Region-Based Object Recognition: Pruning Multiple Representations and Hypotheses

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

We address the problem of object recognition in computer vision. We represent each model and the scene in the form of Attributed Relational Graph. A multiple region representation is provided at each node of the scene ARG to increase the representation reliability. The process of matching the scene ARG against the stored models is facilitated by a novel method for identifying the most probable representation from among the multiple candidates. The scene and model graph matching is accomplished using probabilistic relaxation which has been modified to minimise the label clutter. The experimental results obtained on real data demonstrate promising performance of the proposed recognition system.

1 Introduction

Recognising known objects in cluttered scenes has been the principal goal of computer vision since its inception. The object recognition problem is inherently difficult due to noise and the effect of other factors which include geometric transformation of the measurements, geometric distortion, occlusion and clutter. In this paper we address the problem of object recognition using the model based approach. We take the view that objects can be represented as a union of planar surfaces. This modelling is completely appropriate for the recognition of polyhedral objects as well as objects which include some planar patches in their surfaces. Furthermore, the method is applicable to nonplanar objects in the cases when the deviation from planarity of an object face can be absorbed into other geometric distortions.

Model based object recognition involves two major problems, namely that of object representation and the closely related problem of object matching. A large number of representation techniques have been proposed in the computer vision literature. Despite their essential differences they can be broadly classified into two categories: feature based, and appearance based. Although appearance based techniques[10] have positive merits, they can not cope with occlusion and local distortion problems. For this reason we have opted for a feature based representation.

The matching process tries to establish a correspondence between the features of an observed object and a hypothesised model. The various object recognition techniques proposed in the literature deploy different mechanisms for matching the scene and model

BMVC 2000 doi:10.5244/C.14.75

features. They also differ in terms of how a consistent interpretation is derived from scene-model feature matches and how the pose is estimated from a consistent interpretation. The techniques range from the alignment methods [4][5] to geometric hashing [7] [6] or hough transform methods [9]. In an earlier work [11] it was argued that an effective object recognition method should be based on the extraction of relatively simple features as only such features can be reliably detected in complex images. The distinctiveness of such features can be enhanced by relational measurements. As a suitable tool to achieve matching the evidence combination method of relaxation labelling was adopted. The method uses an attributed relational graph for representing both scene and model. Only unary and binary relations which are made invariant to any pertinent geometric transformation group are employed. The actual implementation of the method used as features interest points found on the texture boundaries. The method was shown to work well in experiments involving both synthetic 2D and real 3D objects in cluttered backgrounds. However, it was found that the recognition performance of the system decreased for complex scenes where the size of objects becomes inevitably small. The main reason for the degradation was the failure to extract interest points reliably. It was noted that in contrast to interest points, regions remain stable over a large range of scales.

In this paper we use region based features for attributed relational graph representation of scene and model objects as suggested in[12] [2]. As in [2] we make the representation more resilient to occlusion by using multiple representations for each node of the scene graph. However, in contrast to our previous work we do not involve all the representations of the scene graph in the iterative relaxation labelling algorithm used for matching. Instead we adopt the most probable representation for each scene node before starting the labelling process. A methodology for identifying this representation from the multiple candidates is proposed. A further speed up of the matching algorithm is gained by hypothesis pruning at each step of the matching process.

The paper is organised as follows. In the following Section we described the proposed representation. The relaxation algorithm adopted for the attributed relational graph matching is described in Section 3. The experiments carried out and the results obtained are presented in Section 4 . Section 5 draws the paper to conclusion.

2 Representation

We first overview the invariant representation of the scene and model images adopted. It is based on regional features. Accordingly we regard an image of the scene or model as a set of regions. Then for each region we provide a basis matrix which allows us to transform the region to a normalised space in which the corresponding regions of model and scene should have identical appearance subject to noise. In this manner we construct an Attributed Relational Graph in which normalised scene regions are considered as graph nodes and binary relations between neighbouring region pairs constitute graph edges. It has been shown in [12] that a matrix $\mathbf{B}$ which possesses the following properties can be used to transform a region to an invariant space:

1. $\mathbf{B}$ is a non-singular matrix.
2. Matrices $\mathbf{B}$ and $\mathbf{B}'$ associated with corresponding regions R, R' respectively are related as $\mathbf{B}' = \mathbf{B}\mathbf{T}$ where $\mathbf{T}$ is a transformation matrix which maps R to R'. In other word $\mathbf{R}' = \mathbf{R}\mathbf{T}$.

Such a matrix is called basis matrix.
Using the basis matrix, $\mathbf{B}$, barycentric coordinates of an arbitrary point P of region R can be defined as :

$$C_{\mathbf{B}}(P) = P\mathbf{B^{-1}} \tag{1}$$

The defined coordinate system is transformation group invariant since an arbitrary point P of region R and the corresponding point P' of region R' have the same barycentric coordinates:

$$C_{\mathbf{B}'}(P') = C_{\mathbf{B}T}(P\mathbf{T}) = (P\mathbf{T})(\mathbf{BT})^{-1} = C_{\mathbf{B}}(P) \tag{2}$$

The barycentric coordinates can be used as unary relations of region R. Similarly, a binary relation matrix $\mathbf{B_{ij}}$ associated with a pair of regions R_i and R_j can be defined as $\mathbf{B}_{ij} = \mathbf{B}_i\mathbf{B}_j^{-1}$. Let $\mathbf{B'_{ij}}$ be a binary relation matrix for regions R'_i and R'_j in the scene which correspond to regions R_i and R_j respectively of a model, we can readily show that:

$$\mathbf{B'_{ij}} = \mathbf{B'_i}\mathbf{B'^{-1}_j} = (\mathbf{B_i T})(\mathbf{B_j T})^{\mathbf{-1}} = \mathbf{B_i T T^{-1} B_j^{-1}} = \mathbf{B_{ij}} \tag{3}$$

Thus the binary relation matrices are transformation group invariant as well.
To find the basis matrix of an arbitrary region R, we initially address the problem of finding an affine transform that modifies the region R to region r in a normalised space where the modified corresponding regions are comparable.In order to uniquely define such transform we impose the following constraints on it:

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.

To simplify the transformation task we split it into two sub-tasks. First, using a similarity transformation matrix $\mathbf{T}_s$, the reference points (x_0, y_0) and (x_1, y_1) of R are mapped to points $(1, 0)$ and $(0, 0)$ in the new region R_0 respectively. The matrix $\mathbf{T}_s$ can readily be shown to be:

$$\mathbf{T}_s = \frac{1}{k_n}\begin{pmatrix} x_1 - x_0 & y_0 - y_1 & 0 \\ y_1 - y_0 & x_1 - x_0 & 0 \\ x_0^2 + y_0^2 - y_0 y_1 - x_0 x_1 & x_0 y_1 - y_0 x_1 & k_n \end{pmatrix} \tag{4}$$

where $k_n = (x_1 - x_2)^2 + (y_1 - y_2)^2$ is the square of distance between the two reference points. Second, we determine the affine transform, $\mathbf{T}_a$, that modifies new region R_0 to the normalised region r. Such a matrix can be calculated taking into account the relations between the second order moments of R_0 and r :

$$\mathbf{T}_a = \begin{pmatrix} 1 & 0 & 0 \\ -u_{1,1}/u_{0,2} & k & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{5}$$

where $u_{1,1}, u_{0,2}$ are the second order moments of the normalised region r and $k = k_n/(Area \quad of \quad region \ R)$. The transformation matrix that will map region R to the normalised region r will then be given as $\mathbf{T} = \mathbf{T}_s\mathbf{T}_a$.

Returning to the main problem, it is easy to check that the matrix $\mathbf{T}^{-1}$ possesses the basis matrix properties.Thus we can define the basis matrix transformations as :

$$\mathbf{B} = \mathbf{T}_a{}^{-1}\mathbf{T}_s{}^{-1} \tag{6}$$

Thus for each region the basis matrix $\mathbf{B}$ is determined in terms of the local $(\mathbf{T}_s)$ and global$(\mathbf{T}_a)$ geometric features of the region. In other words the basis matrix of a region captures geometric features of the region. This information from each pair of regions, R_i, R_j, is conveyed to the binary relation matrix, $\mathbf{B_{ij}}$, assosiated with the region pair.
In this representation we regard the given images as a set of regions and consider the centroid of a region as one of the required reference points. The way the second reference point is selected is different for scene and model. In the case of the 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 and consequently a number of representations for each scene region are provided. The selection of more than one point on the boundary is motivated by the fact that an affine transformation may change the ranking and distort the position of high curvature points.
Now we represent all of the model images in a graph as $\breve{\mathbf{G}} = \{\Omega, \breve{\mathcal{X}}, \breve{\mathcal{A}}\}$ and refer to it as the model graph. In this graph $\Omega = \{\omega_1, \omega_2, \cdots, \omega_M\}$ denotes the set of nodes (normalised regions) and $\breve{\mathcal{X}} = \{\breve{\mathbf{x}}_1, \breve{\mathbf{x}}_2, \cdots, \breve{\mathbf{x}}_M\}$ is a set of unary measurement vectors where for each node, ω_i, we have a vector of measurements $\breve{\mathbf{x}}_i$. Furthermore, the binary measurement set $\breve{\mathcal{A}} = \{\breve{A}_{ij}|(i,j) i \in \{0, \cdots, M\}, j \in \breve{\mathcal{N}}_i\}$ represents a set of binary measurement vectors associated with the pair of neighbouring nodes where the neighbouring set $\breve{\mathcal{N}}_i$ associated with the node, ω_i, contains all of the nodes corresponding to the regions in a neighbourhood of the node ω_i.
We consider the vector $\breve{\mathbf{x}}_\mathbf{i} = \left[\ \mathbf{S}, \mathbf{C}\ \right]$ as a unary attribute vector associated with the node,ω_i, where the components of vector $\mathbf{S}$ are the coordinates of a number of equally spaced samples on the boundary of the ith normalised region and vector $\mathbf{C}$ as a representative of the region colour expressed in terms of the YUV components. The binary measurement vector $\breve{\mathbf{A}}_{ij}$ associated with the pair nodes, ω_i, and, ω_j, is defined as follows:

$$\breve{\mathbf{A}}_{ij} = \left[\ \mathbf{B_{ij}}, ColorRelation_{ij}, AreaRelation_{ij}\ \right]$$

where $\mathbf{B}_{ij}$ denotes the binary relation matrix associated with the region pair and vector $ColorRelation_{ij}$ and scalar $AreaRelation_{ij}$ express the colour relations and area ratios of the region pair respectively.
Similarly, the graph, $\mathbf{G} = \{\mathbf{a}, \mathcal{X}, \mathcal{A}\}$, represents the scene image where $\mathbf{a} = \{a_1, \cdots, a_N\}$ is the set of scene graph nodes and, $\mathcal{X}$, and, $\mathcal{A}$, denote the set of unary and binary measurements respectively. Note that the representation of the scene differs from that of the model in the sense that for the scene representation more than one bases are provided for each scene region. The multiple representation for each scene node is defined in terms of a set of unary measurement vectors, $\mathbf{x}_i^k$, which possess the same components as the model ones, with index k indicating that the vector is associated with the kth representation of the ith node. Also related to each neighbouring pair of the regions a_i and a_j binary relation vectors $\mathbf{A}_{ij}^{kl}$ are defined with the properties of the model binary vectors . The multiple unary measurement vectors $\mathbf{x}_i^k$ and binary relations $\mathbf{A}_{ij}^{kl}$ constitute the combined unary and binary relation representation $\mathbb{x}_i = \{\mathbf{x}_i^k | k \in \{1, \cdots, L\}\}$ and $\mathbb{A}_{ij} = \{\mathbf{A}_{ij}^{kl} | k, l \in \{1, \cdots, L\}\}$ where, L, denotes the number of representations used for the scene regions.

3 Matching

For matching we have adopted the relaxation labelling technique of [1, 8, 13] and adapted it to our task. The problem considered is much more complex than the previous application of relaxation methods due to a large number of nodes in the model graph. Similarly to [13] we add a null label to the label set to reduce the probability of incorrect labelling. The essential difference in our matching problem is that the product support function derived in [13] is not applicable due to the scene clutter driving the total support to zero, thus masking the coherent support from consistently labelled objects. For this reason we have adopted the benevolent sum support function to measure the supporting evidence from the neighbouring objects as in [8].

In the previous work [2] we considered all the available representations of the object throughout the labelling process. This made the method complex and time consuming particularly when the number of model nodes increased.To reduce the complexity of the problem, we divide the matching process into two stages: first, finding the best representation of object under a particular label hypothesis and second, updating the label probabilities by incorporating contextual information. In the first stage, we compare the unary attribute measurement of each object of the scene, a_i, with the same measurement for all the admissible interpretations and construct a list, $\mathbf{L}_i$, containing the labels which can be assigned to the object. Simultaneously for each label in this list we find the best representation. For this purpose we propose a novel method of assessing each representation. It is based on the measurement of the mean square distance between the normalised region boundary points stored in a vector, $\mathbf{S}$, and the unary attribute vector of the hypothesised label. In other words the merit of kth representation of object a_i in the context of the assignment of label ω_α to that object is evaluated as:

$$E(\theta_i = \omega_\alpha) = \min_k \left(\mathbf{x}_i^k[\mathbf{S}] - \check{\mathbf{x}}_\alpha[\mathbf{S}]\right)^2$$

In ideal conditions, for the correct basis, the above measurement for the corresponding regions would be zero. However, due to errors in the extraction of the reference points and as a result of the segmentation noise affecting the boundary pixel positions this measurement is subject to errors. Thus criterion function E is compared against a predefined threshold. A label is entered in the list of hypotheses only if the measurement value is less than the threshold. The best representation basis, k, is also recorded in the node label list. Note that the above strategy may introduce more than one representation for each candidate node label. This is likely to happen when the label and object shapes are symmetric. In such cases, we need to compare the support received for this assignment from the neighbouring objects and select the most consistent representation. At the end of this process we will have a label list for each object with the best representation for each label in the list. Hence we do not need to distinguish between different representations by superscript indices on the unary and binary vectors. Instead we index them with a star to indicate that the best representation is being considered.

In the second stage, we consider the possible label assignments for each object, a_i, and iteratively update the probabilities using their previous values and supports provided by the neighbouring objects. In addition to the above pruning measures, at the end of each iteration we eliminate the labels the related probabilities of which drop below a threshold value. This will make our relaxation method faster and more robust as well. Indeed the updating of probabilities of unlikely matches not only takes time but also increases the probability of incorrect assignments due to increased entropy of the interpretation process which is a function of not only probability distribution but also of the number of possible interpretations. Returning to our problem of assigning labels from label set Ω to the set of objects in the scene graph, let $p(\theta_i = \omega_{\theta_i})$ denote the probability of label ω_{θ_i} being the correct interpretation of object a_i. In the iterative process we consider all possible assignments for object a_i and the related probabilities using their previous values and supports provided by the neighbouring objects. As mentioned before we combine the iteration rules in [13] and [8] to derive a new iteration formula defined as:

$$p^{(n+1)}(\theta_i = \omega_{\theta_i}) = \frac{p^{(n)}(\theta_i = \omega_{\theta_i})Q^{(n)}(\theta_i = \omega_{\theta_i})}{\sum_{\omega_\lambda \in \mathbf{L}_i} p^{(n)}(\theta_i = \omega_\lambda)Q^{(n)}(\theta_i = \omega_\lambda)} \tag{7}$$

$$Q^{(n)}(\theta_i = \omega_\alpha) = p^{(n)}(\theta_i = \omega_\alpha) \sum_{j \in N_i} \sum_{\omega_\beta \in \mathbf{L}_j} p^{(n)}(\theta_i = \omega_\beta) p(\mathbf{A}^*_{ij} | \theta_i = \omega_\alpha, \theta_j = \omega_\beta) \tag{8}$$

where function Q quantifies the support that assignment $(\theta_i = \omega_\alpha)$ receives at the nth iteration step from the neighbours of object i in the scene.

In the first step of the process the probabilities are initialised based on the unary measurements. Denote by $p^{(0)}(\theta_i = \omega_{\theta_i})$ the initial label probabilities evaluated using the unary attributes as:

$$p^{(0)}(\theta_i = \omega_{\theta_i}) = p(\theta_i = \omega_{\theta_i} | \mathbf{x}^*_i) \tag{9}$$

Applying the Bayes theorem we have :

$$p^{(0)}(\theta_i = \omega_{\theta_i}) = \frac{p(\mathbf{x}^*_i | \theta_i = \omega_{\theta_i}) p(\theta_i = \omega_{\theta_i})}{\sum_{\omega_\alpha \in \mathbf{L}_i} p(\mathbf{x}^*_i | \theta_i = \omega_\alpha) p(\theta_i = \omega_\alpha)} \tag{10}$$

with the normalisation carried out over labels in the label list $\mathbf{L}_i$. Let ζ be the proportion of scene nodes that will assume the null label. Then the prior label probabilities will be given as :

$$p(\theta_i = \omega_\lambda) = \begin{cases} \zeta & \lambda = 0 \ (null\ label) \\ \frac{1-\zeta}{M} & \lambda \neq 0 \end{cases} \tag{11}$$

where M is the number of labels (model nodes).

We shall assume that the errors on unary measurements are statistically independent and their distribution function is Gaussian i.e.

$$p(\mathbf{x}^*_i | \theta_i = \omega_\alpha) = \mathcal{N}_{\mathbf{x}^*_i}(\breve{\mathbf{x}}_\alpha, \Sigma_u) \tag{12}$$

where Σ_u is a diagonal covariance matrix for measurement vector $\mathbf{x}^*_i$. In the support function, Q, the term $p(\mathbf{A}^*_{ij} | \theta_i = \omega_\alpha, \theta_j = \omega_\beta)$ behaves as a compatibility coefficient in other relaxation methods. In fact it is the density function for the binary measurement $\mathbf{A}^*_{ij}$ given the matches $\theta_i = \omega_\alpha$ and $\theta_j = \omega_\beta$. Similarly, the distribution function of binary relations is centred on the model binary measurement $\breve{\mathbf{A}}_{\alpha\beta}$. It is assume that deviations from this mean are modelled by a Gaussian. Thus we have:

$$p(\mathbf{A}^*_{ij} | \theta_i = \omega_\alpha, \theta_j = \omega_\beta) = \mathcal{N}_{\mathbf{A}^*_{ij}}(\breve{\mathbf{A}}_{\alpha\beta}, \Sigma_b) \tag{13}$$

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

The iterative process will be terminated in one of the following circumstances:

1. If in the last iteration none of the probabilities changed by more than threshold ϵ.
2. If the number of iterations reached some specified limit

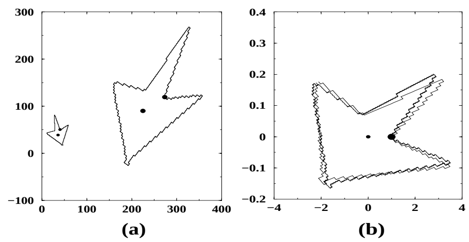

Figure 1: Demonstration of region normalisation

4 Experimental Results

In this section we present the results of experiments designed to demonstrate the advantages of the proposed method of recognition of objects in cluttered test images. Initially to show how regions in our method are normalised, let us consider the two boundaries in figure 1-a which are related by affine transformation. The marked points represent the extracted reference points on each region. For each region a point of high curvature and the centroid of the region constitute the region reference points. After normalisation we expect the corresponding reference points to coincide in the predetermined coordinates $(0, 0)$ and $(0, 1)$ and the boundaries should be closely aligned with each other. The effect of normalisation step is shown in figure 1-b for the two boundaries in figure 1-a. The sligth error in the boundaries alignment is due to the error in the extraction of the region reference points.

Next we show how the proposed method of node reference basis selection can be used to prune the list of candidate labels assigned to the object. To illustrate its effectiveness we show in figure 2 the boundary of a normalised scene region and compare it with the boundary of three model regions. Each row of the figure presents the comparison for the two different representations of the scene region. In order to get a good agreement (low value E) both the interpretation and the representation have to be correct. For a good fit the value of E is at least two orders of magnitude smaller than any other combination of label and representation. We now proceed to demonstrate the recognition ability of the proposed method in a realistic scenario. As shown in figure 3, the objects to be recognised are traffic signs. Close-up views of the objects are used as the model images.

The recognition method is applied to a number of outdoor test images taken in different illumination conditions. Examples are shown in figure 4. As can be seen, the imaging viewpoints are such that the objects of interest in the test images are significantly smaller than the model images. We used the colour segmentation method proposed in [3] to extract image regions. Then based on the extracted regions we represented both the scene and model images in terms of the scene and model graphs. In contrast to the previous work[2] where we chose three representations for each object node, in this application we chose nine representations to provide greater robustness to noise and change of view point. This increased multiplicity is motivated by two considerations. First as our objects have only a few regions each, we can not risk miss-assigning even one node of the model. This is particularly important also from the point of view of object disambiguation as in some of the model images one specific region will allow us to distinguish between two models. Second, as in the proposed matching algorithm only one of the object representations is

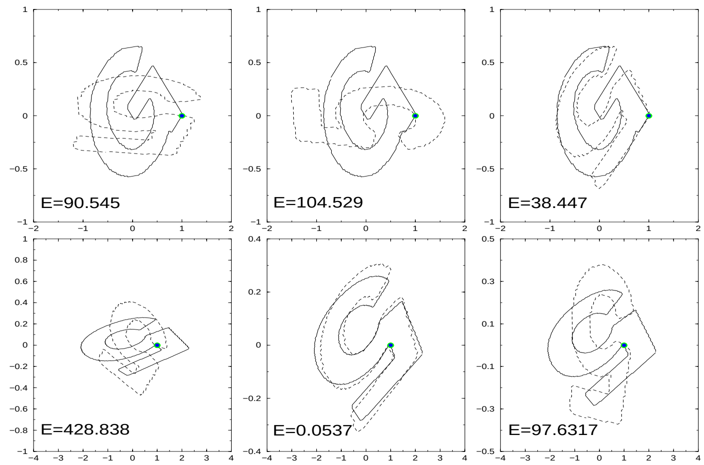

Figure 2: Sensitivity of the proposed criterion function, E, for assessing candidate representations

involved in the labelling process, increasing the number of candidate representations is more tolerable than in the previous work.

The impact of the proposed evidence combination algorithm was significant both computationally and performance wise. In the first step, using the proposed criterion, E, at least 30% of incorrect hypotheses have been eliminated from the label list of each object. The iterative labelling process, as a result of continuously pruning unlikely hypotheses, converges very fast so that at the end of third iteration the most of the irrelevant regions in the scene will have been assigned the null label. After five iterations the difference between the corresponding label probabilities in two successive steps is negligible and algorithm terminates. It should also be mentioned that the matching algorithm is not very sensitive to parameters Σ_u and Σ_b which have been determined experimentally.

In spite of the large number of extracted regions in each scene image, the method is able to recognise 90% of the model regions correctly and among all the irrelevant scene regions only 1% do not take the null label. This compares very favourably with the previous method where these figures were 80% and 7% respectively.

5 Conclusions

We addressed the problem of model-based object recognition in computer vision. In our approach the model and scene were represented in the form of Attributed Relational Graphs using multiple representations at each node of the scene ARG. We argued that this enhanced the stability of the scene image representation in the presence of noise and significant scaling. We proposed a novel method to identify the most probable representation among the available candidates at each scene node. By reducing the number of representations involved, the relaxation matching process was less likely to get stuck in local optima. Further speeding up of the labelling process was accomplished by pruning the candidate model labels associated with the scene nodes at each iteration. The experimental results obtained on real data demonstrated that the proposed method can deliver

Figure 3: The traffic signs as the library of model images

Figure 4: The test images

good recognition performance even under severe viewing conditions. It also convincingly outperforms our earlier method.

References

[1] Rosenfeld A., Hummel R., and Zucker S. Scene labeling by relaxation operations. *IEEE Transaction on system man and cybernetics*, pages 420–433, 1976.

[2] Ahamdyfard A.R and Kittler J. Region-based representation for object recognition by relaxation labelling. In *WORKSHOP ON Syntactical and Structural Pattern Recognition*, Alicante, Spain, September 2000.

[3] Comaniciu D. and Meer P. Robust analysis of feature spaces: Color image segmentation. *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 750–755, 1997.

[4] Huttenlocher D.P. and S. Ullman. object recognition using alignment. *first international confrance of computer vision*, pages 102–111, 1987.

[5] Huttenlocherr D.P and Ullman S. Recognizing solid objects by alignment with an image. *International Journal of Computer Vision*, pages 195–212, 1990.

[6] Wolfson H.J. Model-based object recognition by geometric hashing. *International confrence on computer vision*, pages 526–536, 1990.

[7] Wolfson H.J. and Lamdan Y. Geometric hashing: A general and efficient model-based recognition scheme. *Second International Conference on Computer Vision*, pages 238–249, 1988.

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

[9] Wong K.C., Sim H.C, and Kittler J. Recognition of two dimensional objects based on a novel generalized hough transform method. *IEEE International Conference on Image Processing*, 3:376–379, 1995.

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

[11] Z. Shao and J. Kittler. Shape recognition using invariant unary and binary relations. *Advances in Visual Form Analysis*, pages 529–538, 1997.

[12] Z. Shao and J. Kittler. Shape representation and recognition using invariant unary and binary relations. *Image and Vision Computing*, 17:429–444, 1999.

[13] 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.