

# Probabilistic Relaxation Labeling: A Short Survey on Object Recognition

Abbas Zohrevand  
Department of Computer Engineering  
Kosar University of Bojnord  
Bojnord, Iran

## ABSTRACT

Object recognition problem can be defined as classifying input object(s) to number of predefined classes. Object recognition is one of the most important sections in computer vision. While this field has been studied from long time ago, but it still suffers from several challenges such as: occlusion, rotation, distortion illumination, and scaling. The conventional object recognition system has two phases. Firstly: extraction of the most important (informatics or key points) parts from object image (scene image) and predefined class image (model image), secondly matching between object and model. The Probabilistic Relaxation Labeling (PRL) is one of the popular probabilistic approaches in matching among model and scene. In this paper we review two phase and report the most important works based PRL.

## Keywords

Object recognition, Probabilistic Relaxation Labeling, image descriptor, model image, scene image.

## 1. INTRODUCTION

Object recognition is intrinsically a difficult problem, both conceptually and technically. It is due to a host of factors, which are intrinsic to the process of object sensing using imaging techniques. These factors include geometric transformation of the measurements as a result of changing view point[1]. Object recognition systems involve two main problems, object representation and object matching[2]. The representation of 3D objects using their 2D views has received significant attention from many researchers[2], [3]. These methods can be broadly classified into two classes: feature-based and appearance-based. The feature-based methods represent an image of each object using local features. These features are often chosen among relatively simple image primitives (e.g. interest points, lines) as such features can be reliably detected in clutter images. The objective in these methods is to find correspondence between two images of an object based on the geometric configuration of the extracted features. Using geometric restriction enables these methods to recognize an object using single image or its few dominant views. Moreover, the reliance on local features

makes them robust to occlusion and object deformation. As their success depends on how reliably and accurately the image features are extracted, many attempts have been made to extract image descriptors being robust in a wide range of imaging conditions[4], [5].

In contrast to the feature-based recognition paradigm, the appearance-based approach models an object using its appearance as manifest in the associated image intensity function. In the seminal paper of [3], an object is learnt using many of its views. Exploiting the correlation between the respective views, the set of object images are then represented into a low-dimensional eigenspace. Despite the remarkable success achieved in modeling real objects, the sensitivity to outliers (occlusion and changing background) was considered as one of the major drawbacks of this method. However, a significant progress has been made[6]–[8] towards solving the occlusion problem. For instance, instead of using all image data to determine the projections into eigenspace, Leonardis et al. [6] proposed to use subsets only. A different approach is advocated by Huang et al.[7], who first divide the image into its constituent parts, and each part then individually associated with its adjacent parts are represented in eigenspace. The fundamental problem of the appearance-based approach namely the need for a large number of object views remains a serious drawback. In the next section the most conventional object recognition system which uses both of these approaches (feature and appearance-based) has been reviewed.

## 2. OBJECT RECOGNITION SYSTEM

The whole diagram for object recognition system is shown in Figure 1. As seen in this figure, any object recognition system has three sections. Firstly: description which model and object (scene) images described by image descriptor, respectively. Secondly, an Attributed Relational Graph (ARG) constructed for model and scene image and thirdly two constructed graphs matched by Probabilistic Relaxation Labeling framework. In the next subsection each section reviewed

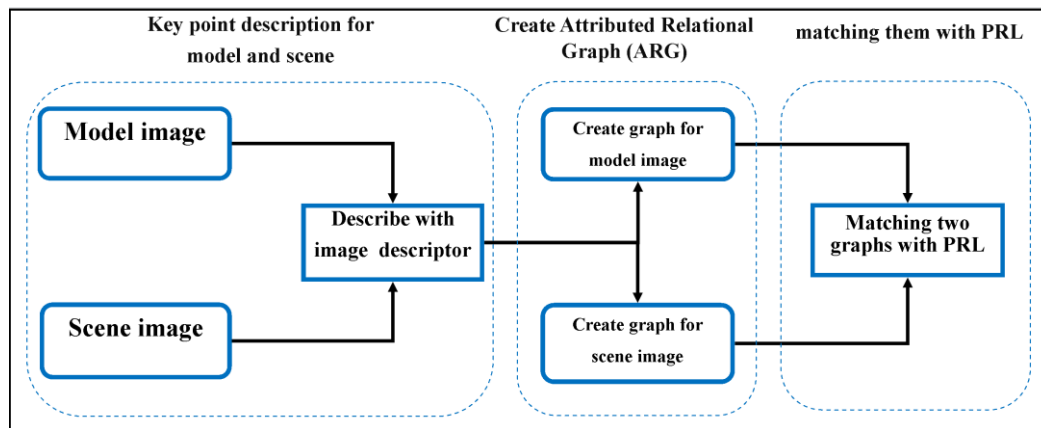


Figure 1. The whole diagram for object recognition system which contain three stages such as: key points extraction, create ARG graph from model and scene and matching corresponding created graphs.

## 2.1 Image descriptor

In order to utilize both the advantages of feature-and appearance-based approaches, methods based on the matching of local descriptors have been emerged. In contrast to local features, a local descriptor characterizes a part of an image more uniquely. The advantage of such representation is that it allows a direct matching of model and scene descriptors to identify an object. Furthermore, the use of local descriptors that are invariant to geometric transformation enables these methods to reduce the number of object views required for object modeling. Among proponents of this approach, Schmid et al[9], represented an object image using a set of differential features (called local jets) around each detected corner of the image. Lowe detects local extremums of image intensity at multi-scales, referred to as stable points[10]. Shokoufandeh et al. represent an object view in terms of the salient regions extracted at multiple scales using a wavelet transform[11]. Lowe presents a scale invariant feature transform (SIFT) to detect image salient points and describe each extracted point using local features in images[10]. SIFT key-points of objects are first extracted from a set of reference images and stored in a database. An object is recognized in a new image by individually comparing each feature from the new image to this database and finding candidate matching features based on Euclidean distance of their feature vectors. SIFT can robustly identify objects even among clutter and under partial occlusion, because his SIFT feature descriptor is invariant to uniform scaling, orientation and partially invariant to affine distortion and illumination changes. Other researchers presented other ways based on SIFT to improve it, for example Morel[12] proposed Affine SIFT (ASIFT). In compare to SIFT, ASIFT was more invariant in affine transformation. Original SIFT descriptor use gray level image, so Ancuti[13] proposed SIFT-CCH, which can be applied on color images. Bay et al. [14] present a method so called Speeded Up Robust Features (SURF) to extract a set of robust local features from image. It is partly inspired by the SIFT descriptor, the standard version of SURF is several times faster than SIFT and claimed to be more robust against different image transformations than SIFT.

## 2.2 Matching with PRL

In the computer vision literature, different approaches for image feature matching have been reported. Basically, the matching of object features in the scene image to those in the model of object is highly dependent on how object is represented in two images (model and scene). The matching

in many feature-based methods is addressed as the problem of determining the geometric transformation which maps the object image from the scene to its model. The global consistency is checked by measuring alignment between model and the transformed scene features. Techniques range from the alignment method[15], which is relatively time consuming to geometric hashing [4], [16] or Hough transform methods[17] the success of which is largely depends on the ability to extract invariant features[15]. A common criticism for these matching approaches is that a small error in the estimation of transformation parameters may result in a considerable miss-match between corresponding features[15].

In contrast, matching in appearance-based methods is carried out in the feature space. For this purpose, efficient algorithms for matching high-dimensional features have been proposed[3]. However, in the case of methods employing distinctive local descriptors, each of the scene descriptors can be individually matched against the descriptors of all object models. The model with the highest matching score determines the identity of the object in the scene[9]. Despite the simplicity of this approach, which offers a fast matching process, the lack of distinctive descriptors and presence of similar objects in the database degrade its performance. Similarity between descriptors in the scene and model is utilized to hypothesize the transformation between the model and the scene[16]. The hypothesis is then verified by checking global consistency of the descriptors. However, the problem of feature misalignment which is a shortcoming of feature-based matching still remains. In[11], authors represent salient regions in a hierarchical graph in which node represents a region. The link between a pair of nodes indicates that regions associated with nodes are extracted at different scales and the center of one region lies inside the other region. They suggested to use topological and geometrical constraints for matching of two graphs[11]. However, even topological relations may be affected by imperfect segmentation and also they are not very discriminative.

In an earlier work it was argued that an effective object recognition method should be based on extraction of relatively simple features as only such features can be reliably detected in complex images[18]. The distinctiveness of such features can be enhanced by relational measurements. However, these relations should be of low order to minimize computational complexities of both feature extraction and model matching, and to maximize the probability of the features being observable. As a suitable tool to achieve the latter objective,

the evidence combination method of relaxation labeling was adopted[19]. The method uses an Attributed Relational Graph (ARG) for representing both scene and model. Only unary and binary relations who are made invariant to any pertinent geometric transformation group are employed. For example an object image (model and scene) is represented using its key-point descriptor, which is extracted based on SIFT descriptor. To matching, an Attributed Relational Graph (ARG) is created, which in ARG each node represents a key-point and an edge in this graph connects a pair of key-points in a predefined neighborhood (Figure 2). The graph nodes and edges are characterized using unary and binary measurements, respectively.

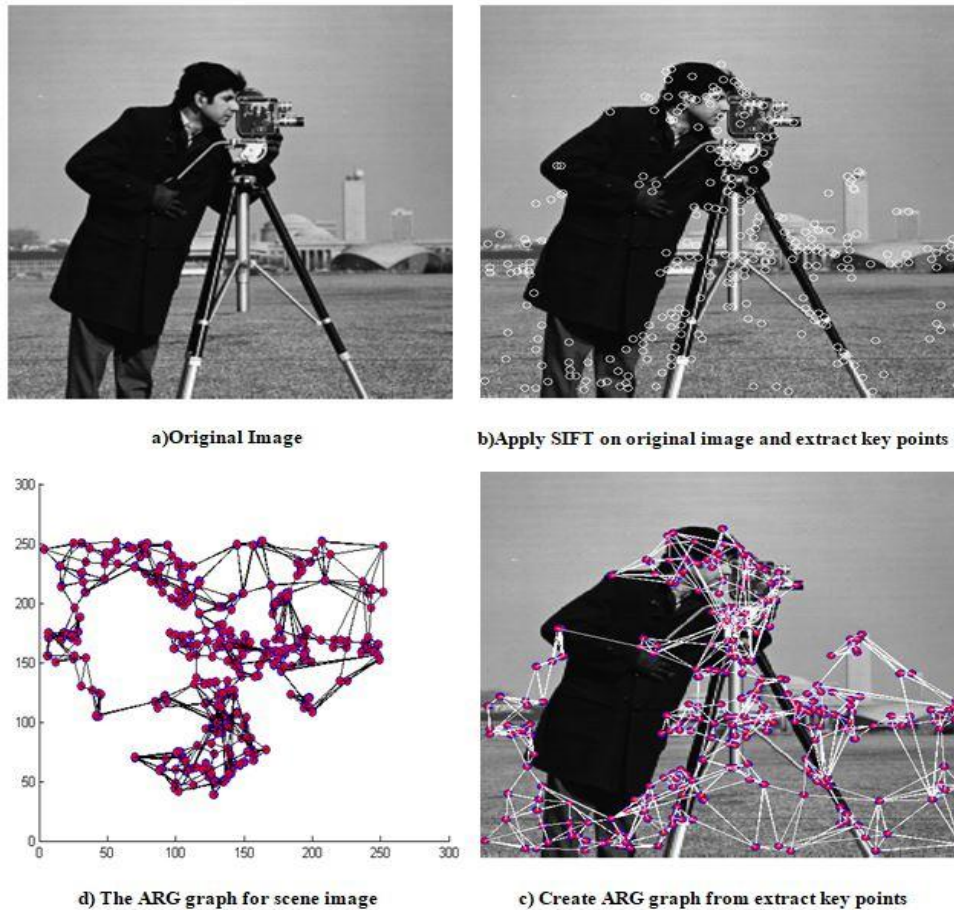
Scene labeling using relaxation techniques, starts a new approach in the contextual interpretation of spatial data using soft computing processes. An early Waltz attempt to discrete labeling[20]. Rosenfeld et al. developed a model to relax discrete labels by probabilistic assignments[21]. Ahmadyfard et al. proposed a novel approach for matching model and scene[22]. They represent each model and the scene in the form of attributed relational graph (ARG). 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 minimize the label clutter. The experimental results obtained on real data demonstrate promising performance of the proposed recognition system. Their model called asymmetric because of it use NULL node in matching. If any node in scene does not exist in model graph, this node labeled NULL by asymmetric probabilistic relaxation labeling. Kostin et al[23] proposed object recognition system using graph-matching which include two-stage process: extracting suitable object descriptor from an scene and corresponding models, and matching graphs constructed from these two sets of object descriptors. They focused mainly on the second issue of graph matching, for which they use a technique based on probabilistic relaxation graph labeling. The presented approach was evaluated on two standard data sets, SOIL-47 and COIL-100, in both of which objects must be recognized from a variety of different views. In fact they adopted PRL proposed by Ahmadyfard[1] method based on symmetric probabilistic relaxation graph labeling. Kostin in his approach extracted Image primitive by SIFT descriptors. They don't use NULL node (symmetric) in matching, so result may end up

with a set of unmatched nodes in scene graph. Hummel et al. placed the probabilistic relaxation into continuous optimization by demonstrating that ending consistent labeling was equivalent at maximizing a local average consistency[24]. Hence, the problem could be solved by traditional continuous optimization techniques such as gradient descent. They showed that in their method objects can be recognized from different views. Zohrevand et al[25] use SIFT descriptor for extract object primitive from model and scene and presented model and scene as ARG(for creating ARG graph, see Figure 2). Then, they use asymmetric model with NULL node for matching.

### **3. OTHER APPLICATION OF PRL**

Beside object recognition, PRL can be used in other application such as: super resolution, object tracking, object retrieval and etc. Chevalier et. al[26] used PRL for object retrieval in video. They adopted the problem of matching of objects in video in the context of the rough indexing paradigm. Their proposed approach was based on matching of region adjacency graphs (RAG) of pre-segmented objects. The video data had low resolution in the context of the rough indexing paradigm; hence, segmentation is inaccurate, and the RAGs vary with the time. The contributions of their approach were a graph matching method for such RAGs based on relaxation labeling techniques. In their approach, modification of similarity among regions according to neighborhood consistency compensate for the inaccuracy of segmentation. In compared to another region-based technique, their approach shows superior performance on real sequences.

Image Registration is a critical preprocessing step in applications like medical images, super resolution and remote sensing. Dense image registration requires several deformations and transformations such as contrast changing, scaling, rotation and displacement. However, in most presented approaches, only some of these transforms are considered which results in incorrect output. Amiri et.al[27] propose a new method for image registration based on relaxation labeling. In their proposed approach, for each pixel of a test image sample, the best match (nearest pixel) in a reference image is found. They use blocks of reference image as features, then they look for the closest candidate in the test image. In next step, a relaxation labeling framework was applied to these candidates for selecting the best match between candidate pixels. Experimental results show that their proposed method achieves satisfactory performance in terms of visual quality, PSNR values and Bad pixel evaluation criteria.



**Figure 2. Creating ARG for sample image.**

Remote sensing image allows the monitoring of the surface and atmosphere of the earth on different scales. The development of imagery sensors has increased the availability of high-quality remote sensing images. Yousefi et.al[28] proposes a method to combine contextual, structural, and spectral information for classification. Their method is an integrated method for automatically classifying urban-area objects in very high-resolution satellite imagery. The approach addresses three aspects. First, the Gabor wavelet is applied to the image along with morphological operations, with the sparsity of the outcome considered. A Bayesian classifier then categorizes the different classes, such as buildings, roads, open areas, and shadows. There are some false positives (wrong classification), and false negatives (non-classification) in the initial results. These results can be corrected by the relaxation labeling categorization of the unknown regions.

#### **4. CONCLUSION**

This paper presents some application of Probabilistic Relaxation Labeling method on object recognition and similar task on computer vision. PRL is iterative approach which tries to find the best match among model and object(s) like model in clutter scene. There are a lot of challenges in object recognition such as: occlusion, rotation, distortion illumination, and scaling. While, there are many works which tries to solve object recognition, but due to several factors which mention above object recognition still is open problem.

#### **5. REFERENCES**

- [1] A. R. Ahmadyfard and J. Kittler, "Using relaxation technique for region-based object recognition," *Image Vis. Comput.*, vol. 20, no. 11, pp. 769–781, 2002.
- [2] R. T. Chin and C. R. Dyer, "Model-based recognition in robot vision," *ACM Comput. Surv.*, vol. 18, no. 1, pp. 67–108, 1986.
- [3] H. Murase and S. K. Nayar, "Visual learning and recognition of 3-D objects from appearance," *Int. J. Comput. Vis.*, vol. 14, no. 1, pp. 5–24, 1995.
- [4] H. J. Wolfson, "Model-based object recognition by geometric hashing," in *Computer Vision—ECCV 90*, Springer, 1990, pp. 526–536.
- [5] S. Procter and J. Illingworth, "Foresight: Fast object recognition using geometric hashing with edge-triple features," in *Image Processing, 1997. Proceedings., International Conference on, 1997*, vol. 1, pp. 889–892.
- [6] A. Leonardis and H. Bischof, "Robust recognition using eigenimages," *Comput. Vis. Image Underst.*, vol. 78, no. 1, pp. 99–118, 2000.
- [7] C.-Y. Huang, O. I. Camps, and T. Kanungo, "Object recognition using appearance-based parts and relations," in *Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference on, 1997*, pp. 877–883.
- [8] R. C. Nelson and A. Selinger, "A cubist approach to object recognition," in *Computer Vision, 1998. Sixth*

- International Conference on, 1998, pp. 614–621.
- [9] [C. Schmid and R. Mohr, “Local grayvalue invariants for image retrieval,” *Pattern Anal. Mach. Intell. IEEE Trans.*, vol. 19, no. 5, pp. 530–535, 1997.
- [10] D. G. Lowe, “Object recognition from local scale-invariant features,” *Computer Vision*, 1999. The Proceedings of the Seventh IEEE International Conference on, vol. 2. pp. 1150–1157 vol.2, 1999.
- [11] A. Shokoufandeh, I. Marsic, and S. J. Dickinson, “View-based object recognition using saliency maps,” *Image Vis. Comput.*, vol. 17, no. 5, pp. 445–460, 1999.
- [12] J.-M. Morel and G. Yu, “ASIFT: A new framework for fully affine invariant image comparison,” *SIAM J. Imaging Sci.*, vol. 2, no. 2, pp. 438–469, 2009.
- [13] C. Ancuti and P. Bekaert, “SIFT-CCH: Increasing the SIFT distinctness by Color Co-occurrence Histograms,” in *Image and Signal Processing and Analysis, 2007. ISPA 2007. 5th International Symposium on*, 2007, pp. 130–135.
- [14] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, “Speeded-up robust features (SURF),” *Comput. Vis. image Underst.*, vol. 110, no. 3, pp. 346–359, 2008.
- [15] W. E. L. Grimson and D. P. Huttenlocher, “On the sensitivity of the Hough transform for object recognition,” *Pattern Anal. Mach. Intell. IEEE Trans.*, vol. 12, no. 3, pp. 255–274, 1990.
- [16] D. G. Lowe, “Distinctive image features from scale-invariant keypoints,” *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.
- [17] A. Rosenfeld and S. W. Zucker, “Scene Labeling,” no. 6, pp. 420–433, 1976.
- [18] A. Ahmadyfard and J. Kittler, “Region-based representation for object recognition by relaxation labelling,” in *Advances in Pattern Recognition*, Springer, 2000, pp. 297–307.
- [19] K. Mikolajczyk and C. Schmid, “Indexing based on scale invariant interest points,” in *Computer Vision*, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on, 2001, vol. 1, pp. 525–531.
- [20] D. L. Waltz, “Generating semantic description from drawings of scenes with shadows,” 1972.
- [21] A. Rosenfeld, R. A. Hummel, and S. W. Zucker, “Scene Labeling by Relaxation Operations,” *Systems, Man and Cybernetics, IEEE Transactions on*, vol. SMC-6, no. 6, pp. 420–433, 1976.
- [22] A. Ahmadyfard and J. Kittler, “Region-Based Object Recognition: Pruning Multiple Representations and Hypotheses,” in *BMVC, 2000*, pp. 1–10.
- [23] A. Kostin, J. Kittler, and W. Christmas, “Object recognition by symmetrised graph matching using relaxation labelling with an inhibitory mechanism,” *Pattern Recognit. Lett.*, vol. 26, no. 3, pp. 381–393, 2005.
- [24] R. A. Hummel and S. W. Zucker, “On the Foundations of Relaxation Labeling Processes,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. PAMI-5, no. 3, pp. 267–287, 1983.
- [25] A. Zohrevand, A. Ahmadyfard, A. Pouyan, and Z. Imani, “A SIFT based object recognition using contextual information,” in *Intelligent Systems (ICIS), 2014 Iranian Conference on*, 2014, pp. 1–4.
- [26] F. Chevalier, J.-P. Domenger, J. Benois-Pineau, and M. Delest, “Retrieval of objects in video by similarity based on graph matching,” *Pattern Recognit. Lett.*, vol. 28, no. 8, pp. 939–949, 2007.
- [27] M. Amiri, A. Ahmadifard, and V. Abolghasemi, “A probabilistic framework for dense image registration using relaxation labelling,” in *Signal Processing and Intelligent Systems (ICSPIS), International Conference of*, 2016, pp. 1–5.
- [28] B. Yousefi, S. M. Mirhassani, A. AhmadiFard, and M. Hosseini, “Hierarchical segmentation of urban satellite imagery,” *Int. J. Appl. Earth Obs. Geoinf.*, vol. 30, pp. 158–166, 2014.