Enhancement of ARG Object Recognition Method

ABSTRACT

We address the problem of object recognition in computer vision. The Attributed Relational Graph (ARG) method proposed in [**?**] is modified to extend its applicability under object scaling condition. This consists in using representation stage, unary and binary attributes which do not involve measuring the shape of regions.The experimental result shows the new method can cope with object scaling better.In the new method we also modify the probabilistic relaxation labelling method used in[**?**] in order to minimise its sensitivity to the number of spurious nodes in the contextual neighbourhood.

1 Introduction

The recognition of objects in computer vision has many applications.The methods which address this problem are classified into two categories[4]: feature-based and appearance-based approaches.In feature-based methods the model and scene are represented using geometric features extracted from the images. In contrast appearance-based methods use a collection of 2D views of objects. Each image of an object is then represented using raw pixel information[1].The appearance-based methods have two major advantages. First of all, they do not need to extract geometric features from the image. Second, the manner in which they represent objects is applicable to a wide class of objects[5]. Although appearance-based techniques have some positive merits, they cannot cope with occlusion and local distortion problems.
In an earlier work [**?**] we proposed a recognition method in which each object is represented in terms of its image regions. The regions extracted are normalised in an affine invariant manner.The normalised regions of the image are 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[**?**]. The regions are characterised using shape and region appearance properties. Object recognition is achieved by comparing the scene ARG to the graph of object models using relaxation labelling. The results obtained in several experiments involving images of objects taken from different viewing angles in a cluttered environment were very promising. However for objects imaged from extreme viewing angles, and also under sever scaling, our method tends to fail (like most of the existing methods).A further investigation revealed that in these situations the shape of a segmented region is rather unreliable.
In this paper we extend the ARG approach so that it can cope with such extreme conditions.For this purpose we propose a new ARG representation in which we utilise unary and binary measurements without using any shape information.The additional benefit of this representation is that we do not need to normalise the regions. In the new method we also modify the probabilistic relaxation labelling method used in[**?**] in order to minimise its sensitivity to the number of spurious nodes in the contextual neighbourhood.The experimental results show the new method can better cope with object scaling. However the recognition results for objects with a few extracted regions is worse than those of the original method.The results also show that a significant improvement in terms of the number of incorrect matches is achieved.
In the next two sections we overview the ARG method[**?**] from the representation and matching points of view. We introduce our new representation and matching in Section 4.Section 5 demonstrates the experimental results.We draw the paper to conclusion in Section 6.

2 Representation in the original method

In this section we first overview the invariant representation of the scene and model images which was proposed in[**?**]. In this representation an image of the scene or model is regarded as a set of regions.For this purpose image of an object is segmented using region growing approach[2].Each region of the image is transformed to a normalised space in which the corresponding regions of model and scene have identical appearance subject to noise.Assuming that two reference points are selected for each region, the affine transformation which normalises region R is computed by imposing the following constraints:

1. the reference points (x_0, y_0) and (x_1, y_1) of the region R are to be mapped to points $(1, 0)$ and $(0, 0)$ of r respectively.

2. the normalised region r is to have a unit area and the second order cross moment equal to zero.

Recalling that we need two reference points to normalise each region, the centroid of the region is chosen as one of the these points. The ways the second reference point is selected are different for scene and model. For a region in the object model the highest curvature point on the boundary of the region is chosen as the second reference point, while in the scene for each region a number of points of high curvature are picked.

Normalised regions of an image are represented in the form of an Attributed Relational Graph.All object models are represented jointly in an ARG referred as the composite model graph. Each node of this graph, $\breve{a}_i$, represents a normalised region which is described by unary measurement vector $\breve{\mathbf{x}}_\mathbf{i}$. This vector consist of two components: vector $\breve{\mathbf{S}}_\mathbf{i}$ contains a number of equally spaced samples on the boundary of the associated normalised region and the representative colour of the region expressed in form of the YUV coordinate system $\breve{\mathbf{C}}_\mathbf{i}$. Based on the adjacency of the regions in an image, each node $\breve{a}_i$ has a number of neighbours listed in set $\breve{\mathcal{N}}_i$.There is an edge between $\breve{a}_i$ and each of its neighbours,$\breve{a}_j$. The edge is described by binary measurement vector $\breve{\mathbf{A}}_{ij}$ consisting of three components: binary matrix $\mathbf{B_{ij}}$, $\overline{ColorDis_{ij}}$ and $AreaRatio_{ij}$. The matrix, $\mathbf{B_{ij}}$, is defined in terms of the transformations by which the associated regions $\breve{a}_i$ and $\breve{a}_j$ are normalised,i.e. $\mathbf{B}_{ij} = {T_i}^{-1}T_j$. This measurement is affine invariant[**?**]. The vectors $\overline{ColorDis_{ij}}$ and scalar $AreaRatio_{ij}$ describe the colour relations and area ratios of the corresponding regions respectively.

Similarly to the object models the scene image is represented in an ARG referred as the scene graph. In the scene ARG the unary and binary measurements are defined similar to the model ARG measurements. The only difference lies in the use of the multiple representation of the scene nodes. Since each scene node is multiply represented, associated with scene node a_i there is an array of unary measurement vectors $\mathbf{x}_i^k$ and similarly each pair of neighbours, $\breve{a}_i$ and $\breve{a}_j$, is described using an array of binary relation vectors $\mathbf{A}_{ij}^{kl}$. In other word we have: $\underline{\mathrm{x}}_i = \{\mathbf{x}_i^k | k \in \{1, \cdots, L\}\}$ and $\underline{\mathrm{A}}_{ij} = \{\mathbf{A}_{ij}^{kl} | k, l \in \{1, \cdots, L\}\}$ where, L , denotes the number of representations used for the scene regions.

3 ARG matching

The graph matching in[**?**] is regarded as the problem of assigning a label, by association of the node with a node in the model graph to each node of the scene graph. During model building each node of the model ARG is allocated a label representing an object primitive.The set of model labels is denoted by $\Omega = \{\omega_0, \omega_1, \cdots, \omega_M\}$. The label ω_0, called null label, is added to the set as a wild card for assigning to the scene nodes for which no other label is appropriate[6]. The labelling of the scene graph is accomplished in two stages: first, the best representation of each scene node under a particular label assignment hypothesis is selected; second, the label probabilities are updated by incorporating contextual information. In the first stage, for each scene node, a list of candidate labels based on the similarity of the unary attributes of the scene node and the model nodes is constructed. Simultaneously for each label in this list the best representation of the scene node is determined. It is based on the measurement of the mean square distance between the normalised region boundary points of the scene region and those of the hypothesised label. At the end of this process a label list,Ω^i, for each object,a_i, with the best representation for each label in the list is provided.

The second stage of matching is performed using the relaxation technique proposed in[6] which was adapted to our task. Two major modification were proposed: the use of sum operator instead of product in the support function, and label pruning at the end of each iteration. The first modification was motivated by the fact that the product support function derived in [6] is not applicable due to the scene clutter which may drive the total support to zero. For this reason we have adopted the benevolent sum support function to measure the supporting evidence from the neighbouring objects as in [3]. We proposed the label pruning at the end of each iteration to speed up the algorithm convergence.

In the relaxation technique[6] all the possible label assignments for each node ,a_i, are considered. The probability of assigning a label to a scene node is initially computed by measuring the similarity between the unary measurement vector of the hypothesised label and the scene node. The label probabilities are then iteratively updated using their previous values and supports provided by the neighbouring nodes.The modified iteration formula is given as:

$$p^{(n+1)}(\theta_i = \omega_{\theta_i}) = \frac{p^{(n)}(\theta_i = \omega_{\theta_i}) Q^{(n)}(\theta_i = \omega_{\theta_i})}{\sum_{\omega_\lambda \in \Omega_i} p^{(n)}(\theta_i = \omega_\lambda) Q^{(n)}(\theta_i = \omega_\lambda)} \tag{1}$$

$$Q^{(n)}(\theta_i = \omega_\alpha) = \sum_{j \in \mathcal{N}_i} \sum_{\omega_\beta \in \{\Omega^i \cap \Omega_\alpha\}} p^{(n)}(\theta_i = \omega_\beta) p(\mathbf{A}_{ij}^* | \theta_i = \omega_\alpha, \theta_j = \omega_\beta) \tag{2}$$

where function Q quantifies the support for the assignment of label ω_α to a_i, received from the neighbours of object i in the scene at the nth iteration step. Set Ω_α denotes the labels whose associated nodes are listed in $\breve{\mathcal{N}}_\alpha$.

In the support function, Q, the term $p(\mathbf{A}_{ij}^* | \theta_i = \omega_\alpha, \theta_j = \omega_\beta)$ is the probability distribution of the binary relation vector $\mathbf{A}_{ij}^*$ given the matches $\theta_i = \omega_\alpha$ and $\theta_j = \omega_\beta$.Note that $\mathbf{A}_{ij}^*$ is the binary measurement vector associated with the pair nodes a_i,a_j given for the best representation of the two nodes. It is assumed that the distribution function is centred on the model binary measurement $\breve{\mathbf{A}}_{\alpha\beta}$ and deviations from this mean are modelled by a Gaussian, i.e.

$$p(\mathbf{A}_{ij}^* | \theta_i = \omega_\alpha, \theta_j = \omega_\beta) = \mathcal{N}_{\mathbf{A}_{ij}^*}(\breve{\mathbf{A}}_{\alpha\beta}, \Sigma_b) \tag{3}$$

where Σ_b is the covariance matrix of the binary measurement vector $\mathbf{A}_{ij}^*$.

The iterative process will be terminated if either in the last iteration none of the probabilities changed by more than a given threshold or the number of iterations reached some specified limit.

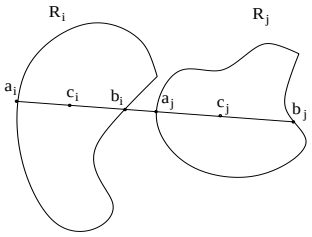


Figure 1: Binary measurements associated with pair of regions

4 New ARG representation and matching

The shape of a region is one of the main descriptors by which the region can be represented. The distinctiveness of this feature is the main motivating factor for using it as a cue for object recognition. However, poor image segmentation may distort the boundaries of segmented regions.As a result, the shape information becomes unreliable. This becomes a particularly serious problem in situations such as sever image scaling or object pose transformation. In order to extend the ARG object recognition method to such sever conditions we propose a new representation in which shape of regions is no longer utilised. Each extracted region R_i is individually described using its representative colour vector in the YUV space, which is referred as unary measurement vector $\mathbf{x_i}$. The relationship between a pair of regions R_i, R_j which is described using geometric and colour measurements contributing the binary measurement vector, $\mathbf{A}_{ij}$, defined as follows: Let us consider a pair of regions R_i and R_j in Fig 4. The line which connects the centroid points c_i and c_j intersects the regions boundaries at a_i, b_i, a_j and b_j. Under affine transformation assumed here, the ratio of the segments on the line remains invariant. Using this property, we define $m_1 = \frac{a_i a_j}{c_i c_j}$ and $m_2 = \frac{b_i b_j}{c_i c_j}$ as two elements of the binary measurement vector. In addition, the area ratio $AreaRatio = A_i/A_j$ and the distance between colour vectors $\overline{ColourDis} = \overline{C_i} - \overline{C_j}$ are used as complementary components of vector $\mathbf{A}_{ij}$. All the elements used of the vector are affine invariant.

As before the matching is accomplished in two stages.Unlike in the previous method, the first stage of matching is a simple task.The list of admissible labels is created by measuring the Euclidean distance between the colour vector of a candidate label and that of the scene node, a_i. If the distance is below a predefined threshold the label is added to the list of candidate labels Ω^i associated with the scene node.

In the second stage we apply the relaxation technique[6] with the support function modified as follows:

$$Q^{(n)}(\theta_i = \omega_\alpha) = \prod_{j \in \mathcal{N}_i} \{ \sum_{\omega_\beta \in \{\Omega^i \cap \Omega_\alpha\}} P^{(n)}(\theta_j = \omega_\beta) P(\mathbf{A}_{ij} | \theta_i = \omega_\alpha, \theta_j = \omega_\beta) + \sum_{\omega_\beta \in \Omega^i - \{\Omega^i \cap \Omega_\alpha\}} P^{(n)}(\theta_j = \omega_\beta)\eta \} \quad (4)$$

The support function differs from the previous formula in two respects. First we find an alternative way to alleviate the problem with the product support function[6] which drives the total support to zero. Our proposal is to set the binary distribution function to a constant value, η_{PDF}, when the function is below η_{PDF}. This threshold is estimated experimentally by inspecting the binary relation distribution function for the corresponding pair of nodes. Although the sum support function in(2) also tries to deal with the above problem, its tendency to collect incoherent support from clutter contributions is undesirable.

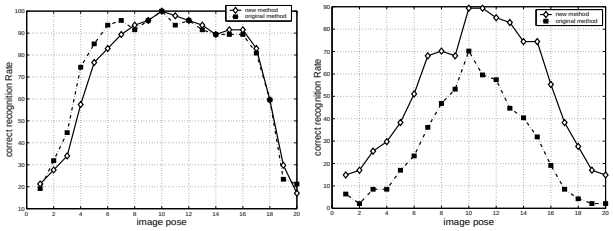
Second, we measure the consistency of labelling node a_j (a neighbour node of a_i) in the context of the assignment ω_α to a_i using two terms: the first term measures the contribution from ω_α neighbours whose binary relation with ω_α is available (the main support) and the second term is added to balance the number of contributing terms via the other labels in Ω. The latter part is added to reduce the undesirable effect of the number of contributing terms to the support for each neighbouring node a_j. The parameter η in this term plays the role of the binary relation distribution function $P(\mathbf{A}_{ij}|\theta_i = \omega_\alpha, \theta_j = \omega_\beta)$. Its value is constant for all the pairs of model nodes ω_α and ω_β that are not neighbours. Note that this value is set to be identical to the threshold η_{PDF}.

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 this measure is used as an object matching score.

5 Experimental Results

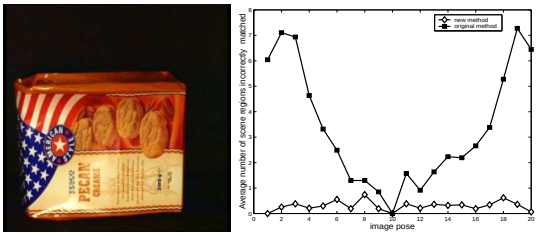
We designed two experiments to investigate the performance of the new method in comparison with the previous method. The first experiment was carried out on SOIL-47 database(Surrey Object Image Library) which contains 47 objects.Each object has been imaged from 21 viewing angles spanning a range of up to ± 90 degrees.Fig 3(a) shows the frontal view of an object in this database. The database is available online[7].In this experiment we model each object using its frontal image while the other 20 views of the objects are used as test images. 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 noise level in the original images was determined by manually delineating a set of homogeneous regions from which we estimated the parameters of a Gaussian noise model in the respective colour channels.The scaling parameter was chosen so as to produce test image size of 37.5% of the original image set. Note that throughout the experiment we used the full size images as the object models.

In fig 2 the recognition rate for two methods are plotted



(a) (b)

Figure 2: The Correct recognition rates for the two methods a) without object scaling b) scaled by factor 0.375



(a) The frontal view of an object in soil-47 database

(b) Average number of scene regions which are not correctly matched

Figure 3:

as a function of object pose.The results show that when the size of objects in the test and model images is comparable, the previous ARG, which benefits from the distinctiveness of the region shape features, performs better than the proposed method.In contrast, when objects size in the scene becomes considerably small, the performance of the proposed method is very good whereas the previous method fails.In the proposed method the contextual information seems to suffice to interpret the scene correctly. As a second criterion for comparison, we measured the number of scene regions incorrectly matched. In fig3(b) this measure is plotted for the two methods. As the results reveal the rate of incorrect matching for the new method is remarkably low. This success derives from two factors. First, the undesirable effect of the number of node neighbours contributing to contextual information is reduced. Second the use of the product operator in the support function as proposed in [6] with our modification, we avoid accumulating spurious support from incoherent contextual information.

In second experiment we tested the two methods on images of traffic signs.In this experiment we use 50 images of 8 traffic signs as test images.Fig 4 shows two samples of these images. As in the previous experiment a frontal image of each object is used as the object model. Using the original method, the object(s) in the test images are correctly recognised in 87% of cases whereas this rate declines to 80% for the new method. This is not very surprising. The objects in the soil-47 are more complex than the traffic signs. As a result the large number of regions constituting an object in

(a) (b)

Figure 4: Two test images of the traffic signs

soil-47 allows the method to rely on contextual information in the matching stage. This in not the case for traffic sign objects for which only a few regions are extracted.

6 Conclusion

The Attributed Relational Graph (ARG) method proposed in [**?**] was modified to cope with object scaling.We proposed a new representation does not involve measuring region shape.We confirmed experimentally that this modification improved the robustness.However the experiments revealed that when an object ha a small number of regions,the original approach was superior. We also modified the support function proposed in the original method to alleviate the undesirable dependence of the contextual support on nodes. The matching results showed that a significant reduction in incorrect matches was achieved.

References

[1] Leonardis A. and Bischof H. Robust recognition using eigenimages. *Computer Vision and Image Understanding*, 78(1):99–118, 2000.

[2] R. Haralick and L. Shapiro. Image segmentation techniques. *Computer Vision, Graphics and Image Processing*, pages 100–132, 1985.

[3] Kittler J. and Hancock E.R. Combination evidence in probabilistic relaxation. *International Journal of Pattern Recognition and Artificial Intelligence*, 3(1):29–51, 1989.

[4] Pope R. Model-based object recognition a survey of recent research. Technical report, 1994.

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

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

[7] www.ee.surrey.ac.uk/EE/VSSP/demos/colour/soil47/.