Recognition of Occluded Objects: A Cluster Structure Paradigm[1]

Bir Bhanu and John C. Ming

Department of Computer Science, University of Utah
Salt Lake City, Utah 84112

Abstract - Clustering techniques have been used to perform image segmentation, to detect lines and curves in the images and to solve several other problems in pattern recognition and image analysis. In this paper we apply clustering methods to a new problem domain and present a new method based on a cluster-structure paradigm for the recognition of 2-D partially occluded objects. The cluster-structure paradigm entails the application of clustering concepts in a hierarchical manner. The amount of computational effort decreases as the recognition algorithm progresses. As compared to some of the earlier methods, which identify an object based on only one sequence of matched segments, the new technique allows the identification of all parts of the model which match with the apparent object. Also the method is able to tolerate a moderate change in scale and a significant amount of shape distortion arising as a result of segmentation and/or the polygonal approximation of the boundary of the object. The method has been evaluated with respect to a large number of examples where several objects partially occlude one another. A summary of the results is presented.

Index Terms - Clustering, Occlusion, Recognition, Segment Matching, Sequencing, Shape Matching

I. Introduction

Recognition of partially occluded objects is of prime importance for industrial machine vision applications and to solve real problems in military domain and factory automation [3]. The problem of occlusion in a two-dimensional scene introduces errors into many existing vision algorithms which cannot be resolved. Occlusion occurs when two or more objects in a given image touch or overlap one another. In such situations vision techniques using global features to identify and locate an object fail because descriptors of part of a shape may not have any resemblance with the descriptors of the entire shape.

Because we know that occlusion will be present in all but the most constrained environments, several methods have been developed [1, 2, 4, 7, 14, 16]. The approaches to solve the occlusion problem can be classified either as boundary based [4, 8, 14, 16] or local feature based [7]. The latter approach makes use of the available local features such as holes, corners, etc. These techniques are computationally intensive. They can not handle minor distortion in the shape, change in scale and do not give good matching results over a wide range of industrial objects. Some of these factors led Price [15] to use a conceptually simple technique to solve the occlusion problem by following the *order* of matched segments in the model and the apparent object; the apparent object is formed as a result of occlusion of two or more objects. Price uses a device called a disparity matrix. Initially, the algorithm assumes that we have a linear border approximation of a given model and an image (apparent object). Price's method then compares every segment in the model with every segment in the image. If the segment pairs are compatible in terms of length and angle between successive segments, the rotational offset between the two segments is entered into the disparity matrix. The entry is indexed by the segment number in the model and the segment number in the image. If the segments are incompatible, an error code is placed at the appropriate location in the matrix. After all line segments have been compared, the matrix contains the offsets, or disparities, for all line segments. By traversing this newly formed matrix diagonally, the longest sequence in the matrix that contains compatible entries is found. From this longest sequence, the Price method then computes the transform dictated by the segment pairs in the sequence. This value is the final result of the procedure.

Unfortunately, the Price technique is very expensive computationally. Since it must treat every entry in the disparity matrix as a possible starting location of a sequence, it traverses the matrix once for every entry that exists. While this fact does not pose a problem in the simplest cases, matching takes a long time for models and images with more than about 20 or 30 segments each. Another major problem encountered by the Price procedure concerns the ability to use more than one sequence in the overall matching of the model to the image. It uses only the longest sequence found in the traversal because it cannot determine the compatibility between more than one sequence. In the cases where a large amount of occlusion is present, Price's technique will not be very successful. Thus, while Price's early attempt to solve the occlusion problem met with limited success, it did not fully deal with the problem.

In order to overcome the factors which caused the Price method to fail, we have used a cluster-structure paradigm which allows the recognition of a given object and provides information about the orientation and position of the objects in the image [5]. Some techniques and structures used by Price have been retained, but the method of matching is entirely new. The cluster-structure paradigm applies the clustering concept in a hierarchical manner, which reduces the amount of computational effort as the recognition algorithm progresses. Basically the technique consists of three steps: clustering of segments, finding sequences of segments in appropriately chosen clusters and clustering of sequences. Sequences of segments are found in one pass over the data. The length of each of these segments as well as the angle between successive segments comprise the only information needed by the algorithm to recognize objects and

[1]This work was supported in part by NSF Grants DCR-8506393, DMC-8502115, ECS-8307483 and MCS-8221750

CH2282-2/86/0000/0776$01.00 © 1986 IEEE

find their position and orientation. The motivation behind the cluster-structure paradigm and how the clustering techniques can be useful in solving the occlusion problem is given in section II. Section III presents the algorithmic description of the new method. Section IV provides a number of examples illustrating the capabilites of the technique. Finally section V presents the conclusions of the paper.

II. Principle of the Technique

In an unstructured environment, for example parts going over a conveyor belt, occlusion may occur in a large number of images that have to be processed. This problem requires the development of an algorithm which can easily handle occluded scenes. Figure 1 shows the block diagram of the clustering based occlusion algorithm. It is based on the premise of successive applications of clustering to the model and image data and the use of structural information in an environment in which the data reduction takes place as the recognition process progresses and the confidence in recognition is increased.

Clustering, in its most general form, groups a set of objects into subsets where objects in a subset are more similar than the objects in other subsets [11, 12]. Clustering techniques are commonly used in Pattern Recognition and Image Processing. (For a recent review see [13].) A significant problem inherent in using the clustering techniques involves the choice of the number of clusters to be used at any given time. Since the choice depends on the structure of the data that is being clustered, the number of clusters cannot usually be a constant value in a given application. As an example, when applying clustering techniques to solve the occlusion problem, the number of clusters must be altered, depending on the lighting conditions, the segmentation techniques used, the amount of occlusion present, and many other factors. Fortunately, there are measures which can be used to find the intrinsic number of clusters [6, 9, 10, 13]. These performance measures determine the scattering of the samples within each individual cluster as well as the distance between each of the cluster centers themselves. This information is held in a matrix form known as a scatter matrix [12]. The scattering of the samples in a particular cluster is defined as within-cluster scatter matrix, S_w. The overall position of all clusters in relation to each other becomes the between-cluster scatter matrix, S_b. By definition, the β value for a certain clustering equals the trace of the within-cluster scatter matrix multiplied by the trace of the between-cluster scatter matrix, i.e., $\beta=Tr(S_w)Tr(S_b)$. As the number of clusters increases, the value of β will reach a maximum and then slope towards 0. The number of clusters at which the value of β is a maximum is the desired value and gives the best results. Thus, by setting the number of clusters to be one, clustering the samples, computing the β value, comparing the β value with its last value, and continuing until the maximal β value is reached, a program can find the best number of clusters in a given data set.

In determining the number of clusters using the above method, a clustering technique, such as the K-Means Algorithm could be used. After every sample has its feature vector computed, the algorithm creates an arbitrary known number of K cluster centers, into which all samples will be placed. In order to determine which center to place a given sample in, the algorithm computes the distance from the sample to each of the cluster centers. This distance is merely the Euclidean distance from the sample to the cluster center in the feature space. The sample belongs to the cluster which is closest to it. When every sample has been assigned to a unique cluster, the algorithm recomputes the value of each of the cluster centers. The new cluster center is the average of all samples which are currently in that cluster. After the new cluster centers have been determined, the algorithm then redistributes all of the samples again, using the new centers this time. The process continues until no further changes take place in the location of the cluster centers. At that point, the samples in each of the clusters are said to be compatible with each other. Thresholds help to determine when cluster centers have become stable. Since the Euclidean distance may be affected by the choice of the features present in the feature vector, the feature values are normalized so that each feature contributes equally to the overall distance.

The above techniques are computationally efficient, even when the number of samples is very high, thus allowing the vision process to be fast as well as accurate. Since the clustering method groups all sets of compatible matches into a single cluster, regardless of their position in the image, it can find multiple sequences in a model which may match in the given image. While the Price method was only able to find a single best matching sequence, the new procedure will find as many sequences in the image as possible. Also, since the traversal of matrices in Price's method is computationally expensive, the new technique improves the speed of the matching as well. The use of clustering and the performance measures in the body of the algorithm are discussed in the next section.

III. Algorithm Description

As shown in Figure 1, the algorithm consists of the following main computational steps:

1) Disparity Matrix
2) Initial Clustering
3) Sequencing
4) Final Clustering
5) Transform Computation

Each of these steps will be described individually in this section. Two sets of data are assumed to be given. The first set contains the object model data, which is a set of vertices that define the boundary of the model. This model is the object that we are searching for in the image. The second data set contains the description of the image that has been acquired. This data is also a set of vertices on the boundary that describe the scene that was taken by a video camera.

Disparity Matrix: The first step of the algorithm consists of the formation of the disparity matrix. From the set of vertices for the object and the image, the algorithm determines the length of each segment and the angles between successive segments. At this point, every segment in the object model will be compared with every segment in the image. If the segment lengths and successor angles are compatible, the rotational and translational disparity between the pair of segments are computed. These values are stored in the disparity matrix, indexing the values according to the segment number in the model and the image. This process proceeds until all segments have been compared. Now the range of rotational and translational values present in the matrix are determined, and the disparity matrix is normalized over their appropriate range. The normalized values are kept in a normalized disparity matrix, since the initial disparity matrix needs to be retained for later use. This matrix is similar in structure to the matrix used by Price [15]. However, in addition to the rotational offsets, we also place translational offsets in the disparity matrix.

The computation time required to complete this step comprises about 10 to 20 percent of the total execution time. Since all of the values must be compared with each other, the exact percentage depends on the total number of segments present in the model and image.

Initial Clustering: After all of the normalized values have been placed into the disparity matrix, the algorithm clusters these values where the feature vector is merely the normalized rotational and translational offsets for each of the pairs of line segments. The initial number of clusters is one and in the application of K-means algorithm the first sample becomes the first cluster center. The clustering proceeds as described in the previous section. At each step, all of the samples are clustered, the value of the new cluster centers are recomputed, and this process continues until none of the cluster centers change their positions. Now for the current cluster results, scatter matrices are computed and the value of β is determined. The algorithm then compares the current β value with the last β value. If the value has decreased, then the previous β value and the number of clusters become the final result of this processing step.

This step of the algorithm takes the most computational time of all of the steps due to the large amount of samples that are clustered. For example, if the model contains 25 segments and the image contains 100 segments, the disparity matrix will contain 2500 entries. Out of this number, 500 samples may be present in the disparity matrix which satisfy the length and angle thresholds. If the program has to cluster these samples 3 times until β is maximized, 1500 distances must be computed. However, this amount of computation is far less than the comparable computation that would need to be done by the Price method.

After the number of clusters have been determined and the results are known for that particular value, the program selects the cluster with the largest number of samples. The data in this cluster will be used by the rest of steps in the algorithm to determine the location and orientation of the model in the image. However, since some of the other clusters may contain approximately the same number of samples as the largest one, the program also uses any cluster which is within 20% of the largest cluster. Each cluster is considered separately and the final transform comes from the cluster which yields the highest confidence level. The confidence metric is discussed below. The program now passes each cluster that has been selected to the following algorithm steps, one at a time.

Sequencing: Since the clustering results provide no information concerning the physical structure of the model, this information must be provided at this time. Using the samples in the current cluster, the program finds all sequences in these samples. For instance, if the previous sample indicates that segment 1 in the model matches segment 27 in the image (represented by the notation [1,27]), the program then searches for the pair [2,28], since this pair should logically follow the first pair on the borders of the model and the image, respectively. Since there may be some missing and extra segments in the model and the image as a result of segmentation, polygonal approximation, and various other reasons, we allow up to 2 extra or 2 missing segments when finding the sequences. This procedure continues until all possible sequences have been located in the data of the current cluster. This step provides the only structural information within the algorithm and cannot be omitted.

Any samples in the current cluster which were not placed in any sequence are discarded. Since these samples are not members of any sequence, they usually represent the extraneous data in the cluster that was mentioned earlier. The program also removes any sequences which have a segment count of less than three. Three segments make the basic local shape structure. This removal insures that arbitrary data included in the initial clustering and sequenced by the current step is not included in the final processing steps. Because of their small length, these sequences are assumed to be invalid and have low confidence. Even if the sequences indicated valid matches, their removal from the set of sequences does not introduce any error into the final matching that will be computed.

The final task to be accomplished at this step of the algorithm is to compute the rotational and translational average of each sequence that has been located. These averages are merely the averages of all of the samples that are present in each sequence. These sequences and their averages will be used in the final clustering step of the program.

The sequencing step requires the second largest amount of execution time within the entire program. Since it is still very costly to check the possibility of a sequence occurring at any given sample, the program must check every sample in order to locate the best choices. However, because the clustering results have greatly reduced that amount of choices that need to be checked, this step takes far less time than does the Price method. It is a one pass algorithm over the data.

Final Clustering: Using the sequences and the sequence averages obtained from the previous step, the algorithm clusters theses values to find those sequences which lead to the same rotational and translational results. As with the initial clustering, the program uses the iterative technique of clustering, evaluating, clustering, etc. After the value of β has reached its maximum, the program again selects the cluster which contains the largest number of sequences and passes this cluster to the final program step.

While the initial clustering step had to deal with a large number of samples, this step of the program uses a trivial portion of the total program time. In all the examples discussed here, the number of sequences is less than 100, with an average somewhere near 30 to 40. Also, since the sequencing step has eliminated a good deal of the erroneous data, the β value quickly reaches it maximum and this steps ends.

Transformation Computation: After all clusters which were selected have been sequenced and clustered a second time, the program determines the confidence level of the transformation determined by each cluster. The cluster with the highest confidence level is selected as the final transformation cluster. The program assembles the set of matched segments included in the sequences in this cluster. These segments are sorted into increasing model segment number so that the sequences will indicate successive segments around the object boundary. The final output of the program is the rotation and the vertical and horizontal translation necessary to locate the object model within the image. The program also produces a confidence level which indicates the likelihood that the final matching is correct. The confidence level is found by dividing the cumulative length of all segments in the final matching by the total length of all segments in the object model. So, if the confidence factor is 80/200, we are 40% sure of the program results. This factor will be used by later versions of the program to decide if further processing should be done in order to insure the proper results. Confidence levels of 10 percent or more usually lead to the correct transformation. Of course, its value depends upon the degree of occlusion within the image.

IV. Results

Image Acquisition & Polygonal Approximation: In order to determine the ability of the program to find objects in an occluded scene, a set of 14 models was obtained and used in the matching algorithm. The models consist of a set of tools such as a hammer, screwdriver, pliers, wrench, and so on. The model for each of these tools was created by first acquiring an image of the object lying by itself on a backlit table. The image is obtained with a camera and a

commercially available digitizer. The digitized picture of the scene is 576 pixels wide by 720 pixels high. After the image has been digitized, it is then transferred to a Vax 11/780 computer for the remainder of the processing.

Once the image has been obtained, the program finds the border of the tool using a simple border follow algorithm. This procedure traverses the boundary of the object in the image and marks the pixels which lie on the border. The number of boundary points for the tools range from 567 to 1425 pixels. After locating the boundary of the object, the program uses a curvature maxima algorithm to compute an initial border approximation of the object. This procedure uses the local curvature at every border pixel in conjunction with a smoothing factor to find the approximation of the object. These approximations range from 18 to 52 segments in length. The smoothing factor, which controls the size of neighborhood around a boundary point, is 8 for all of the models.

Using both the initial border approximation and the border pixels themselves, the program then uses the split-merge technique to determine the final border approximation for each of the objects. This method splits all border segments with bad approximations and combines all pairs of adjacent segments that do not cause a dramatic change in the model representation. The number of final border segments varies from 5 to 33 for the 14 models that were used.

The task of obtaining images to be processed proceeds in exactly the same manner as in the model acquisition. The images are obtained using the same hardware and are then moved to the Vax 11/780. These images are then processed with the border follow, curvature maxima, and split-merge algorithms. Some examples of the images collected in this test run appear in Figure 2. For this particular experiment, 20 images were collected and then processed. In these 20 images, 56 instances of the 14 models are present. The number of boundary points varied from 1123 to 4025 pixels. The smoothing factor used to obtain the initial image approximation is 24 for all images. The number of curvature maxima segments ranges from 21 to 71. The final number of border segments received from the split-merge procedure range from 26 to 71 segments.

Model Based Recognition: Once all of the models and images have been collected, the clustering algorithm is used to locate the models in the images. When the clustering program was run on the 20 images that were collected, the results were very good. Of the 56 models present in these images, 47 (84%) models were correctly matched. 4 of the 56 models were mistakenly matched to a different model. The remaining 5 model instances could not be matched. Figure 3 shows the matching results for the five images that were shown in Figure 2. Solid lines show the polygonal approximation of the images using the split-merge algorithm. The dotted lines show the polygonal approximation of the model at its matched location in the image.

Out of the 47 models that were correctly matched in the images, 7 of these matches were not successfully found until the polygonal approximation of the appropriate model was improved. Of the 5 model instances that were not located in the images, the failure to find these objects is due to the substantial difference in the polygonal representations of the particular tool in the model and in the image. When the polygonal approximations of the object become too diverse, clustering is not able to overcome this problem. However, if the representation of the model is improved within the image, matching occurs and the transformation determined by the program is better. For example, in Figure 3c, the pipe wrench has been properly matched, but the transformation of the model is not very good due to the large difference in polygonal approximation. Figure 4 shows the results after the representation of the model has been substantially improved within the image. Now the new transformation is only slightly improved. However, more model segments were matched in this case which led to a higher confidence level. Note that a change in scale occurred between the models and their representations in the image during the process of image acquisition.

Execution times for the clustering method range from .5 to 3.2 seconds on a Vax 11/780. No attempt was made to optimize the C code. After each matching has been determined, the program also computes a confidence level based on the segments which contributed to the matching. This value is computed by dividing the cumulative length of all matched segments by the cumulative length of all segments in the model. For this set of images, confidence levels vary from 0 to 98%. The error analysis of the matchings which were correct yield a mean rotational error of -0.14^{o} and a standard deviation of 8.68^{o}. The mean translational errors are -11.83 and -7.65 pixels for translation in x and y, respectively.

To contrast this new method with Keith Price's earlier work, we have run Price's algorithm on several of the examples used above. Figure 5 shows the results of this matching. Out of the 7 models present in Figure 5, only 2 were properly matched. These results can be compared with the clustering results in Figure 3a and 3e. The clustering method, which obviously yields far better results, also finds the matches in much shorter time. The model transformations also tend to be better because clustering can find several sequences along a boundary which may contribute to the final transformation. Price's method, on the other hand, only selects the longest sequence.

V. Conclusions

Based on the results presented in this paper, we conclude that the cluster-structure paradigm is a robust approach to solve the occlusion problem in 2-D. The data and the amount of computation reduce in a systematic and hierarchical manner. Since the technique does not limit itself to a single sequence of line segments on the border of an object, it can locate all of the matched segments of the model, which accounts for the high success rate. The program was not highly successful in the instances of severe occlusion, where a given model has only about 5% of the total number of segments visible in the image. In those situations, even an expert would have problems locating a model within the image.

Future modifications to this algorithm will include the ability of the algorithm to handle cases when the model has a large amount of symmetry and considerations for grasping strategies by a manipulator.

References

[1] N. Ayache. A Model-Based Vision System to Identify and Locate Partial Visible Industrial Parts. In *IEEE Conf. on Computer Vision and Pattern Recognition*, pages 492-494. June, 1983.

[2] B. Bhanu. Shape Matching of Two-Dimensional Occluded Objects. In *Proc. 6th International Conference on Pattern Recognition*, pages 742-744. Munich, West Germany, October, 1982.

[3] B. Bhanu. Automatic Target Recognition: State-Of-The-Art Survey. *IEEE Trans. on Aerospace and Electronic Systems* to appear, July 1986.

[4] B. Bhanu and O.D. Faugeras. Shape Matching of Two-Dimensional Objects. *IEEE Trans. on Pattern Analysis and Machine Intelligence* PAMI-6(2):137-155, March, 1984.

[5] B. Bhanu and J.C. Ming. *Recognition of 2-D Occluded Objects and their Manipulation by PUMA 560 Robot.* Technical Report UUCS85-111, Department of Computer Science, University of Utah, August, 1985.

[6] B. Bhanu, A.S. Politopoulos and B.A. Parvin. Intelligent Autocueing of Tactical Targets. In A. Oosterlinck and P.E. Danielsson (editor), *Architecture and Algorithms for Digital Image Processing, Proc. SPIE 435*, pages 90-97. August, 1983.

[7] R.C. Bolles and R.A. Cain. Recognizing and Locating Partially Visible Objects: The local-Feature-Focus Method. *The International Journal of Robotics Research* 1(3):57-82, Fall 1982.

[8] W.K. Chow and J.K. Aggarwal. Computer Analysis of Planar Curvilinear Moving Images. *IEEE Trans. Computers* C-26:179-185, February, 1977.

[9] G.B. Coleman. *Image Segmentation by Clustering.* Technical Report USCIPI 750, Image Processing Institute, University of Southern California, Los Angeles, California, 1977.

[10] E. Diday. *Problems of Clustering and Recent Advances.* Technical Report 337, IRIA Laboria, Domaine de Voluceau, Rocquencourt, France, January, 1979.

[11] R. Dubes and A.K. Jain. Clustering Techniques: The User's Dilemma. *Pattern Recognition* 8:247-260, 1976.

[12] R.O. Duda and P.E. Hart. *Pattern Classification and Scene Analysis.* Wiley-Interscience, New York, 1973.

[13] K.S. Fu and T.Y. Young, Eds. *Handbook of Pattern Recognition & Image Processing.* Academic Press, 1985. Cluster Analysis, Chapter by A.K. Jain.

[14] M.W. Koch and R.L. Kashyap. A Vision System to Identify Occluded Industrial Parts. In *IEEE Int. Conf. on Robotics and Automation*, pages 55-60. March, 1985.

[15] K.E. Price. Matching Closed Contours. In *Proc. 7th Int. Conf. on Pattern Recognition*, pages 990-991. July-August, 1984. Also in Tech. Report 104, Intelligent Systems Group, University of Southern California, Los Angeles, October 19, 1983, pp. 29-37.

[16] J.L. Turney, T.N. Mudge and R.A. Volz. Recognizing Partially Occluded Parts. *IEEE Trans. on Pattern Analysis and Machine Intelligence* PAMI-7:410-421, July, 1985.

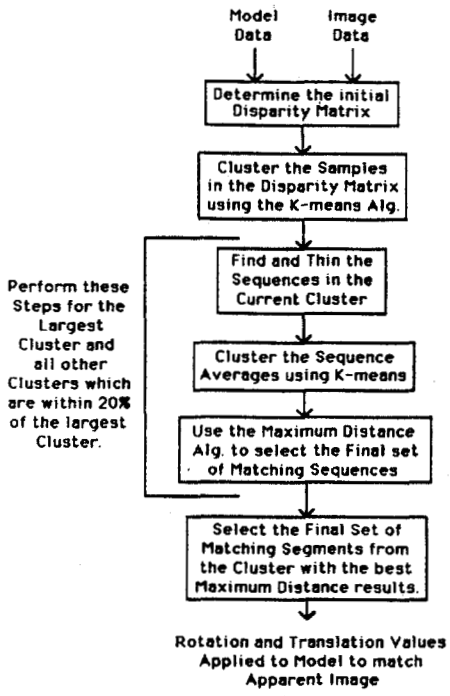

Fig. 1 Block diagram of the clustering based occlusion algorithm.

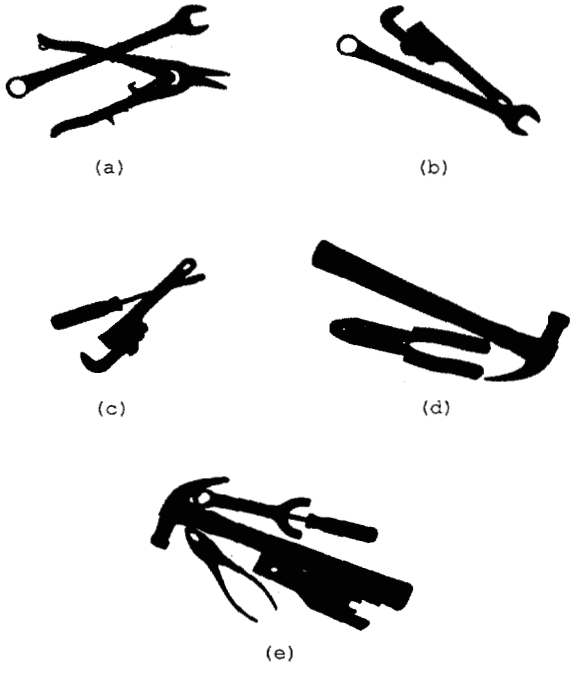

Fig. 2 Images of the occluded objects (a to e).

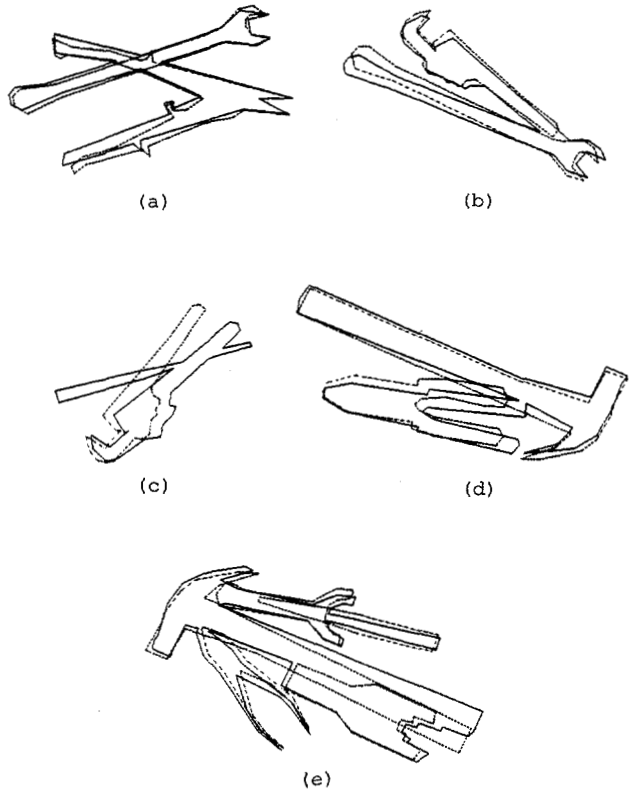

Fig. 3 Results of matching for the occluded images shown in Fig. 2.

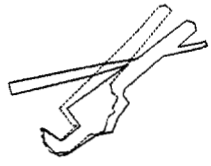

Fig. 4 Results of matching using the improved polygonal approximation for the model wrench only.

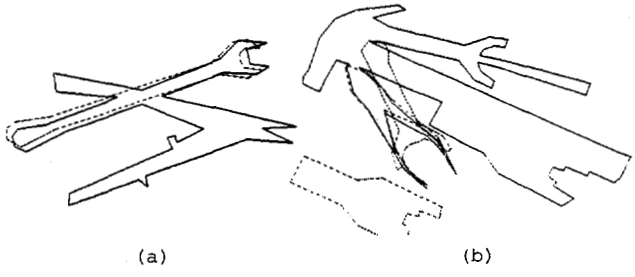

Fig. 5 Results of matching using the Price algorithm. Figs. 5(a) and 5(b) correspond to Figs. 3(a) and 3(e), respectively.