A New Approach for Multi-object Industrial Scene Analysis

S.Chaudhury*, A. Gupta+, S. Subramanian*, G. Parthasarathy+

* Dept. of Computer Sc. and Engg., + Dept. of Electronics & Electrical Communication Engg., I.I.T, Kharagpur, INDIA.

ABSTRACT

A system based on a new approach for recognition of partially obscured randomly positioned planar objects in a multi-object scene is presented in this paper. The approach is based on abductive reasoning. Interpretation of the image is generated from the observed spatial relations between instantiated primitives. The scheme is general and robust enough to accomodate uncertainties in feature and relation detection. The algorithm is capable of recognising objects with minimum of supportive evidence.

Introduction

Recognition of occluded objects is of prime interest for industrial automation and military applications. Several approaches using local and relational features or constraints[1,2,3] have been proposed. In this paper, we present a new technique for the recognition problem which is general enough to accomodate any kind of feature and relation detectors.

Problem formulation

An object is characterised by the set of primitive features present in it and spatial relations among them. With the assumption that the objects are planar and lies on a plane perpendicular to the camera axis, image of a multi-object scene contains identifiable local primitive image structures having one to one correspondence with the visible subparts of the objects involved. From the primitive image structures, spatial relations among them and knowledge of the object models which can cause their presence, inferences are drawn about the objects present in a scene. This is essentially an inference of the form Given a fact 'B' and an association 'A causes B', infer 'plausible A'.

Since primitive image structures and their spatial relations are invariant under rotation and translation, it forms the basis of our representational scheme. For analysis of images of occluded objects where only partial information about relative configuration of the primitives may be available, the set of observed spatial constraints among the primitives is used to form interpretations. Interpretations are organised as generator sets[4] and at any time it fully explains the partial evidence available at that instant. A suitable updation procedure has been developed to update the interpretations for new evidences.

Interpretation

Let R^+ be the set of all instantiated relations and O^+ be the set of instantiated objects such that

$$O^+ = \bigcup_{i=1}^{m} O_i^+$$

$$O^+_i = \{ O^+_{lj} \mid O_l \in cause^R[R^+_i] \}$$

where, m is the number of instantiated relations, $cause^R(R^+_i)$ is the set of objects possesing the relation R_i and O^+_{lj} is the jth instance of the lth object such that O^+_{lj} is a 4-tuple $(O_l, x_j, y_j, \theta_j)$. (x, y, θ) is the transformation vector calculated for the object O_l from the spatial constraints.

An interpretation of an image is defined as the set of object instances O_I such that

i) $O_I \subseteq O^+$

ii) $R^+ \subseteq man^R(O_I)$ where $man^R(O_I)$ is the set of relations associated with any of the object in O_I

iii) for all $O'_I \subset O_I$, $R^+ \not\subseteq man^R(O'_I)$.

With these three conditions an interpretatin becomes an irredundant set of object instances which explains the presence of all the instantiated relations in the image.

Evaluation of interpretation:

Many interpretations are generated at every stage but the evidential support for each one of them is different. It is necessary to rank the interpretations in order of their evidential support because the control algorithm works in a generate and test manner and at each stage it picks up the strongest interpretation for verification. A suitable evaluation function has been developed which correctly ranks the different interpretations based on their evidential support.

Generation of interpretation:

Interpretations are organised in the form of generator set, as proposed by Reggia et al.[4] Let $s_1, s_2, s_3, \ldots, s_n$ $(n >= 1)$ be pairwise disjoint nonempty subsets of O^+, then $s_I = \{ s_1, s_2, \ldots, s_n \}$ is a generator of alternative interpretations. The interpretations generated by s_I is defined as $[s_I] = \{ O_1^+, O_2^+, \ldots, O_n^+ \} \mid O_i^+ \in s_i \}$. Each s_i corresponds to a disjunction and s_I as a whole corresponds to conjunction of disjunctions. A generator set is simply a set of generators.

Updation of interpretation

Let $s_I = \{ s_1, s_2, \ldots, s_n\}$ be a generator of interpretatins. A new relation R_i^+ is instantiated in the image then s_I should be updated by the set $cause^R(R^+_i)$ (call it set D). The updation operation gives a set of generators.

$s_I/D = \{S_{Ik}$ is a generator $|$ $k = 1$ to $n+1\}$ and each $s_{Ik} = (s_{k1}, s_{k2}, \ldots, s_{km})$ where $m=n$ for $k<=n$ and $m = n + 1$ for $k = n + 1$. The individual s_{ki} are formed as given below.

For k = 1 to n

For i = 1 to n

$$s_{ki} = s_i - D \text{ if } i < k$$
$$= s_i \cap D \text{ if } i = k$$
$$= s_i \text{ if } i > k$$

and for k = n + 1

$$s_{ki} = s_i - D \text{ if } i < n$$
$$= D - \bigcup_{j=1}^{n} s_j \text{ if } j = n + 1$$

This operation , as it is , does not always ensure irredundant interpretations. This problem is taken care of in the interpretation algorithm.

Object modelling

Apriori knowledge of the system consist of explicit description of the models in terms of its primitives and spatial relations. The knowledge base should also model the causal association between objects and spatial relations to facilitate working of the interpretation algorithm. Collection of these relations enables organisation of these informations in a standard relational database. The basic entity sets involved are object labels, primitive image structures and spatial rela tions. However,as the definition of spatial relations is based on certain rules, the info rmations in the entity set of spatial rela tions is not needed and so it is not actually stored.

It is evident from the previous section that the following relations between entity sets are needed for proper functioning of the interpretation algorithm : (i)$cause^R(R_i)$, ie,set of objects associated with the relation R_i, and (ii)$man^F(O_i)$,ie, all primitives, with their control points, associated with the object O_i. We also need the information about total number of relations of type R_i associated with an object O_j and total number of relation instances associated with object O_i for calculation of strength of the interpretations. This information is static in nature and it can be precomputed and stored as attribute of entity set or relation to speed up the interpretation algorithm

The relational database has been implemented through multiple ISAM file system written in C and implemented under HP-UX environment.

Interpretation algorithm

The above definitions assumed that no false spatial relation can be instantiated and if a spatial relation is instantiated then the object actually giving rise to it will always be instantiated. In reality, this assumption is not always valid because of noisy data and errors in the feature detection process. The updation of generator with this false set of objects may eliminate the correct object instance from the generator and/or increase cardinality of the interpretations,thereby producing an inconsistant generator set. Here, a strategy is presented to reduce these possibilities and a scheme for circumstantial evidence analysis is incorporated in the updation operation to prevent a generator from becoming inconsistant in case of false alarms.

Interpretation generation is initiated only after a seed set of instantiated primitives have been formed. The seed set discovery is purely a data analysis procedure activated according to some pre-defined heuristic strategies. If the features can be obtained by local computation in image, one strategy calls for identification of all primitives of a particular type (say only corners of any angle) in the entire image. Another strategy does the same but over selected windows. It is assumed that with the seed set of primitives, all the objects likely to be present in the image will be hypothesized so that it can be passed on to the verification stage to find out the objects actually present in the image.

The problem of updating the generator set with a set of objects not including the correct one is tackled by introducing the concept of n-ary relations (where n is variable). A feature belonging to a particular object should satisfy spatial relations with all other features belonging to that object. Therefore, the object set used for updation of generator set is formed by considering the relation of one feature with all other instantiated features in the image. For a feature P_i, its relation with all other instantiated features is considered and for each instantiated spatial relation, an accumulator associated with the corresponding object instance is incremented. The object instance having highest number of voting for the entire set of instantiated primitives satisfies largest set of spatial constraints corresponding to P_i. However, considering that few relations belonging to the currect object instance may be missed and few spurious relations may be instantiated , all the object instances with voting close to the maximum value, are considered for updation of generator set. Interpretations produced by this process , therefore ,explains spatial constraints of nearly identical arity corresponding to individual features.

In case a primitive feature p_i belongs to two different objects actually present in the image, it may result in either both of them or only one of them being included in C_{up}. The generator set formed after updation operation may not include the correct interpretation, even though it is correct with respect to the partial infornation available at this point. This fact will not detrimentally affect correctness of the final interpreation, if atleast one non-overlapped features

p_k of both the objects are visible. The set C_{up} formed for the primitive p_k will contain this object and hence generators will be updated appropriately producing correct interpretations.

A scheme for circumstantial evidence analysis is incorporated in the updation operation. It is used for (i) establishing approximate equivalence between object instances depending on the set of spatial relations supporting these object instances, (ii) determination of reliable criterion for incrementing cardinality of interpretation, and (iii) differing decisions in case of confusing situations.

During updation, objects having identical labels and transformation vectors within predefined cluster radius are clubbed together to form a single instance. A criterion of approximate equivalence between different instances is evolved to take care of false relations. Object instances having most of the spatial relations in common are considered equivalent and are collected in the same set. If number of relations not in common is more than some percentage of the number of common relations, then they are considered different. However, if they can not be put in any of the categories then the situation is considered confusing.

During updation operation object instances belonging to object sets of the generator are removed/not removed either by or - operation if they have been found equivalent to some object of C_{up}. If an element of C_{up} is equivalent to all the elements of an object set then that element of C_{up} is put in it. Situation becomes slightly more complex when 'confusing' condition is encountered. In these situations objects are simultaneiously considered equivalent as well as different and object sets are formed to take into account both the conditions.

Consideration of approximate equivalence, essentially circumvents the problem of eliminating a valid object instance from object set of generators. A valid object may be edged out of C_{up} because of very small number of stray relations but considerations of approximate equivalence, using the bit array which records all supporting evidence, the situation is analysed in a global perspective and decisions are made. Relaxing the criterion for entry of object instances in C_{up} would not have eliminated the problem.

Similar analysis is done for determining the criterion for increasing the cardinality of a generator. Let R_{old} be the set of spatial relations explained by a generator, formed by taking union of the instantiated spatial relations of all the objects in the generator. An object is said to be subsumed in the generator if the set of spatial relations of the object is a subset of R_{old}. When the exact subset criterion is violated but the difference is small, the object is said to be approximately subsumed in the generator. If an object belonging to C_{up} is subsumed by the generator, it is simply removed from C_{up}. If the object is not subsumed by the generator, it proceeds as usual. But if it is a confusing situation then both the cases are taken into account. Normal division of the generator proceeds as usual. But, along with this, the effect of removing the object from C_{up} is also considered. If this results in an empty C_{up}, the original generator is retained, otherwise nothing is done.

There can be a reciprocal situation when generator is subsumed in the object. In either case, if it is the only generator in the generator set, all such objects in C_{up} are collected together in a single object set to form a new generator of unit cardinality, otherwise no action is taken. The object subsuming a generator is always removed from C_{up}. The subsumption operation is also useful for detecting and eliminating redundancies in the interpretation sets.

Once operations on the seed set of features are over and a number of possible interpretations of the image are formed, in the next phase the algorithm predicts uninstantiated image features from models of the objects, searches for these features and modifies the interpretations depending on outcome of the search.

The generator set is organised in the form of m-ary tree. Children of root node correspond to groups of generators all having the same cardinality. Nodes at this level contain maximum of the confidence of constituent generators. The root node stores the confidence of generator which explains maximum number of spatial relations. Whenever updation is performed, confidence values are propagated up. The arc along which values are propagated are marked. When the value propagated to the root node exceeds a threshold or the image has been substantially scanned for features, the algorithm terminates.

Following the marked arc, objects are choosen and put into 'ACTION set'. Now job queue is formed with uninstantiated features of the objects in the action set. The predicted feature location is mapped into a predefined image window. The associated data structure of the window indicates whether the feature has already been searched in that window or not. Job is created only when the feature has not been searched earlier. Priority is assigned to each job and the job with maximum priority is choosen for execution. If feature is not instantiated and gray level in the predicted location correspond to object area then the feature is likely to be obscured. Otherwise this is a negative evidence for the object and a counter of negative evidence is incremented. If the negative evidence is high enough then the object is a stray instance and should be eliminated from the object set in which it belongs to. If, as a result, the object set becomes empty then the generator is not elimінted but its cardinality is decreased and accordingly it is regrouped. Moreover, if the feature is instantiated then its relation with all other features is considered and the generator set is accordingly updated. The algorithm for this stage is nearly identical to the algorithm for creating the initial interpretation set from the seed set of features. But there is one important difference that no more new object instances are considered.

Feature and relation detectors

In this system, only two types of features have been considered, namely corners and holes. Both the corner and hole detectors are based on Hough transform technique and have been designed to work in a given window of the image.

The number of constraints on a relation between two features $f_1(x_1, y_1, 0)$ and $f_2(x_2, y_2, 0_2)$ depends on whether the features are vector feartures (feature with inherent orientation) or simple features (with no preferred direction). Therefore, a relation may have one, two or all the three constraints, namely, distance, relative orientation and orientation difference.

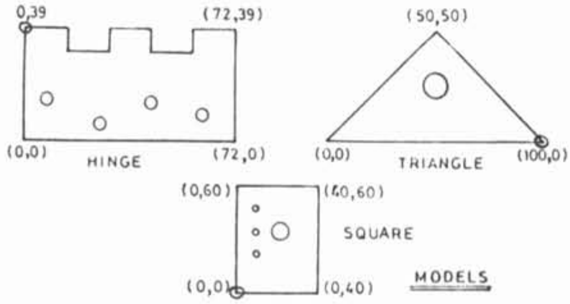
A relation type may be considered as a region in the relation space where the relation space is either 1-D, 2-D or 3-D, depending on the type of relation. Different points within a region are basically different instances of the same relation. Therefore, an instance of a relation with given attribute values is classified to be belonging to a relation type based on the region in which it falls. Since the relation space is divided into finite number of smaller regions based on some fixed rule (in the simplest case it may be divided into uniform regions), determining the region in which a point (an instance of a relation) belongs to, is simple enough.

As a more appropriate stratagy classification rules for features and relations are sunthesized by using notions of unsupervised learning and some heuristic criterion. With actual attributes of features of the training models histograms are constructed with bins of fixed size. For relations 3-d or 2-d data values are mapped on to a suquence using MODESP algorithm[5]. Then classification definitions are extracted from heuristic analysis of the local structure of the histogram or the sequence.

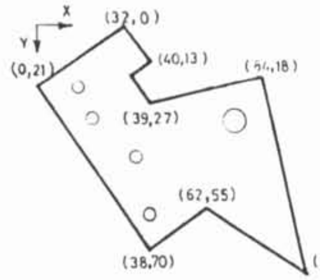
Results and discussion

The entire system is implemented in C on HP-9000 system.Experimentations are performed with multiobject scenes and correct results are obtained. Here two such examples are illustrated. In scene 1,since only few relations for triangle are found,single object interpretation of hinge was also under consideration. But there is no negative evidence for triangle and the corresponding interpretation consisting of hinge and triangle explained largest set of instantiated relations. Hence the given interpretation is correctly choosen. In scene 2, identical objects(hinge) with different transformations are identified correctly along with the square object (sandwitched between them).

Although Hough transform based feature detectors are used,the methodology is general enough to work with other kind of features and feature extractors. Extention of the approach for 3-d objects is under active consideration.

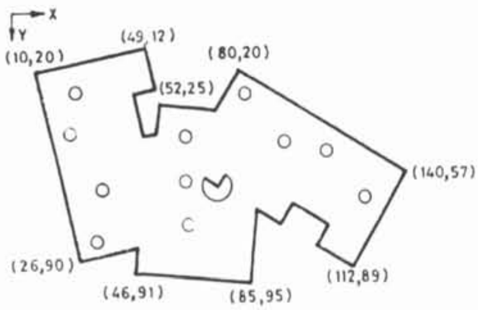


MODELS



SCENE 1

Hinge(31, 0,56),Triangle(93,76,305)



SCENE 2

Hinge(114,87,207), Square(45,90,6), Hinge(47,10,76)

References

1. Bolles, R.C., and Cain, R.A., "Recognizing and Locating Partially Visible Objects : The Focus Feature Method", The International Journal Of Robotics Research, 1, No. 3, Fall 1982, 57-81.

2. Stockman, G, Kopstein, S. and Benett, S., "Matching Images to Models for Recognition and Object Detection via Clustering", IEEE Transactions on Pattern Analysis and Machine Intelligence, 4, No. 3, May 1982, 229-241.

3. Koch W. Mark and Kashyap R.L., "Computer Vision Algorithm to Recognise and Locate Partially Occluded Objects", Engineering Research Centre for Intelligent Manufacuring Systems, Schools of Engineering, Purdwe University, Indiana, U.S.A., TR-ERC 86-2, 1985.

4. J.A.Reggia, Dana S. Nau, P.Y.Wang, A formal model of diagnostic Inference. I. Problem formulation and decomposition, Information Sciences 37, 227 (1985).

5. Kittler and Devijver , " Pattern Recognition : A statistical approach " Prentice-Hall International , 1985.