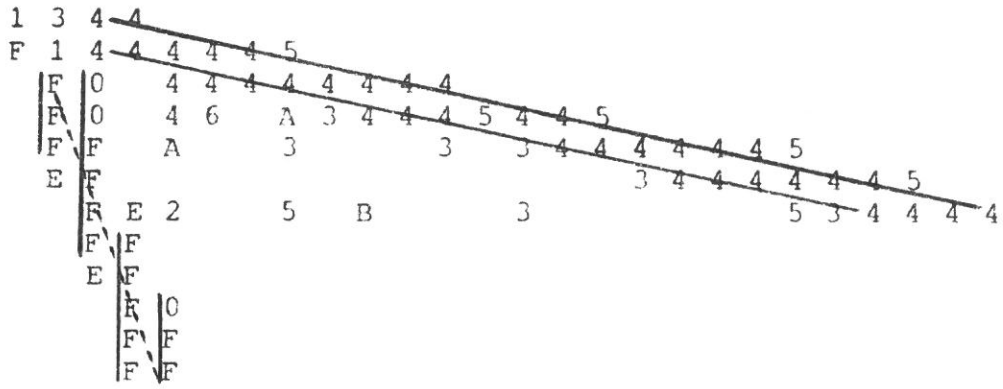

Fig.1: Striche und Kanten in der Gradientenmatrix

3. MODELLE

Wissen ueber Objektformen ist ein einer Modelldatenbasis enthalten. Ein Objektmodell wird durch eine Modellkanten-Relation MKR definiert. Sie gibt an, welche Kanten zu einer bestimmten Objektansicht - gekennzeichnet durch einen Objektnamen - gehoeren. Jede Kante besteht aus einem Kennzeichner, sowie Anfangs- und Endpunktkoordinaten. Kanten sind gerichtet in dem Sinne, dass die dunklere Seite zur Rechten ist. Eine zweite Relation MWR assoziiert zusammenhaengende Kanten mit dem eingeschlossenen Winkel. Dies stellt keine zusaetzliche Information dar, sondern dient nur zur Unterstuetzung der Hypothesenbildung. Die Kanten- und Winkelrelationen lauten formal:

```
MKR = [ < o k <x1 y1> <x2 y2> > ]
MWR = [ < w k1 k2 > ]
```

Dabei bedeuten

o	Objektname
k,k1,k2	Kantenkennzeichner
w	Winkel
x1,x1	x-Koordinaten in der Gradientenmatrix
y1,y2	y-Koordinaten in der Gradientenmatrix

Jedes Modell beschreibt eine "typische" Objektansicht. Es dient dazu, alle Objekte zu erkennen, deren Ansichten im geometrischen Sinne aehnlich sind. Eine Menge von Szenenkanten wird also als Objekt o identifiziert, wenn es eine Translation, Rotation und Dilatation gibt, die die Modellkanten mit den Szenenkanten zur Deckung bringen. Natuerlich brauchen Modell und Wirklichkeit nicht exakt uebereinzustimmen. Eine Modellkante kann durch mehrere Szenenkanten repraesentiert werden, und umgedreht. Der Grad der Uebereinstimmung wird durch eine Konfidenzfunktion

gemessen, die im Wesentlichen den Quotienten aus tatsaechlich sichtbarer und idealerweise vorhandener Kantenlaenge ausgibt.

Neue Modelle koennen auf einfache Weise dadurch erzeugt werden, dass man das Objekt unter ungestoerten Bedingungen mit der Kamera aufnimmt und die Kanten mit dem Segmentierer des Systems extrahiert. Ein interaktiver Modelleditor gestattet kleinere Korrekturen und erzeugt aus der Kantenkollektion die oben beschriebene Relationalstruktur. Modelldateien koennen durch einfaches Konkatenieren erweitert werden.

4. INTERPRETATION

Nach der Segmentierung wird eine Szene nur noch durch die extrahierten Kanten repraesentiert. In der Interpretationsphase wird nun versucht, moeglichst vielen Kanten eine Bedeutung zuzuweisen. Genauer gesagt: Kanten werden mit einem Objektnamen o und den Translations-, Rotations- und Dilatationsparametern t, r, d der erforderlichen Modelltransformation assoziiert. Ein Tupel <o t r d> heisst Hypothese. Eine Interpretation verknuepft Szenenkanten mit Hypothesen und ist definiert als Relation [< k o t r d >]. Einzelne Hypothesen koennen auf natuerliche Weise durch Vergleich von Modell- und Szenenkanten bewertet werden. Es ist jedoch nicht offensichtlich, welche Interpretation insgesamt den hoechsten Konfidenzwert verdient. Dies Problem ist nicht neu und wird haeufig dadurch geloest, dass zusaetzliches Wissen ueber moegliche Objektkonfigurationen, Beziehungen zwischen Objekten, etc. herangezogen wird. Formuliert man dieses Wissen als Zwangsbedingungen, so lassen sich Relaxationstechniken zur Ermittlung der besten Interpretation verwenden [9]. In dem hier beschriebenen System wird gefordert, dass eine Szenenkante nur eine Bedeutung haben kann, also nur Teil eines Objektes ist. Statt Relaxation im ueblichen Sinne wird eine schnelle "best-first" Variante eingesetzt.

Im Folgenden wird zunaechst ueber Hypothesenbildung berichtet. Szenenkanten werden in derselben Form wie Modellkanten abgespeichert, mit Ausnahme der fehlenden Objektnamen. Es werden also sowohl eine Szenenkanten-Relation SKR aufgebaut als auch eine Szenenwinkel-Relation SWR, in der alle Szenenkanten eingetragen sind, deren Anfangs- und Endpunkte genuegend nahe sind. Falls kein Kontextwissen ueber bevorzugte Hypothesen vorhanden ist, werden die Tupel von SWR der Reihe nach (und zwar die laengsten Kanten zuerst) auf eine Winkeluebereinstimmung mit Tupeln von MWR hin untersucht. Jede Uebereinstimmung bestimmt Objektnamen o und Rotation r

einer Hypothese. Translation t und Dilatation d werden mithilfe der Winkelpaare und einer dritten Szenenkante berechnet, die mit einer der verbleibenden Modellkanten uebereinstimmen muss. Die Parameter t,r,d haengen nur von der Lage von Szenenkanten, jedoch nicht von ihrer Laenge ab.

An dieser Stelle wird das 2-stufige bottom-up Verfahren invertiert, und ein top-down Vergleich von Szenenkanten und transformierten Modellkanten findet statt. Daraus ergibt sich ein Konfidenzwert fuer die Hypothese <o t r d>. Zwei Schwellwerte entscheiden darueber, wie die Hypothese weiter behandelt wird. Ist der Konfidenzwert groesser als der hoehere Schwellwert, so wird die Hypothese sofort akzeptiert, und die dazugehoerigen Szenenkanten werden entfernt. Ist er kleiner als die niedrigere Schwelle, so ist die Hypothese nicht akzeptabel und wird verworfen. Bei Konfidenzwerten im Mittelbereich wird angenommen, dass fuer diese Szenenkanten womoeglich noch bessere Hypothesen generiert werden. Die vorliegende Hypothese ist akzeptabel aber unsicher und wird in einer Liste aufbewahrt. Nachdem alle Hypothesen generiert sind, wird die Liste mit der folgenden Prozedur solange bearbeitet, wie sie akzeptable Hypothesen enthaelt.

(i) Akzeptiere die beste Hypothese und entferne sie aus der Liste.

(ii) Entferne zugehoerige Szenenkanten und verringere die Konfidenzwerte der verbleibenden Hypothesen entsprechend.

Das Verfahren hat die Tendenz, eine Interpretation mit maximaler Summenkonfidenz zu erzeugen. Die beiden Schwellwerte beeinflussen das Ergebnis nur unerheblich. Ihr hauptsaechlicher Effekt betrifft den Rechenaufwand, der um so kleiner wird, je mehr Hypothesen sofort akzeptiert werden.

5. ERGEBNISSE

Das System hat viele Testszenen analysiert [7]. Bild 1a zeigt eine Szene mit allerlei Werkzeug. In 1b sind die zur Interpretation benutzten Modelle abgebildet, und in 1c sieht man die gefundenen Kanten (weiss) ueberlagert vom Interpretationsergebnis (schwarz).

Die Bilder 2a-c illustrieren den Anwendungsfall "Griff-in-die-Kiste". 2a zeigt eine Beispielsszene mit 5 uebereinanderliegenden fiktiven Stanzteilen (aus Pappe). Die zugehoerigen Modelle, je eines fuer Ober- und Unterseite, sind in 2b abgebildet. Segmentierung und Interpretation sieht man in 2c. Die Rechenzeit betrug 50

sek fuer die Segmentierung und 7 sek fuer die Interpretation auf einem DECsystem-10 mit KI-Prozessor. Alle Programme sind in SAIL geschrieben, einer ALGOL-artigen Programmiersprache fuer Anwendungen der "Kuenstlichen Intelligenz".

Die Bilder 3a-f zeigen eine "Objektverfolgung". Es handelt sich dabei um ein Gehaeuse, das auf einem simulierten Fliessband bewegt wird, und dessen Lage zu jedem Zeitpunkt bestimmt werden soll. Von der 5 Szenen umfassenden Sequenz zeigen 3a-c die erste, dritte und letzte. Die gesamte Sequenz kann mit nur einem Modell (3d) interpretiert werden, obwohl sich die Ansichten durch perspektivische Verzerrung betraechtlich unterscheiden. 3e und 3f zeigen Segmentierung und ueberlagerte Interpretation der ersten bzw. letzten Szene. Die Interpretationszeit liegt zwischen 3 und 11 sek pro Szene, wenn kein Kontextwissen beruecksichtigt wird. Durch Extrapolation aus der vorhergehenden Szene kann der Suchraum jedoch eingeengt und die Rechenzeit bis um 60% verringert werden. Dies ist ein bescheidener Gewinn, da sowieso nur 1 Modell zu betrachten ist. Bei Interpretationsaufgaben mit zahlreichen Modellen kann Kontextinformation zu viel deutlicheren Zeitgewinnen fuehren.

6. DISKUSSION

Das hier beschriebene System war mit der Zielsetzung entwickelt worden, Methoden und Konzepte der Szenenanalyse fuer Objekterkennung im industriellen Bereich nutzbar zu machen. Als positive Eigenschaften sind hervorzuheben

- Unempfindlichkeit gegen gestoerte Ansichten,
- vielseitige Anwendbarkeit,
- einfache Modellgewinnung.

Diese Eigenschaften sind hauptsaechlich durch das sorgfaeltig ueberlegte Interpretationsverfahren bedingt und auch durch das Konzept, 3-dimensionale Koerper mithilfe von Modellen ihrer typischen Ansichten zu beschreiben.

Demgegenueber steht eine Rechenzeit im Minutenbereich, die fuer die meisten industriellen Anwendungen wohl um den Faktor 60 zu hoch ist. Der entscheidende Engpass ist die Segmentierung mit ca. 50 sek pro Bild. Hier koennen durch besondere Hardware jedoch auch am leichtesten Einsparungen erzielt werden. Als Loesung werden ein einfacher Gradientenbaustein und die Ermittlung von Kanten durch 8 Parallelprozessoren vorgeschlagen. Dadurch duerften Zeiten im Sekundenbereich moeglich werden.

Das Problem der Segmentierungszeit scheint grundsaetzlicher

Natur zu sein und zeigt sich auch darin, dass fast alle kommerziell entwickelte Verfahren zur Objekterkennung nur einfache Schwellwertoperatoren benutzen und entsprechend eingeschraenkt nutzbar sind, z.B. [4]. Die Arbeit von Perkins ist eine Ausnahme [8]. Er benutzt einen Kantenoperator und interpretiert Szenen mit Maschinenteilen in ca. 15 sek auf einer IBM 370/165. Der Autor dieser Arbeit zieht den Schluss, dass entscheidende Fortschritte bei der industriellen Objekterkennung nur mit Unterstuetzung durch spezielle Hardware zu erzielen sind.

Apparatur und unterstuetzende Programme fuer diese Arbeit gehen auf gemeinsame Anstrengungen von R.Bertelsmeier, P.Cord, I.Heer, H.Kemen, H.-H.Nagel, B.Neumann und B.Radig zurueck.

7. LITERATURVERZEICHNIS

[1] H.G.Barrow et al., "Some Techniques for Recognizing Structures in Pictures", in: Frontiers of Pattern Recognition (Watanabe, Hrsg.), S.1, 1972

[2] H.G.Barrow, R.J.Popplestone, "Relational Descriptions in Picture Processing", in: Machine Intelligence VI (Meltzer/Michie, Hrsg.), S.377, 1971

[3] A.Hanson, E.Riseman (Hrsg.), "Computer Vision Systems", Academic Press, N.Y., 1978

[4] R.Karg, "A Flexible Opto-Electronic Sensor", in: Proc. 8th Int. Symp. on Industr. Robots, Deutschland, Boeblingen, 1978

[5] D.Marr, "On the Purpose of Low-Level Vision", AI-Memo 324, MIT, Cambridge, 1974

[6] H.-H.Nagel, "Analysing Sequences of TV-Frames: System Design Considerations", IJCAI-77, S.626, 1977

[7] B.Neumann, "Identifikation von gestoerten Objektansichten unter Verwendung geradliniger Konturapproximationen", FBI-HH-B-42/78, Fachbereich Informatik, Universitaet Hamburg, 1978

[8] W.A.Perkins, "A Model-Based Vision System for Scenes Containing Multiple Parts", Proc. IJCAI-77, S.678, 1977

[9] A.Rosenfeld et al., "Scene Labelling by Relaxation Operations", IEEE Trans. Systems, Man and Cybernetics, SMC-6, S.420, 1976

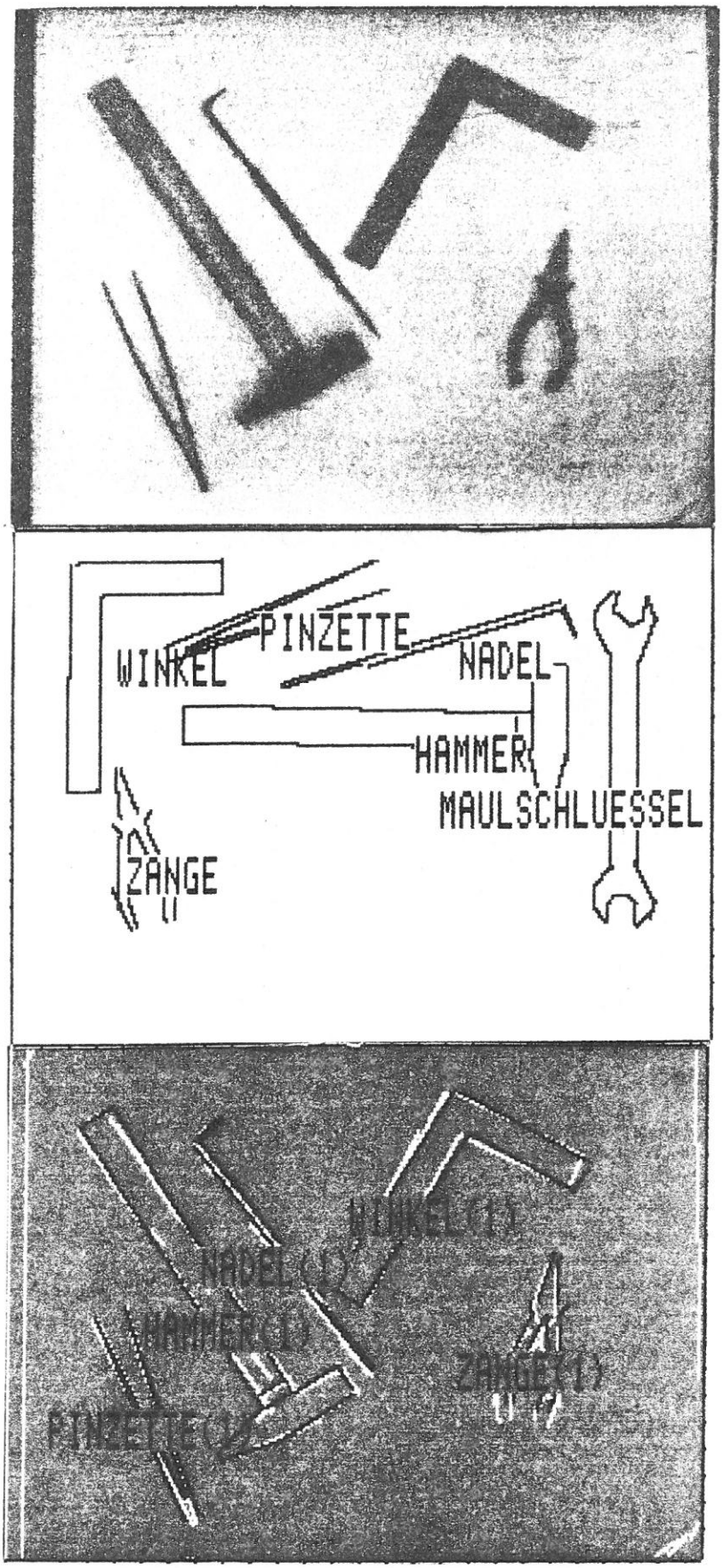

Bild 1: a) Werkzeug-Szene
b) Modelle
c) Segmentierung und Interpretation

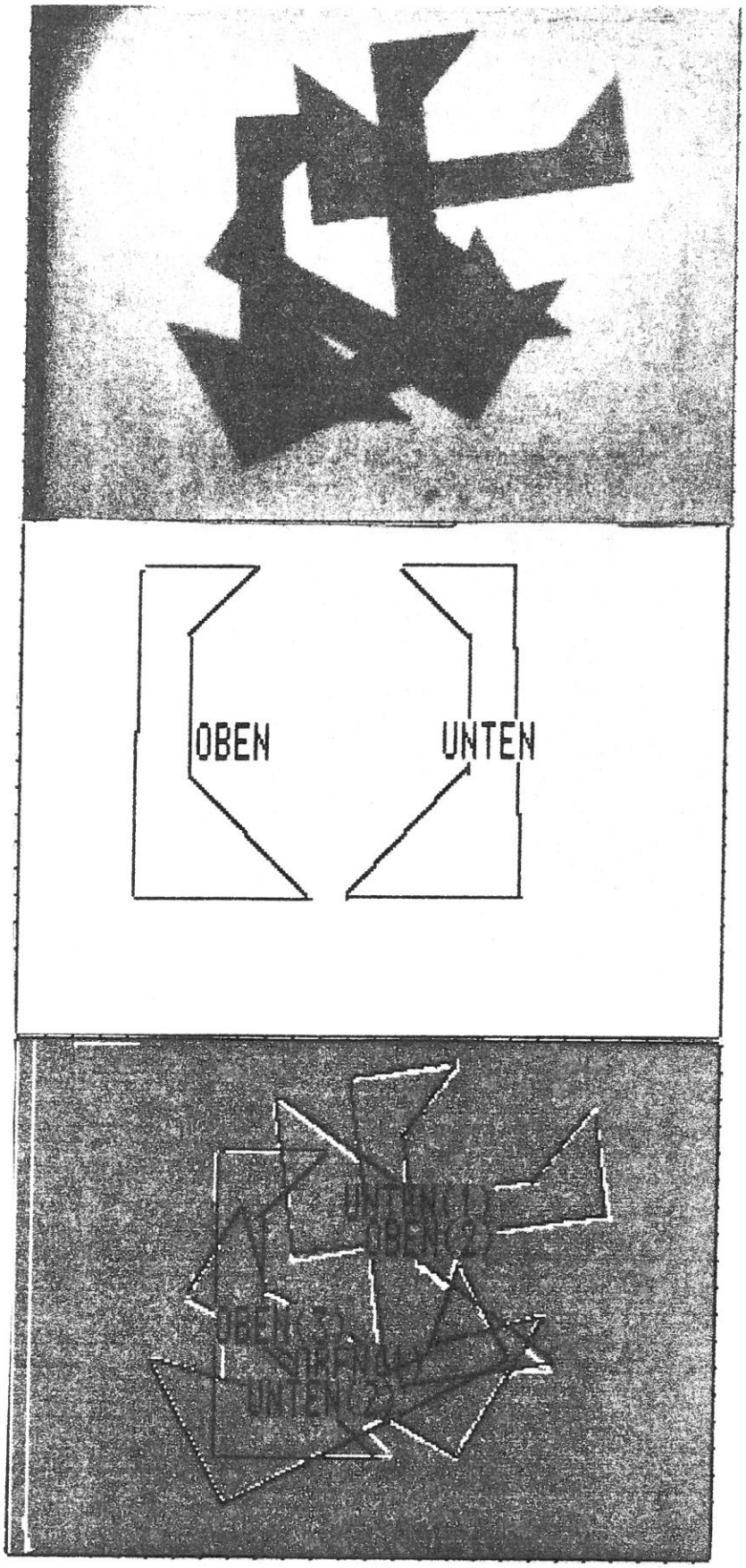

Bild 2: a) Szene mit 5 Stanzteilen
b) Modelle für Ober- und Unterseite
c) Segmentierung und Interpretation

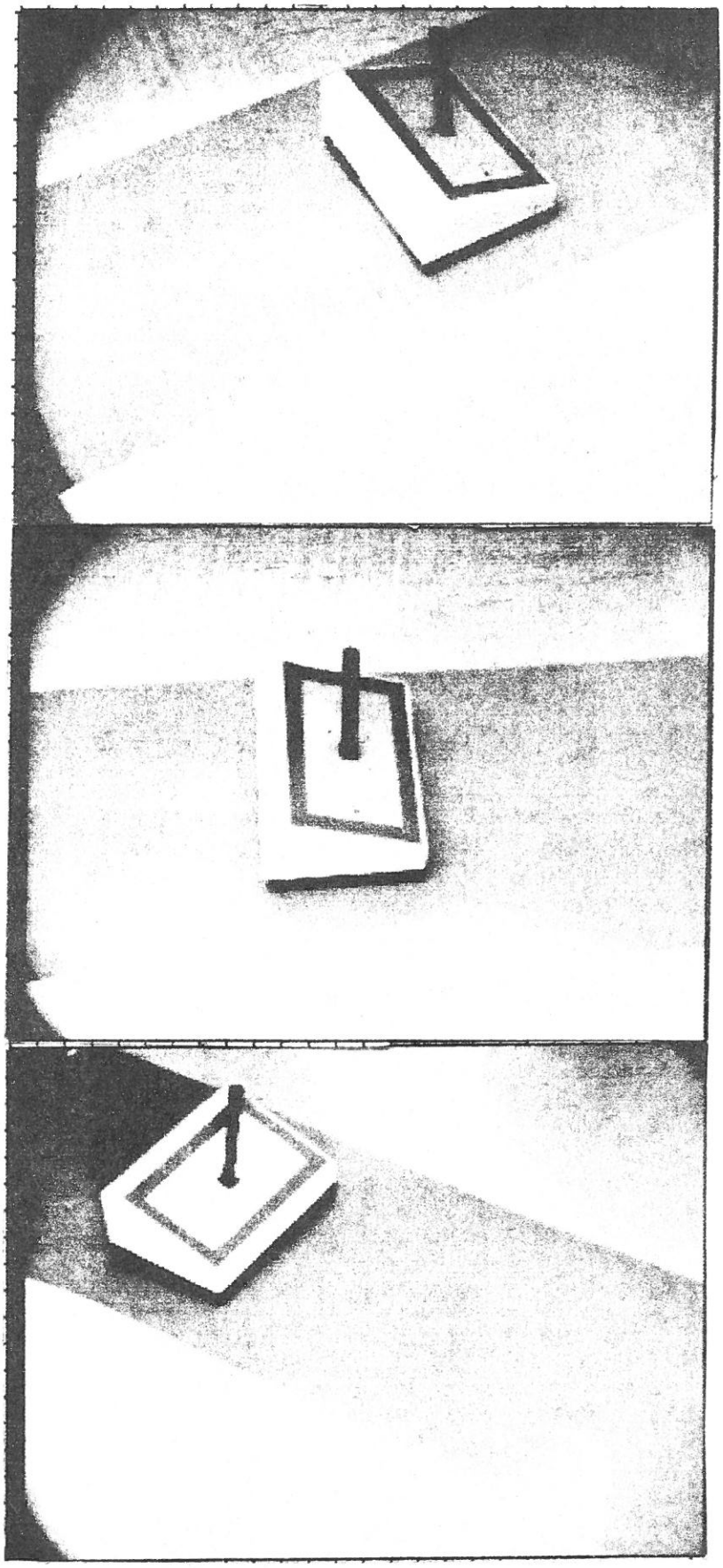

Bild 3: a) 1. Ansicht von "Gehäuse auf Fließband"
b) 3. Ansicht
c) 5. Ansicht

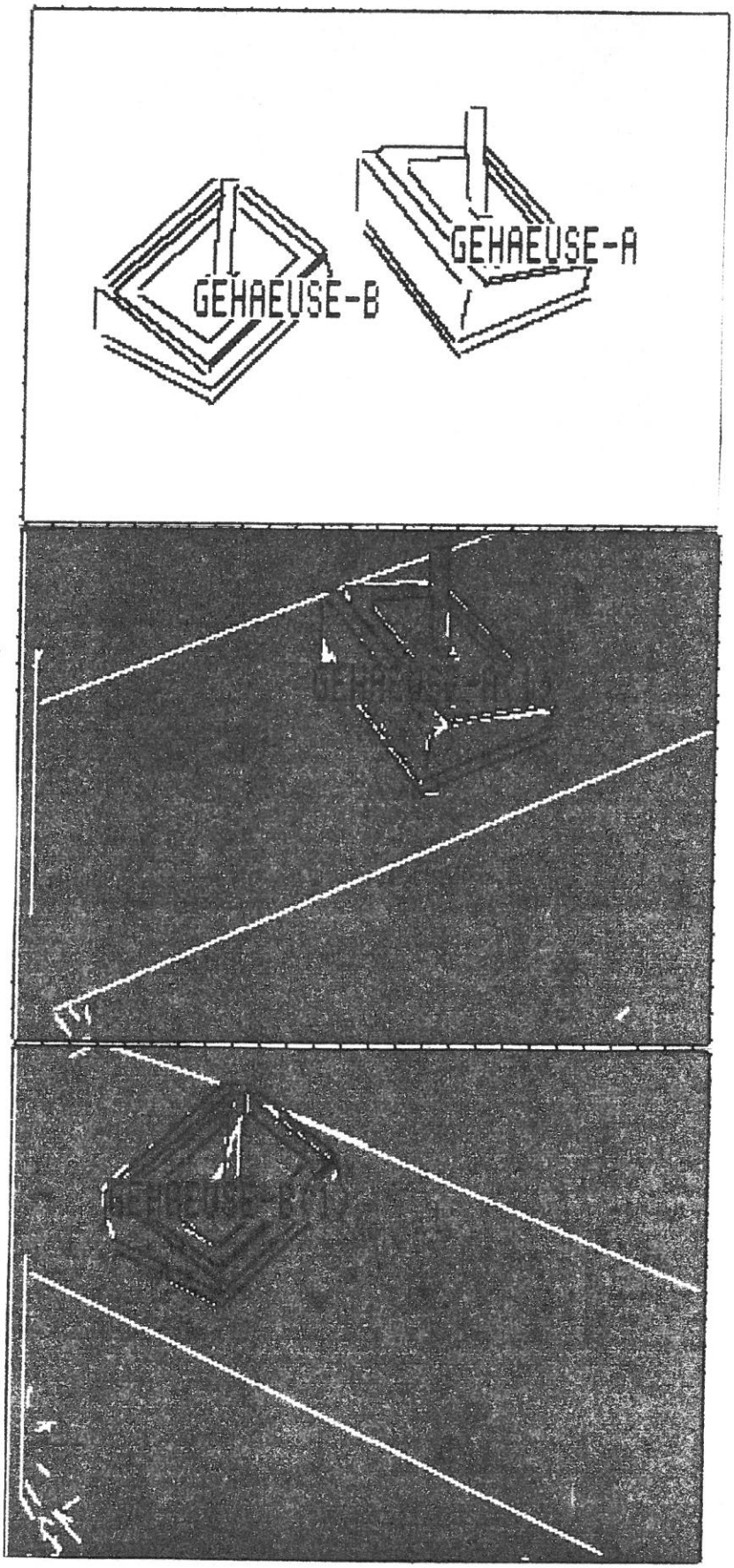

Bild 3: e) Modelle für 2 typische Ansichten
f) Segmentierung und Interpretation von 3a
g) Segmentierung und Interpretation von 3c