MODEL ACQUISITION AND MATCHING IN TAGGED OBJECT RECOGNITION (TOR)

L. M. Soh, J. Matas, J.Kittler
Centre for Vision, Speech and Signal Processing,
Department of Electronic and Electrical Engineering,
University of Surrey, Guildford, Surrey GU2 5XH, United Kingdom
Tel: +44 1483 259838 ; fax: +44 1483 534139
e-mail: {l.soh,g.matas,j.kittler}@ee.surrey.ac.uk

ABSTRACT

General object recognition is difficult. We had proposed a novel solution for object recognition in an unconstrained environment using a tag (Matas[2]). This simplifies the problem by placing tags of special pattern on the objects that allows us to determine the pose easily. A robust calibration chart detector was developed for the first stage of the solution (Soh[1]). This paper investigates the next part of the solution, i.e. the model acquisition and matching using the Chamfer matching algorithm. The algorithm is reasonably simple to implement and very efficient in terms of computation. We experiment with this technique extensively to prove the reliability of the approach. Using this approach, the objective of Tagged Object Recognition (TOR) can be realised and we should be able to perform object recognition wherever a tag is located. The technology will facilitate applications in landmark and object recognition, mobile robot navigation and scene modelling.

1 INTRODUCTION

If a visual recognition system is to interact intelligently and effectively with its environment, it must be able to decide which objects are present in the scene and determine the position of each object with respect to the sensor. The problem of object recognition is complex as objects have different appearance in the image when taken at different viewpoints, illumination, occlusion and background. Matching an object model with the image would result in an extensive search through the large pose space and in the variety of conditions as stated above. The majority of successful 3D object recognition systems to date have concentrated either on a very specific class of objects (e.g. polyhedra) or are limited to environments with controlled illumination, viewing geometry, background etc. [7].

Conversely, if the 3D pose of an object is known, its identity can be confirmed simply by a direct comparison of the image with a model projected onto the image plane. In order to recognise objects without restrictions we have developed a system by placing a readily detectable small pattern of black circles on a white background on the objects of interest. This pattern or chart was based on the design by Tsai[3],[4] for camera calibration and a shareware exists for determining its 3D pose with respect to the camera [5]. A complex but extremely robust technique [1] has been developed to handle detection of the chart over a large range of scales, in very difficult illumination conditions and in the presence of partial occlusion. Wherever the chart is (or charts are) detected in the image, the chart or object pose in the camera coordinate system is known.

With this Tagged Object Recognition (TOR), visual recognition task can be simplified by defining the scene as a collection of tagged objects either fixed or moving. Potential applications of TOR include autonomous vehicle navigation, landmark and object recognition and scene modelling.

In this paper we address the issues of the choice of model features and the problem of acquisition and matching the model reliably. To identify an object that is tagged by the chart we use the edge map in the vicinity of the chart as the model features. In contrast to previous effort, higher level features need not be extracted since the edge map will contain sufficient information about the model in terms of texture, shape and size. There are two ways of acquiring the model. The first method uses a blank background where the model is captured at a controlled environment in a single view. The other utilises two or more images with different background and the model is segmented by correlation of the images. The second method is especially useful when the object is a fixture, e.g. a road sign. For comparing the two binary edge maps, Chamfer matching algorithm is used to obtain the score of the correlation as it is simple, efficient and reliable [6]. The details of the above process are described in the later portion of this paper. The test results of the matching are promising.

2 THE CHOICE OF MODEL FEATURES

Traditional methods of pattern recognition require the use of higher level structures like lines and ellipse as the object pose is unknown. With TOR, the higher level features are not selected as there is no need to minimise the extensive search in the pose space since the pose is determined by other means [3]. Hence, the edge map around the chart is used to represent the object. This also gives a versatile and accurate object description.

3 THE CHAMFER MATCHING ALGORITHM

For matching two binary images, we use the Chamfer matching technique as it is quite insensitive to noise and other disturbances [6]. The Chamfer matching algorithm searches for the best fit of edge points from two different images. The first step of the algorithm requires to generate the distance map from a binary edge map. This distance map associates with each pixel, its distance from the closest edge profile pixel. As the true Euclidean distance is costly to compute, we use a sequential Chamfer Distance Approximation. Two passes over the image are needed for this approximation. The details of the computation of the distance map can be found in Borgefors[6]. By placing another edge map over the distance map, the distances between the two edge maps can be read along the edge profile.

4 MODEL ACQUISITION AND MATCHING

To acquire the object models we have to segment the edge map of each object. The best way is to view an object with a blank background and the edge map recovered from the single view is treated as the model. This method gives the least noise and provides the best description.

Another approach is to correlate two different views of the object in the same canonical frame. First, two images of the object with apparently different background (e.g. different viewpoint of the camera) are captured. The pose of the objects in both images are then calibrated [5] and the edges from the images are extracted. Next, the Chamfer distance map is computered on one of the images using the edge map. Following this, with the pose known, the edge map of the other image is transformed onto the image plane of the distance map. By superimposing the transformed edge map of the other image on the distance map, a set of distances that stand between them can be found along the edge profiles. This set of correlation distances is used to differentiate between the model and the rest of the image.

A problem occurs when the edge profile of one image extends over the other : then, the score does not reflect the best match anymore. To handle this, we match the two images in both ways, i.e. image A against image B then image B against image A. The lowest score is taken as the final score.

For matching, a similar method is applied. However, in practice, we use a prestored distance map and pose from the object database and compare them one by one with the transformed edge map of the test image. The map attaining the minimum distance between itself and the image, if above a threshold, identifies the tagged object.

5 TEST RESULTS

A small database of images with tagged objects and landmarks has been assembled. First, the models are acquired and an example of the process is shown in Fig. 1 and 2. The best match is shown by segmenting the model from the test image. The results demonstrate that object models can be reliably acquired automatically. Next, test images are introduced and matching takes place (see Fig. 3). The figure shows one of the learned objects in a cluttered scene. The tag is first detected and database models are then mapped onto the scene and compared. The outcomes of the test shows a favourable distinction between a mismatch and a hit.

6 SUMMARY AND CONCLUSION

We have successfully developed a robust and efficient technique for model acquisition and matching of binary images using the Chamfer matching technique. The technique facilitates reliable Tagged Object Recognition (TOR), where landmarks and objects are tagged with a special pattern. This provides enabling technology for vision based mobile robot navigation, 3D reconstruction and scene modelling.

References

[1] L. M. Soh, J. Matas and J. Kittler. Robust Recognition of Calibration Charts *IEE 6th ICIPA*, 2:487-491, Jul. 1997.

[2] J. Matas, L. M. Soh and J. Kittler. Object Recognition Using a Tag *IEEE ICIP'97*, 2:887-891, Oct. 1997.

[3] R. Y. Tsai. A Versatile Camera Calibration Technique for High-accuracy 3D Machine Vision Metrology Using Off-the-Shelf TV Cameras and Lenses. *IEEE Journal of Robotics and Automation*, 3:323-344, Aug. 1987.

[4] R. Y. Tsai. Synopsis of recent progress on camera calibration for 3D machine vision. In O. Khatib et al., editor, *The Robotics Review*. MIT Press, 1989.

[5] R. Willson. Tsai Camera Calibration Software. Web Site *http://www.cs.cmu.edu/People/rgw/TsaiCode.html*, Oct. 1995.

[6] G. Borgefors. Hierarchical Chamfer Matching: A Parametric Edge Matching Algorithm. *IEEE PAMI*, 10:849-865, Nov. 1988.

[7] W. Eric L. Grimson. *Object Recognition*. MIT Press, 1990.

Fig. 1 Model Acquisition Using 2 Images with Arbitrary Background and Pose

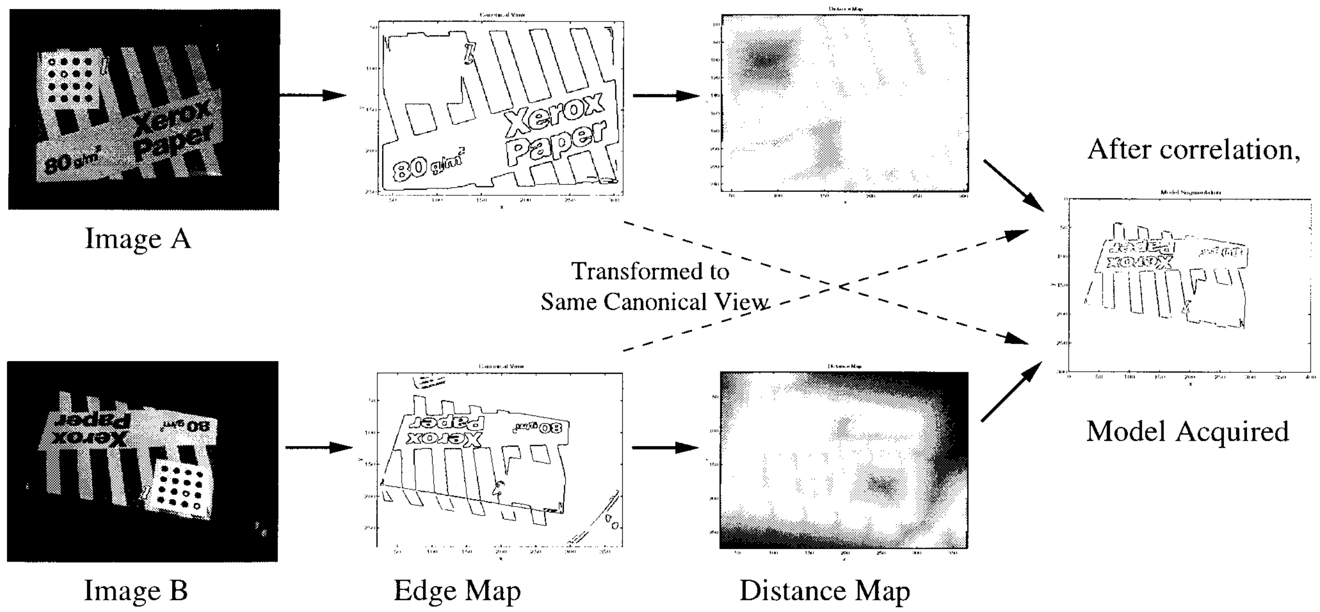

Fig.2 Model Acquisition Using 2 Images with Arbitrary Background and Pose

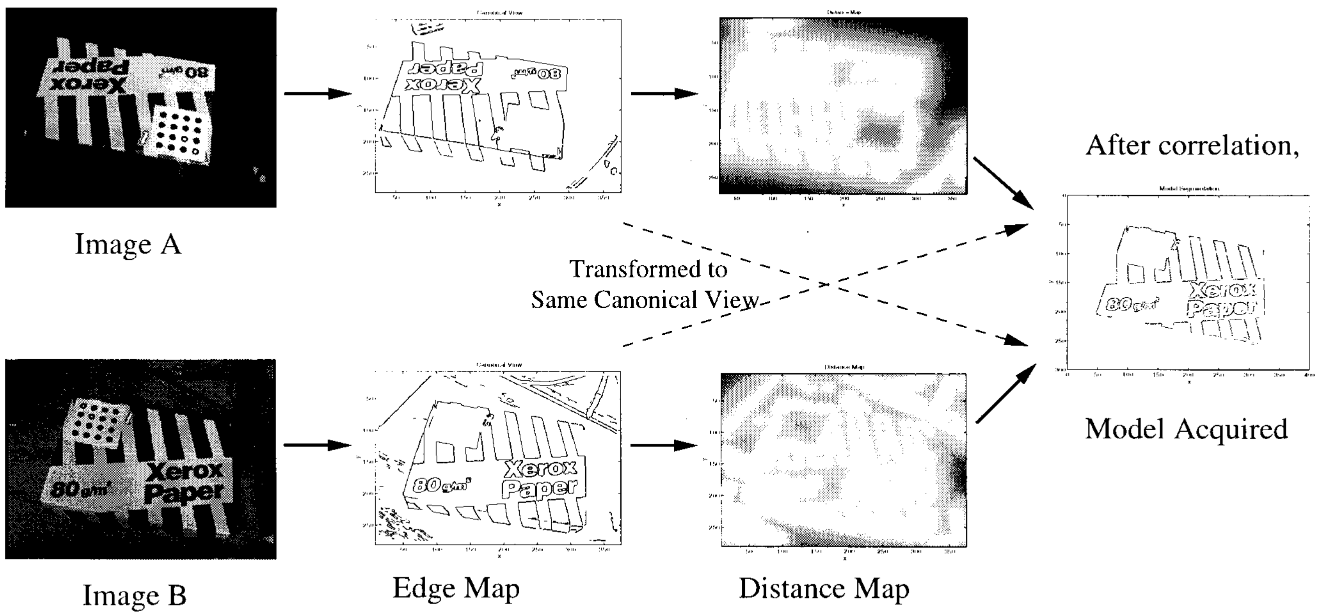

Fig. 3 Matching using Test Images