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.

BMVC 2002 doi:10.5244/C.16.34

In this paper we extend the use of binary measurements proposed in [1] to a more discriminative descriptor based on the image profile between the centroids of regions. We compare this new object representation referred to as Profile-based Attributed Relational Graph Object Recognition (P-ARGOR) with the Affine Invariant Feature Object Recognition (AIFOR) method [9]. In order to make the two methods comparable from the representation point of view, we use the same matching approach for the two methods. For this purpose we represent elliptic regions extracted from an image object in an ARG graph and use the image profile between a pair of regions as the binary measurement between the corresponding nodes in the graph. We refer to this method as Profile-based Affine Invariant Feature Object Recognition (P-AIFOR).

The reasons for selecting these two methods are multi-fold. First of all AIFOR of Tuytelaars et al [9] is an extension of the object recognition method of Schmid et al [7], an acknowledged benchmark against which the performance of other methods is measured. On the other hand ARGOR(or rather its earlier version [2]) has been shown to be the winner in a recent comparative study of matching algorithms for object recognition [5] involving geometric hashing, alignment and attributed relational graph. The attributed relational graph approach was shown to be superior both in terms of recognition accuracy and speed of interpretation. It is therefore of considerable interest to extend the comparison to the realm of object representation for which AIFOR is naturally the best candidate. We will show that the object representation in P-ARGOR is more robust than that in P-AIFOR particularly under severe scaling. The comparison between P-AIFOR and AIFOR reveals that the use of neighbourhood constraints reduces the misclassification rate through all experiments.

In the next section we describe our representation approach. In Section 3 the representations in AIFOR and P-AIFOR are explained. We describe the graph matching approach utilised for P-AIFOR and P-ARGOR in Section 4. In Section 5 we report the results of our experiments and discuss the main findings of the experiments. In the last section we draw the paper to conclusion.

2 Representation in P-ARGOR

In this method an object, or more specifically an image of the object is represented in terms of its segmented regions. The extracted regions are described individually and in pairs using their geometric and colour features. The entire image is then represented in the form of an Attributed Relational Graph (ARG) where each node corresponds to one of the regions and the edges between the nodes capture the region adjacency information. Each extracted region R_i is characterised individually using its (YUV) colour vector and we refer to this description as unary measurement vector $\overline{\mathbf{X}_\mathbf{i}}$. We describe the relation between region R_i and each of its k-nearest neighbours, R_j, using scalar AR_{ij} and vector $\overline{\mathbf{PS}_\mathbf{ij}}$. The AR_{ij} is the area ratio of the two regions which is an affine invariant measurement. As a binary measurement we define vector $\overline{\mathbf{PS}_\mathbf{ij}}$ to characterise the image along a line connecting the centroid of the two regions (R_i, R_j). The use of image profiles has been proposed before by Matas et al [6]. In their method an image of object is represented in terms of the image profiles defined between the image corners. Here we define a coarse measure of image profile instead of using the raw image intensity [6]. The motivation for this is the observation that the centroids of image regions are not accurate enough

to define a profile reliably. Thus we extract image profiles from the segmented images instead of the original images. In other words associated with each pair of regions we construct a vector which describes the line segments along the line connecting the two region centroids. Accordingly vector $\overline{\mathbf{PS}}_{\mathbf{ij}} = \{\mathbf{S_k}|k \in \{1 \cdots Z_{ij}\}\}$ describes Z_{ij} line segments along the line connecting centroid of regions R_i,R_j. Component $\mathbf{S_k}$ in this vector describes the k-th line segment in terms of its normalised length L_k and position $\mathbf{P_k}$ and its representative colour vector $\mathbf{C_k}$ in the YUV system. We normalise the length and position of the line segments by considering that the whole profile is of a unit length. This normalisation provides an affine invariant measure for binary relation. We use vector $\overline{\mathbf{A}}_{\mathbf{ij}} = \{AR_{ij}, \overline{\mathbf{PS}}_{\mathbf{ij}}\}$ to denote the binary relation between R_i and R_j. All the elements used in the binary measurement vector are affine invariant. Using the extracted regions and the associated measurement vectors we construct an Attributed Relational Graph in which a graph node O_i represents region R_i. The measurement vector, $\overline{\mathbf{X}}_{\mathbf{i}}$, is the node unary attribute. The binary measurement vector $\overline{\mathbf{A}}_{\mathbf{ij}}$ describes the link between the pair of nodes O_i, O_j.

Using this approach an object is modelled in the recognition system by an attributed relational graph constructed from its representative image. The graphs of all objects in the model database are collected in a composite graph referred to as the composite model graph. The content of an imaged scene is interpreted by constructing an ARG for the scene image. The resulting representation is referred to as the scene graph. Scene objects are then identified by matching the model and scene graphs.

The matching is achieved by measuring the similarity between unary and binary measurements and enforcing local consistency of interpretation by means of relaxation labelling. The dissimilarity between two unary measurements $\overline{\mathbf{X}}_{\mathbf{i}}$ and $\overline{\hat{\mathbf{X}}}_{\mathbf{k}}$ associated with nodes O_i and $\hat{O}_k$ in the scene and model graphs respectively is measured using the Euclidean distance:

$$UnDis(i,k) = Euclidean(\overline{\mathbf{X}}_{\mathbf{i}}, \overline{\hat{\mathbf{X}}_{\mathbf{k}}}) \tag{1}$$

The dissimilarity between two binary measurement vectors $\overline{\mathbf{A}}_{\mathbf{ij}}$ and $\overline{\hat{\mathbf{A}}_{\mathbf{mn}}}$ from the scene and model graphs respectively are measured as follows. For the area ratio components we use simply the difference between the two corresponding measurements. To measure the distance between line segment descriptors $\overline{\mathbf{PS}}_{\mathbf{ij}}$ and $\overline{\hat{\mathbf{PS}}}_{\mathbf{mn}}$, first we have to match the components of these vectors. As a result of imperfections in image segmentation, the numbers of line segments represented by these vectors are not necessarily identical. For this reason we have to match the vector components in a flexible manner yet this flexibility has to be controlled by imposing some loose constrains. Each component $\mathbf{S}_{\mathbf{p}}$ in $\overline{\mathbf{PS}}_{\mathbf{ij}}$ is linked to at most one component $\mathbf{S}_{\mathbf{q}}$ in $\overline{\hat{\mathbf{PS}}}_{\mathbf{mn}}$ if the distance measure $ComponentDis_{pq}$ is the lowest among the candidates and also below a pre-defined threshold Thr_{BinDis}:

$$ComponentDis_{pq} = W_L * |L_p - L_q| + W_P * |\mathbf{P_p} - \mathbf{P_q}| + W_C * \sum_{b \in \{U,V,Y\}} |\mathbf{C_p}(b) - \mathbf{C_q}(b)| \tag{2}$$

The weighting factors W_L, W_P and W_C are selected to adjust the sensitivity of the distance measure to different descriptors. Having accepted one correspondence the associated components will be deleted from the list of components in the two vectors ($\overline{\mathbf{PS}}_{\mathbf{ij}}$,$\overline{\hat{\mathbf{PS}}}_{\mathbf{mn}}$). At the end of this process we have a set of distance measures $\mathbf{DM}$ associated with the corresponding vector components. From this set the total dissimilarity between

two vectors $\overline{\mathbf{A}}_{\mathbf{ij}}$ and $\overline{\hat{\mathbf{A}}}_{\mathbf{mn}}$ is defined as:

$$\begin{aligned} BinDis(i,j,m,n) &= W_{DM} * \frac{1}{||\mathbf{DM}||} \sum_{\in DM} ComponentDis_{pq} \\ &+W_{AR} * |AR_{ij} - ARmn| + W_D * |1 - ||\mathbf{DM}||/Z_{ij}| \quad (3) \end{aligned}$$

In this formula the first term gives the average measure of dissimilarity between the line segments in the two profiles. The second term conveys the dissimilarity between area ratios AR_{ij} and AR_{mn} and we add the last term to the formula to penalise the unmatched components in $\overline{\mathbf{PS}}_{\mathbf{ij}}$. The weighting factors W_{DM}, W_{AR} and W_D are selected to adjust the sensitivity of the distance measure to different terms in the formula.

3 Representation in AIFOR and P-AIFOR

In the region extraction method proposed by Tuytelaars et al [9], the first step consists of selecting image salient points around which the intensity regions will be formed. As salient points the authors propose to use local image intensity extremas. In the region extraction step, for each detected local extrema the intensity function along certain rays emanating from the extrema is studied. Along each ray, the point at which the intensity function suddenly changes is invariant under affine transformation [8]. The point is detected by evaluating:

$$f_I(t) = \frac{abs(|I(t) - I_0|)}{\max(\frac{\int_0^t abs(|I(t)-I_0|)\,dt}{t}, d)} \quad (4)$$

instead of considering the intensity function, $I(t)$, along the ray directly. In this formula I_0 denotes the image intensity at the local extrema point (origin of rays), t is the Euclidean arc-length along the ray and d is a small number which has been added to prevent a division by zero.The local maximum in this function corresponds to the point of sudden intensity change.

Next, in AIFOR all respective local maxima found using Eq (4) along the rays originating from the same local image extrema are linked to delineate an intensity region. Such a region is invariant because its boundary points are invariant. As an extracted region may have an irregular shape, it is replaced by an elliptic region having the same shape moments up to the second order. Finally the size of the elliptic region is doubled to increase the distinctiveness of the intensity regions [8]. Each extracted intensity region is characterised using a feature vector consisting of nine moment invariants proposed in [8]. These invariants are defined in terms of pixel coordinates and associated colour intensities. The proposed invariants are rotation-invariant which are applied to an elliptic region after the region is normalised to a circle with unit radius.

3.1 Using Context in P-AIFOR Representation

Similar to ARGOR we represent the elliptic regions extracted from an object image in an ARG. Each node of this graph represents an elliptic region described by unary measurement vector $\overline{\mathbf{X}}$. We construct this vector using the nine features as in AIFOR.In order to

make the same source of information available to the two methods we describe the relation between an elliptic region and its k-nearest neighbours using the area ratio and the colour profiles. Unlike in P-ARGOR we use the raw profile information between the two regions. As the profile ending points we select the intensity extremum of the two regions which are close to the centroids of the regions. Recall that in AIFOR each elliptic region is formed at an intensity extrema point. Similar to ARGOR we define binary vector $\overline{\mathbf{A}}_{\mathbf{ij}}$ associated with a pair of elliptic regions R_i and R_j. This vector consists of scalar AR_{ij} and vector $\overline{\mathbf{P}}_{\mathbf{ij}} = \{\mathbf{C_k}|k \in \{1 \cdots Z\}\}$. The components of this vector, $\mathbf{C_k}$, are the colour features at equally spaced samples, Z, along the image profile.

We measure the dissimilarity between two unary vectors $\overline{\mathbf{X}}_{\mathbf{i}}$ and $\overline{\hat{\mathbf{X}}_{\mathbf{k}}}$ associated with nodes O_i and $\hat{O}_k$ in the scene and model graphs, using the Mahalanobis distance between the two vectors as proposed in AIFOR:

$$UnDis(i,k) = \overline{\mathbf{X}}_{\mathbf{i}} \mathbf{\Lambda}^{-\mathbf{1}} \overline{\mathbf{X}}_{\mathbf{k}} \tag{5}$$

The covariance matrix $\mathbf{\Lambda}$ is estimated by averaging the covariance matrices estimated from the feature vectors associated with a selected region in different images of an object. The dissimilarity between two binary vectors $\overline{\mathbf{A}}_{\mathbf{ij}}$ and $\overline{\hat{\mathbf{A}}}_{\mathbf{mn}}$ is measured as follows:

$$\begin{aligned} BinDis(i,j,m,n) = W_P * \sum_{k \in \{1 \cdots Z\}} & Euclidian(\overline{\mathbf{P}}_{\mathbf{ij}}(k), \overline{\hat{\mathbf{P}}}_{\mathbf{mn}}(\mathbf{k})) \\ & + W_{AR} * |AR_{ij} - AR_{mn}| \end{aligned} \tag{6}$$

The factors W_P and W_{AR} control the relative weight of the measurements in the formula.

4 Matching in P-ARGOR and P-AIFOR

As in both methods (P-AIFOR and P-ARGOR) the extracted regions of an image are represented in the ARG form, we adopt the same graph matching approach for the two methods. The only difference is in the definition of dissimilarity measurements as explained in the previous sections. In order to recognise objects in the scene image, the scene graph is matched against the composite model graph. This is in contrast with the methods in which the scene graph is matched against one object model at a time. By this matching strategy, we provide a unique interpretation for each part of the scene [1].

Before describing the algorithm for matching two ARGs let us introduce the necessary notation and the definitions required. We allocate to each node of the scene graph a label. Set $\theta = \{\theta_1, \theta_2, \cdots, \theta_N\}$ denotes the scene labels where θ_i is the label for node O_i. Similarly we use $\Omega = \{\omega_0, \omega_1, \cdots, \omega_M\}$ as the label set for the nodes of the composite model graph. In this label set, ω_0 is the null label which does not refer to any real node. It is added to be assigned to the scene nodes for which no other label in Ω is appropriate [10].

The contextual information in a graph is conveyed to a node from a small neighbourhood. In this regard, node O_j is a neighbour of O_i if the Euclidean distance between the associated regions is below a predefined threshold. We use set $\mathcal{N}_i$ to refer to the nodes in a neighbourhood of O_i. Similarly the labels in the neighbourhood of ω_α are referred by set Ω_α.

By labelling we mean the assignment of a proper label from set Ω to each node of the scene graph. In this regard, $P(\theta_i = \omega_\alpha)$ denotes the probability that node O_i in the scene

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]:

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

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

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

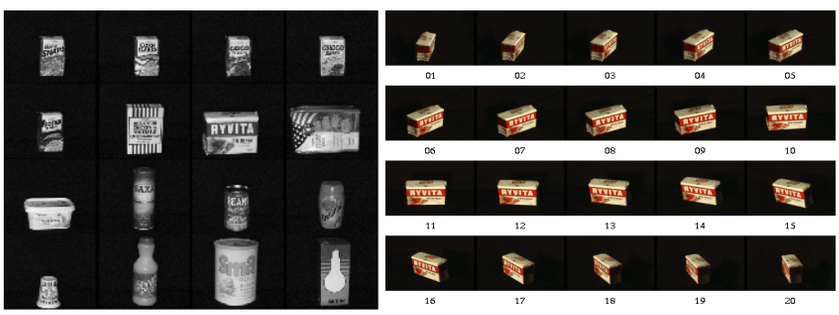
The relaxation labelling technique updates the labelling probabilities in an iterative manner using the contextual information provided by the nodes of the graph. In this formulation $Q(\theta_i = \omega_\alpha)$ is the support function which measures the consistency of the label assignments to the scene 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

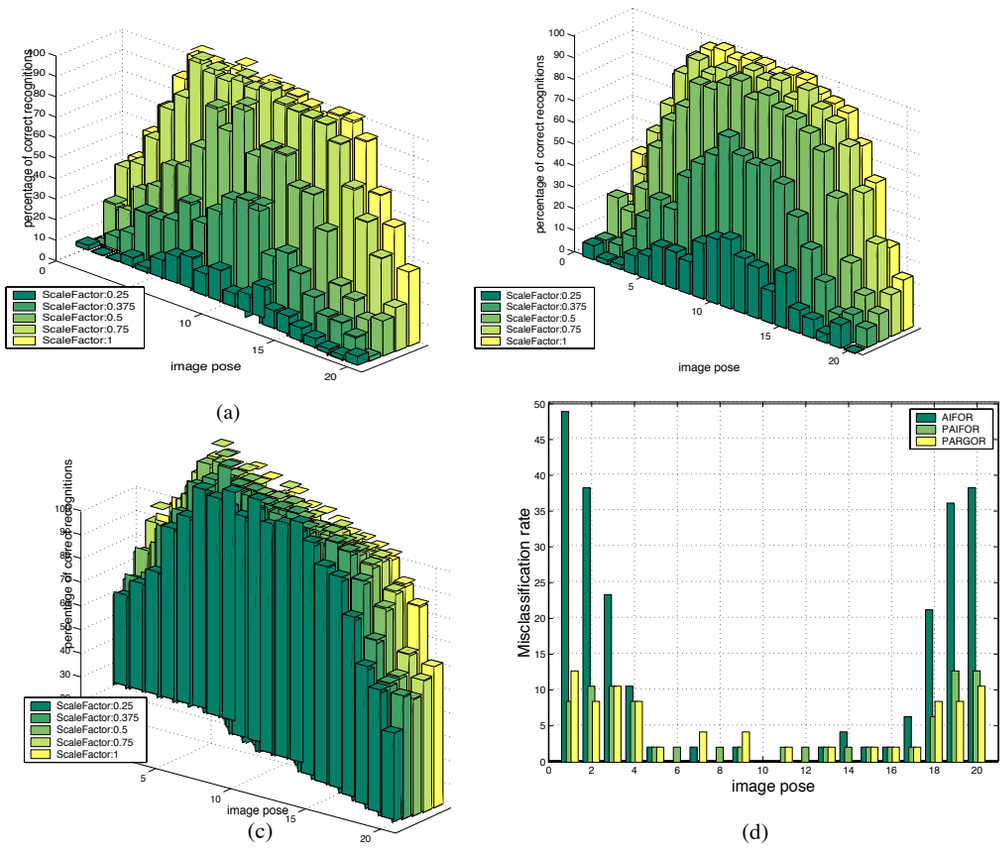


Figure 2: The rate of correct recognition for a) AIFOR b) P-AIFOR c)P-ARGOR for different resolutions of test images in SOIL47. d)The misclassification rate for the above methods (scale factor=1)

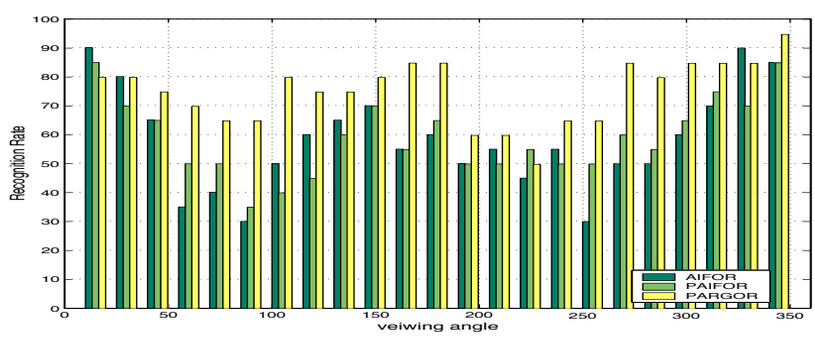


Figure 3: The relative correct recognition rates of the three methods on COIL20

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.

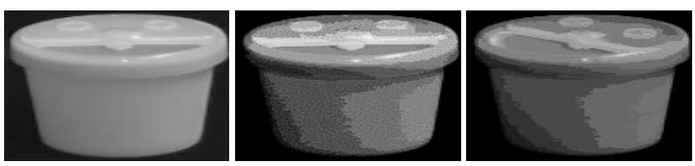
The main gain achieved by graph representation of elliptic regions in P-AIFOR is revealed by considering the misclassification rates in Fig 2(d). As the results show the misclassification rates for P-AIFOR and P-ARGOR in which matching is achieved by means of graphs is significantly better than the rate 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.