BMVC 2002

A Comparative Study of Two Object Recognition Methods

Alireza Ahmadyfard and Josef Kittler
Centre for Vision Speech and Signal Processing
University of Surrey
Guildford, GU2 7XH, UK
A.Ahmadyfard,J.Kittler@eim.surrey.ac.uk

Abstract

An experimental comparative study between two representation methods for the recognition of 3D objects from a 2D view is carried out. The two methods compared are our ARG region-based representation [1] and the elliptic region-based method of Tuytelaars et al[9]. The results of the experiments conducted show that the former method outperforms the latter particularly under sever scaling and also when applied to objects with curved surfaces.

1 Introduction

Object recognition is a subject that has been studied in computer vision for decades. In this paper we are concerned with model-based recognition methods in which an object is represented using its 2D image(s) [3]. These methods can be broadly classified into two categories: feature based and appearance based. In an attempt to combine the advantages of feature and appearance based approaches, methods based on the matching of local features have recently been proposed. Schmid et al [7] extract the image corner points using the Harris detector and describe the image in a neighbourhood of an extracted point by a set of similarity transformation invariant features named local jets. In [9] Tuytelaars et al extend the invariance of object descriptors to affine transformation. They construct elliptic regions around the points of local intensity extrema [9]. Each extracted region [9] is described using moment invariant features defined in terms of the three colour bands. In the approaches proposed by both Schmid and Tuytelaars the local features (regions) of the scene are directly matched against those of object models. The model which gives the best match defines the identity of the object in the scene.
In [2] we proposed a recognition method in which each object is represented in terms of its image regions. The regions are normalised in an affine invariant manner and subsequently represented by an Attributed Relational Graph (ARG) where each node and link between a pair of nodes are described using unary and binary features respectively [2]. Object recognition is achieved by comparing the scene ARG to the graph of object models using relaxation labelling. In a recent work [1] we adopted new binary measurements which characterise a pair of regions using the ratio of line segments on the line connecting the region centroids.

graph takes label ω_α. Obviously the majority of labels in Ω are not admissible for O_i. Therefore in the first stage of matching we compile a list of admissible labels for any scene node O_i denoted by Ω^i. This list is constructed by considering the unary dissimilarity measure between each scene node and all nodes in the model graph. Note that we include the null label in the label list of all the scene nodes, as it can potentially be assigned to any node in the scene. In the second stage of matching the modified labelling probability updating formula is applied [1]:

$$P^{(n+1)}(\theta_i = \omega_\alpha) = \frac{P^{(n)}(\theta_i = \omega_\alpha) Q^{(n)}(\theta_i = \omega_\alpha)}{\sum_{\omega_\lambda \in \Omega} P^{(n)}(\theta_i = \omega_\lambda) Q^{(n)}(\theta_i = \omega_\lambda)} \tag{7}$$

$$Q^{(n)}(\theta_i = \omega_\alpha) = \prod_{j \in \mathcal{N}_i} \{ \sum_{\omega_\beta \in \{\Omega^i \cap \Omega_\alpha\}} P^{(n)}(\theta_j = \omega_\beta) P(\overline{\mathbf{A}}_{\mathbf{ij}} | \theta_i = \omega_\alpha, \theta_j = \omega_\beta) \tag{8}$$

$$+ \sum_{\omega_\beta \in \Omega^i - \{\Omega^i \cap \Omega_\alpha\}} P^{(n)}(\theta_j = \omega_\beta) \eta \} \tag{9}$$

The relaxation labelling technique updates the labelling probabilities in an iterative manner using the contextual information provided by the nodes of the graph. In this formulation $Q(\theta_i = \omega_\alpha)$ is the support function which measures the consistency of the label assignments to the scene nodes in the neighbourhood of O_i, assuming O_i takes label ω_α. The labelling consistency is expressed as a function of the binary measurement vectors associated with the centre node O_i and its neighbours. We evaluate the distribution function in terms of the dissimilarity between the corresponding binary vectors assuming that the degree of similarity is modelled by a Gaussian:

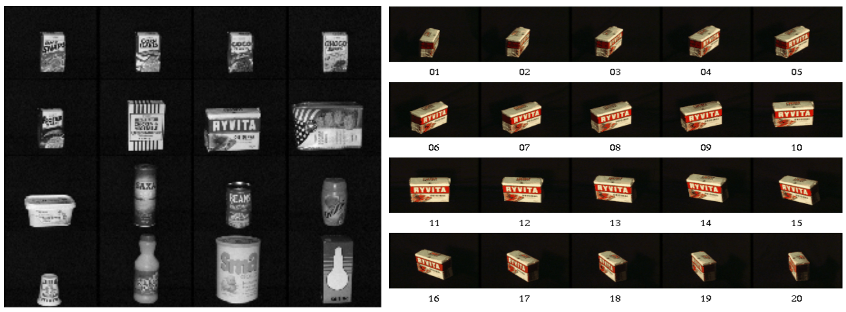
$$P(\overline{\mathbf{A}}_{\mathbf{ij}} | \theta_i = \omega_\alpha, \theta_j = \omega_\beta) = \mathcal{N}_{BinDis(i,j,\omega_\alpha,\omega_\beta)}(0, \sigma) \tag{10}$$

The support function consists of two parts: the first part measures the contribution from Ω_α neighbours (the main support) and the second part is added to balance the number of contributing terms via the other labels in Ω [1]. η is a parameter which plays the role of the binary relation distribution function $P(\overline{\mathbf{A}}_{\mathbf{ij}} | \theta_i = \omega_\alpha, \theta_j = \omega_\beta)$ when the model nodes ω_α and ω_β are not neighbours.

Upon termination of the relaxation labelling process, we have a list of correspondences between the nodes of the scene and model graphs. We count the number of scene nodes matched to the nodes of each object model and use this measure as an object matching score.

5 Experiments and results

We designed two experiments to compare the three methods(P-ARGOR, AIFOR and P-AIFOR). The aim of the first experiment was to assess the relative performance of the methods under affine transformation. For this purpose we used SOIL47 (Surrey Object Image Library) database which contains 47 objects each of which has been imaged from 21 viewing angles spanning a range of up to ±90 degrees. Fig1(a) shows the frontal view of the objects in the database. Note that the majority of the objects in this database have planar surfaces which is the requirement of both recognition methods. The database is available online [4]. In this experiment we model each object using its frontal image



(a) (b)

Figure 1: a) A number of objects in SOIL47 database b) An object in SOIL47 database imaged from 20 viewing angles

while the other 20 views of the objects are used as test images (Fig 1(b)). Furthermore to test the recognition methods under object scaling, we simulated this transformation by re-sampling each test image of the database using the **resize** function in Matlab. As this function automatically filters out the noise of the camera and image digitisation process we restored the original noise level by adding a Gaussian noise to the re-sampled images. The scaling parameter was sampled so as to produce test image sizes of 25%, 37.5%, 50%, 75% of the original image set. Note that throughout the experiment we used the full size images as the object models. We evaluate the recognition methods in terms of two performance criteria: the correct recognition and the false rejection rates. The first criterion gives the average percentage of cases in which objects in test images are correctly recognised. As a complementary measurement of performance we consider the misclassification rate. Figs 2(a),2(b) and 2(c) show the correct recognition rates as a function of object pose for AIFOR, P-AIFOR and P-ARGOR methods respectively. The graphs are parametrised by test image size. The relative misclassification rates of the three methods for the test images at scale factor one are plotted in Fig 2(d).

The aim of the second experiment was to compare the recognition methods when applied to objects with non-planar surfaces. Although the extracted regions and their associated features in both representation methods are invariant only for planar surfaces, it is useful to evaluate their sensitivity to deviation from this condition. For this purpose we test the recognition methods on the COIL20 database. The database is well known and frequently used for benchmarking. It contains 20 objects imaged from viewing angles ranging from -180 to $+180$. Similarly to the previous experiment, the frontal view of each object was used as the object model. As test images, we used 24 images of each object taken from different viewpoints in the ± 180 range. We plot the correct recognition rates for the methods under comparison in Fig 3.

Let us now elaborate on the above results. Referring to the SOIL47 results in Fig 2 the recognition rate for P-ARGOR overall is superior to P-AIFOR and AIFOR. For moderate scaling (scale factors 1 and 0.75) when an object is viewed from viewpoints close to frontal view the three methods perform similarly. Once the viewpoint considerably

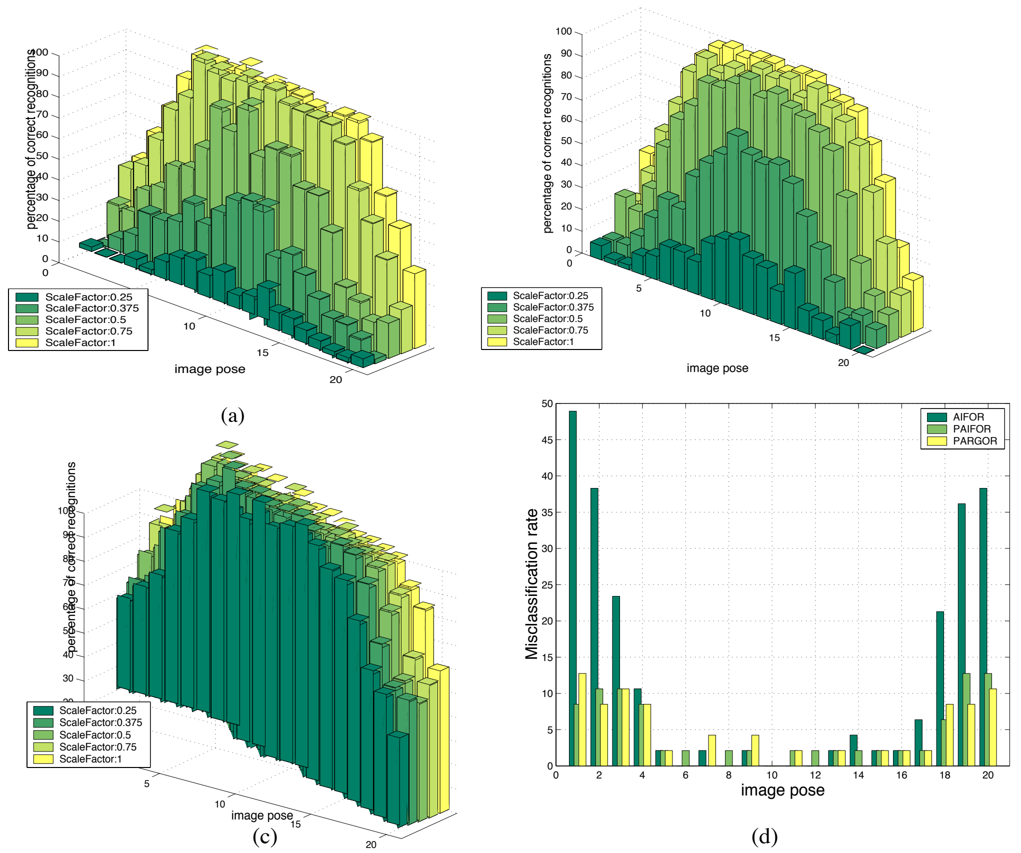


Figure 2: The rate of correct recognition for a) AIFOR b) P-AIFOR c)P-ARGOR for different resolutions of test images in SOIL47. d)The misclassification rate for the above methods (scale factor=1)

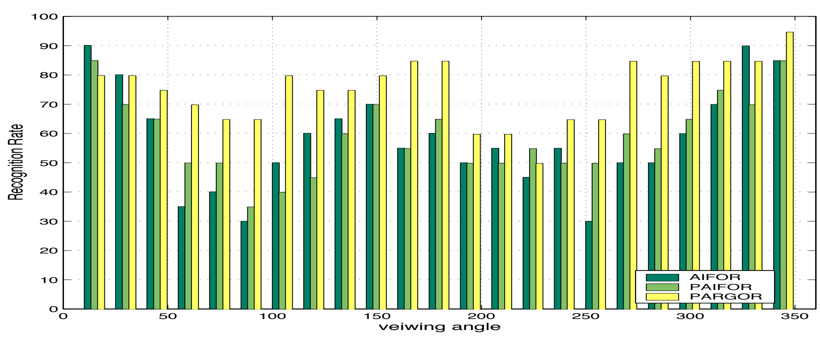


Figure 3: The relative correct recognition rates of the three methods on COIL20

deviates from the frontal view P-ARGOR performs better than two the other methods. Surprisingly in this situation AIFOR performs slightly better than P-AIFOR. The superiority of P-ARGOR to AIFOR stems from the use of context which provides more information for region matching. Although P-AIFOR also uses neighbourhood constraints in matching, since the intensity extrema are not invariant to viewpoint change the profile information is not reliable.

As the size of object in the test images becomes smaller the difference in performance between P-ARGOR and the other two methods becomes more notable. There are three major factors affecting AIFOR. First of all, the success of detecting local extrema depends on the size of object in the image. It is because both the support domain of the noise cancellation filter and the size of the search window for local extrema have to be adaptively selected based on the object size in the image. Secondly, the first extrema of function f_I (Eq (4)) used as the reference point along a ray is not a stable point under scaling. Thirdly the accuracy of moment invariants also depends on the image resolution or the size of objects in the image. In comparison to the elliptic regions, the segmented regions are less vulnerable to object scaling. As the results show, in this condition P-AIFOR performs better than AIFOR.

The main gain achieved by graph representation of elliptic regions in P-AIFOR is revealed by considering the misclassification rates in Fig 2(d). As the results show the misclassification rates for P-AIFOR and P-ARGOR in which matching is achieved by means of graphs is significantly better than the rate in AIFOR particularly once objects are viewed from severe viewpoints.This characteristic of P-ARGOR and P-AIFOR is the benefit of the way contextual information is used during the matching stage.

The results of the experiment on the COIL20 database in Fig 3 also show the superiority of P-ARGOR to the two other methods when objects have curved surfaces. In these circumstances the points with local intensity extrema do not remain stable in different views of an object. In fact the position of such points totally depends on the direction from which the scene is lit. Note that some of the objects in the COIL20 database are almost symmetric and do not have any texture on object surfaces. The process of image segmentation for these objects produces a number of regions which reflect surface shading. As a result, for different images of an object a number of regions will still be in correspondence. Fig 4 exemplifies this effect on one of the objects in the database.

The results in Fig 3 also show that as expected, the recognition rate of both methods falls off for objects imaged from close to ± 90 degrees with respect to the frontal view(object models). The recognition rate again increases for viewing angles close to ± 180 degrees because some objects in the database are almost symmetrical.

6 Conclusion

An experimental comparative study of two representation methods for recognition of 3D objects from a 2D view has been carried out. The methods investigated were the modified ARG region-based representation [1] and the elliptic region-based method of Tuytelaars et al[9]. We improved the binary measurements used in [1] by characterising the image along the line connecting the centroids of a pair of regions. The same source of information was made available to the other method. The modified methods referred to as P-ARGOR and P-AIFOR respectively were assessed from two points of view: the correct

(a) (b) (c)

Figure 4: a) The object model b) The segmented image(frontal veiw) c) The segmented image(45 degree rotation on space)

recognition and the misclassification rates. The result of the experiments showed that P-ARGOR is superior to P-AIFOR, particularly under significant scaling. The test on the COIL20 database showed that P-ARGOR is less sensitive than the P-AIFOR to deviations from surface planarity.

References

[1] A. Ahmadyfard and J. Kittler. Enhancement of ARG object recognition method. To appear in EUSIPCO 2002.

[2] Ahamdyfard A.R and Kittler J. Region-based object recognition: Pruning multiple representations and hypotheses. In *British Machine Vision Conference*, Bristol,UK, September 2000.

[3] Pope A.R. Model-based object recognition, a survey of recent research. Technical report, University of Columbia, 1994.

[4] http://www.ee.surrey.ac.uk/EE/VSSP/demos/colour/soil47/.

[5] J. Kittler and A. Ahmadyfard. On matching algorithms for the recognition of objects in cluttered background. In *Lecture Notes in Computer Science*, volume 2059 of *Springer*, 2001.

[6] J. Matas, J. Burianek, and Kittler J. Object recognition using the invariant pixel-set signature. *BMVC*, pages 606–615, 2000.

[7] C. Schmid and R. Mohr. Local grayvalue invariants for image retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(5), 1997.

[8] T Tuytelaars. *Local, Invariant Features For Registration and Recognition*. PhD thesis, KATHOLIEKE UNIVERSITEIT LEUVEN, Dec 2000.

[9] T. Tuytelaars and L. Van Gool. Wide baseline stereo matching based on local, affinely invariant regions. *BMVC*, pages 412–425, 2000.

[10] Christmas W.J., Kittler J., and Petrou M. Structural matching in computer vision using probabilistic relaxation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 749–764, 1995.