

Diss. ETH ex. B

Diss. ETH No. 10140

Contributions to a 3-D Robot Vision System: Silhouette Analysis and Evidential Reasoning

A dissertation submitted to the
SWISS FEDERAL INSTITUTE OF TECHNOLOGY
ZURICH

for the degree of
Doctor of Technical Sciences

presented by
MARTIN PETER
Dipl. Inf. Ing. ETH Zürich
born February 27, 1957
citizen of Thalwil Kanton Zürich

accepted on the recommendation of
Prof. Dr. O. Kübler, examiner
Prof. Dr. H. Bunke, co-examiner

1993



Gate

Abstract

This contribution originates in the framework of the interdisciplinary, application oriented project “Cooperating Robot” (COR). It contributes to the realization of a robot vision system for the task defined in the project: to build a robot capable of recognizing and clearing objects placed on a tray in a cafeteria. It addresses problems of acquiring 3-D information of transparent and specularly reflecting objects, robust and fast multi-knowledge source recognition scheme and precise object position determination.

Silhouettes as data source are used not only because of the transparent and specularly reflecting materials the objects are made of, but also because of the desired speed of operation of the system. It is demonstrated, how a gray level image of a silhouette can be segmented into a list of approximating contour segments. With the help of relations between these segments, a graph structure is obtained as symbolic representation of a silhouette.

Subsequently the method of graph matching for object recognition well known for range data is presented and applied to graphs obtained from silhouettes. This leads to an implementation of a small recognition system capable of recognizing objects in a scene composed of multiple objects from a known set of 11 objects like cup, glass, plate etc. Ambiguities in segmentation but also the small structural difference between some of the graphs of the objects from the dish set diminish the usability of this method. Therefore, the entirely different method of evidence accumulation for object recognition is introduced. It is based on a completely unbiased use of features like an observed height or a found diameter of a circle. For each feature an independent rule is formulated that states how much the given feature votes for or against a certain object hypothesis. Each object is represented in this rule base by a vector of evidence weights. For recognition, all rules are applied to the features detected

in the scene. This results in a scene vector with each element containing the evaluation result of the corresponding rule. A similarity measure between the scene vector and all object vectors is then computed to identify the object in the scene. This scheme is very efficient and well suited to integrate features from multiple sources. Specially for a 3-D vision system this is of importance since there is no way to compute a 3-D position of an object just from one view. For that reason the COR vision system is built around two silhouettes and one view from above. It is shown, how features from these three views can be integrated to an evidence accumulations scheme and how identity and position of objects can be computed.

The complete system with robot and vision system has been realized and its functioning has been proved. Results of the recognition process obtained during the operational phase of the system are presented.

Zusammenfassung

Die vorliegende Arbeit entstand im Rahmen des interdisziplinären, anwendungsbezogenen Projekts “Kooperierender Roboter” (COR). Sie lieferte einen Beitrag zur Realisierung eines Robot Vision Systems für die im Projekt definierte Aufgabe: einen Roboter zu bauen, der Gegenstände auf einem Mensatblett erkennt und abräumt. Es werden Teilprobleme des Bildinterpretations-System dieses Roboters wie das Beschaffen von dreidimensionalen Daten von Gegenständen, die zum Teil transparent und/oder spiegelnd waren, ein schneller und robuster Erkennungsmechanismus und eine genaue Positionsbestimmung der Gegenstände beschrieben.

Wegen der transparenten und spiegelnden Materialien der Objekte und auch wegen den strengen Laufzeit-Anforderungen werden als Datenquelle Silhouetten der auf den Tablettis platzierten Objekte verwendet. Es wird demonstriert, wie das Grauwert-Bild einer Silhouette in eine Liste von approximierenden Kontursegmenten segmentiert werden kann. Aus dieser erhält man dann mit Hilfe von Relationen zwischen den Segmenten eine Graphenstruktur als symbolische Beschreibung der Silhouette.

Anschliessend wird die von Tiefendaten her bekannte Methode des Graph matching zur Erkennung von Objekten vorgestellt und auf Graphen, die aus Silhouetten erhalten werden, angewendet. Die Implementierung eines einfachen Erkennungssystems für 11 Objekte wie Teller, Tasse, Trinkglas etc. wird beschrieben. Mehrdeutigkeiten der Segmentierung, aber auch die geringen Strukturunterschiede zwischen manchen der Graphen, die von Objekten der Mensatablette erhalten wurden, schränken die Leistungsfähigkeit dieser Methode jedoch stark ein. Als eine grundsätzlich verschiedene Alternative wird daher das Verfahren der Evidenzakkumulation zur Objekterkennung eingeführt. Es beruht auf einem völlig gleichartigen Gebrauch der Merkmale,

etwa einer beobachteten Höhe oder eines gefundenen Kreisdurchmessers. Für jedes dieser Merkmale wird unabhängig von allen andern eine Regel aufgestellt, die für jedes bekannte Objekt besagt, wieviel das Merkmal für oder gegen das betreffende Objekt spricht. Jedes Objekt ist somit in diesem Regelsystem durch einen Vektor mit Evidenzgewichten beschrieben. Für den Erkennungsprozess werden alle Regeln auf die gefundenen Merkmale einer Szene angewendet was als Resultat einen Vektor liefert, dessen Elemente besagen, ob die Vorbedingung der entsprechenden Regel erfüllt oder nicht erfüllt ist. Ein Vergleich dieses Szenenvektors mit allen Objektvektoren liefert schliesslich das Erkennungsergebnis. Dieses Verfahren ist sehr effizient und eignet sich ausgezeichnet für das Verarbeiten von Merkmalen, die von mehreren verschiedenen Quellen abgeleitet werden. Gerade dieser Punkt ist entscheidend für ein operationelles 3-D Vision System, da aus nur einer Ansicht der Szene keine dreidimensionale Positionsbestimmung möglich ist. Für das Vision System des Projekts COR werden daher drei Ansichten, zwei Silhouetten und eine Aufsicht verwendet. Es wird demonstriert, wie die Merkmale aus diesen drei Bildern in einem Evidenz-Akkumulations-Schema integriert werden können und wie damit Identität und Position der Objekte bestimmt werden kann.

Das komplette System mit Roboter- und Vision-System wurde realisiert und seine Funktionstüchtigkeit demonstriert. Resultate vom Erkennungsprozess in der operationellen Phase werden gezeigt.