Object Recognition Using Wavelet Based Salient Points

S. Arivazhagan and R. Newlin Shebiah*

Department of Electronics and Communication Engineering, Mepco Schlenk Engineering College, Sivakasi, 626 005, India

Abstract: In this paper an efficient method to learn and recognize objects from unlabeled natural scenes using patch based object representation is proposed. In the domain of object recognition, it is often the case that images have to be classified based on objects which make up only a very limited part of the image. Hence Patches (local features) are used to describe properties of certain region of an image. The proposed algorithm directly matches the parts distributed in a reference image that contains the object to those extracted in the test and hence it reports better matching. The experimental evaluation of the proposed method is done using the well-known Caltech database.

Keywords: Object Recognition, Wavelet Transform, Salient points, zero tree representation.

1. INTRODUCTION

The human visual system allows us to identify and distinguish thousands or possibly millions of objects quickly and effortlessly, even in complex cluttered surroundings and in the presence of partial occlusions. Duplicating this ability with computers has many practical applications.

Computer programs that tackle it must cope with a lot of difficulties: They must be able to recognize arbitrary objects, the object to be recognized may be occluded by other objects and therefore be only partially visible, the object can appear at any position and with any size in the image and the appearance of the object is not restricted to any "prototypical" appearance [1]. The proposed method allows for recognizing objects under these challenging circumstances and provides excellent results on Caltech database.

Generic object recognition systems do not include any information about specific objects. Rather, they learn to recognize arbitrary objects by inspecting a set of training images and train a model on these. This model is then used to recognize the objects in unseen images. For each of the training images a set of features are derived. Each feature describes properties of either the whole image (global feature) or a part of the image (local feature). Usually, local features are more successful in capturing the content of complex images. To reliably recognize objects under varying circumstances (for example, objects appearing at different scales, rotation, and translation) the features ought to be chosen such that they are invariant with respect to these aspects. From the features of the training images, the parameters of an underlying statistical model are estimated. Using these features and the trained model the object recognition system outputs which of the trained objects is contained in the image, or, in the detection case, if the object in question is contained in the image or not. Once trained, the performance of object recognition systems is measured on a set of test images. The recognition rate on this set denotes the ratio of correctly classified images to all images in the test data set.

In complex images, the information provided by the global features is not sufficient and therefore they are not well suited in this context. Hence local features like patches are better suited for complex images, because they represent restricted regions of the image. Beneficial properties of local features are: Inherent translation invariance, Robustness to object variance and occlusion and possible scale invariance [1].

The work on object recognition using parts started in late 90's. Burl *et al.* Proposed Part detection *via* matched filtering. Here the inputs are the manual selection of candidate parts, eyes, nose tip, mouth corners etc, and then a probabilistic shape model is applied [2]. Weber *et al.* proposed a method where, Forstner interest points get extracted, and then features are calculated and clustered using k-means clustering algorithm [3]. For classification, the joint probability density of the detected part locations is evaluated. Agarwal and Roth (2002) and Agarwal *et al.* (2004) used a technique where, features are clustered in an image and the occurrence of cluster members at a specific spatial relationship is coded in a binary vector [4, 5]. As a classifier, winnow are used. For localization, a sliding window approach is used to calculate a classifier activation map, i.e., the probabilities, that an object at a certain location is present. Extraction of scale and affine covariant parts, calculation of Scale-invariant feature transform (SIFT) features and clustering of the features with Gaussian Mixture Model (GMM), were used in Dorko and Schmid (2003) where each Gaussian represents a cluster. Part classifiers are built using Neural Network with Gaussian kernel density and discriminative parts are found with two criteria: classification likelihood or mutual information [6]. For classification, the *n* most discriminative object parts are used and the final decision is done whether the number of "activated" positive part classifiers is above a certain threshold, which is determined for each class. About 30 Kadir & Brady regions get extracted and normalized to 11x11 pixels in the work done by Fergus *et al.* [7]. The approach is similar to Weber *et al.* (2000),

*Address correspondence to this author at the Department of Electronics and Communication Engineering, Mepco Schlenk Engineering College, Sivakasi, 626 005, India; E-mails: newlinshebiah@yahoo.co.in; r.newlinshebiah@gmail.com

1876-8253/09

nition Results. Finally, Section 4 gives the Discussion and Conclusion of the proposed method.

2. METHODOLOGY

The two sections that involved in this work are Feature Extraction and Feature Matching. The block diagram of the proposed method is given in Fig. (**1**).

Among the various possible methods of representing the image features, salient point detection by combining wavelet transform and zero tree representation of the wavelet coefficients is preferred. Patches are extracted over the salient points to represent the local features. The extracted patches are manipulated and turned into feature vectors where PCA dimensionality reduction is applied to extract the appropriate features from the patches. Brightness normalization is done to achieve homogeneous brightness of the image. The patches from all training images are then jointly clustered with a K-means algorithm such that 1024 clusters are obtained. Similarly for the test image, the patches are extracted and they are labeled with the cluster centroids of training images.

By using the proposed algorithm, optimal matching is obtained between the test image and the given training images that contain the object of interest.

2.1. Feature Extraction

Salient Point Detection

These salient points are literally the points on the object which are almost unique. These points maximize the discrimination between the objects. The characteristics of salient points as proposed by Haralick and Shapiro [15] are are: Distinctness, Invariance, Stability, Uniqueness and Interpretability.

Numerous algorithms for interest point detection have been proposed. The earliest method that is still widely used today is the Harris corner detector [16]. Harris corners are found using the eigenvalues of the second-moment matrix. They are rotationally invariant, but not scale-invariant. Then to extract the corner points, Chen *et al.* used two different resolutions [17]. Loupias used Wavelet Transform to extract both the global as well as the local variations [18].

The wavelet transform is used to extract salient points [18,19]. Orthogonal wavelets are used in our implementations, since it leads to complete and non-redundant representation of the image. The extension of the wavelet model to two dimensions leads to three different wavelet functions related to three different spatial orientations (horizontal, diagonal and vertical). In this framework, the extension of our salient point extraction is straightforward. The algorithm is as follows:

1. Wavelet decomposition of the image at level j using compact support wavelet such as**,** Haar.
2. Process the approximation subband with a threshold to extract highest salient points. The threshold T is given by

$T=2^{\lfloor \log_2(\text{maximum pixel value}) \rfloor}$.

Using wavelet transform, salient points are detected for smoothed edges also. The salient points are not gathered in textured regions. This method leads to a more complete image representation than corner detectors. Salient point extraction from an airplane image using wavelet transform with threshold processing is shown in Fig. (**2**).

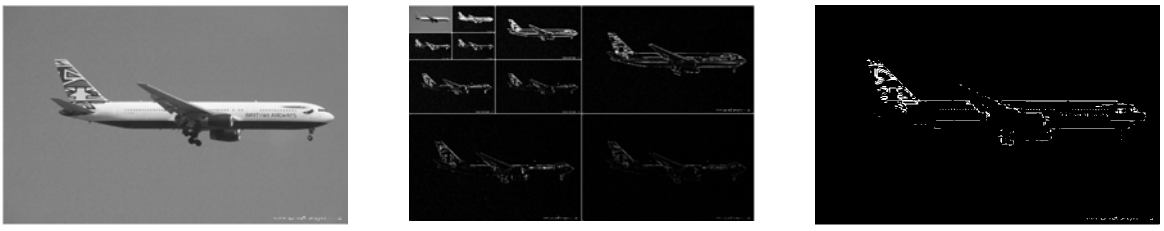
Patch Extraction

Patches are the squared sub images extracted from the image over the salient points. For the images of the Caltech database patch size of 11×11 pixels perform well [1]. If the objects appear across all images at roughly the same size, then all patches can be extracted at the same chosen patch size. Anyway, this assumption is unlikely to hold in many cases. A possibility to address this scale difference of the objects is to extract the patches at different scales. Surely, extracting patches at multiple sizes increases the amount of data to deal with. Hence in this experiment, patches are extracted at only 11×11 pixels, but overlapping patches are allowed. Fig. (**3**) marks the patches of size 11×11 extracted from the image over the salient points.

PCA Dimensionality Reduction [20]

In the simplest case, the pixel values of the patches can be used without any further processing as components of the feature vectors. For an $n \times n$ patch, there are n^2 gray-level pixel values and thus we obtain feature vectors with n^2 components. Therefore, a feature or dimensionality reduction of the feature vectors is desirable. A commonly used reduction method is the Principal Component Analysis (PCA). The steps involved in PCA dimensionality reduction are as follows:

1. Mean vector $\boldsymbol{\mu}$ and the Covariance matrix $\boldsymbol{\xi}$ are computed for all patches of the training images.
2. Find Eigen values and Eigenvectors and sort according to decreasing absolute Eigen values.



(a) Input Airplane Image *(b) Three Level Wavelet Decomposed Image* *(c) Salient Points*

Fig. (2). Salient points extraction for air plane image.

(a) Input Image

(b) Patches (11×11)

Fig. (3). Patches over salient points.

3. Select the top *m* eigenvectors as principle components.

The first 40 coefficients are sufficient to maintain most of the information a patch contains.

Brightness Normalization

As long as non-artificial images or images which have already been normalized beforehand are used, it is quite normal that the objects appear in different images under different lighting conditions. It seems natural to apply a brightness normalization to achieve a more homogeneous brightness of the objects, which can either be done on the whole image or on the extracted patches. Normalizing patches is done by discarding the first PCA Component of the patches after the PCA transformation. The first PCA component approximates the effect of brightness normalization as the "energy" of a patch, i.e., its overall brightness is reflected mainly in the first PCA coefficient.

K-Means Clustering [21]

K-means is one of the simplest unsupervised learning algorithms that solve the well known clustering problem. The procedure follows a simple and easy way to classify a given data set through a certain number of clusters (assume k clusters) fixed a priori. The main idea is to define *k* centroids, one for each cluster. Algorithmic steps for K-Means clustering are as follows:

1. Set the number of clusters.
2. Determine the centroid of each cluster.
3. Determine the distance of each cluster to the centroid.
4. Group the vectors based on minimum distance.
5. Continue from step-2 until converges

The feature vectors from all training images are then jointly clustered with K-means clustering algorithm such that 1024 clusters are obtained. Then we discard all information for each patch except the centroid of the clusters.

Detection Framework

A wide variety of techniques have been developed for solving point pattern matching problems, including Hough transforms [22], geometric hashing [23], minimizing Hausdorff distance [24], etc. While these existing approaches have found numerous practical applications, they have some difficulties and limitations. Hough transforms, for example, require careful choice of parameters like bin size in order to reduce the risk of losing solutions.

The basis of the progressive RAST [25] algorithm is the RAST algorithm [26-28], which is guaranteed to find globally optimal solutions under given error models. RAST algorithms are closely related to (hierarchical) Hough transforms but have more desirable combinatorial properties for object recognition. Effective and simple to implement variants of RAST algorithms are based on interval arithmetic.

The steps involved in the Proposed Algorithm are as follows:

1. Salient points of the reference image are assigned with the Label from the cluster Centroids of training dataset.
2. Calculate the number of salient points in the image (*N*).
3. Divide the labeled reference image in to four sub images and calculate the number of salient points in each sub image, if it is greater than (*N/4*) of the total salient points further subdivide else reject that region.
4. Set up the priority queue.
5. Find the correspondence between the labels of test image and the members of the priority queue and increment a counter if more than 80% of the members gives a zero value.
6. If the counter value is more than 80% of the total number of sub images, then the test image is said to be matched with the reference image else mismatched.

The priorities of regions are compared based on two rules:

1. Smaller regions have higher priorities. This rule forces depth-first searching.
2. If two regions have the same size, the region with larger salient points has higher priority.

When a domain is subdivided in to small regions, the subdivision should meet the following two requirements:

1. The regions do not intersect with each other.
2. The union of the regions equals to the point set domain

3. RESULTS AND DISCUSSION

The proposed method was evaluated on the Caltech database (Visual Geometry Group) [29].

Table 1. Recognition Rate for Single Class Objects

Object	Number of Images Used for Training	Number of Images Used for Testing		Recognition Rate (%)	
		Positive Images	Negative Images	Positive Images	Negative Images
Airplane	250	824	25	98.66	100
Motorbike	200	626	25	97.12	92
Face	100	345	25	98.20	96
Car	30	96	25	95.83	84

objects (Airplanes, Faces, Cars and Motorbikes) and a set of background images not containing any of these objects are considered. The images are of various sizes and for the experiments they were converted to gray scale and resized to 225×520.

The Airplane dataset consists of 1074 images, Motorbike dataset contains 826 images, 445 images in Face dataset and 126 images in Car dataset (Visual Geometry Group). Images that have the object of interest are used for training. Nearly one fourth of the positive i.e., images with object are used for training and the remaining three fourth of the images are used for testing. Negative images, i.e., background images that do not contain the object of interest are also tested to verify the robustness of the algorithm.

In the experiments, the decision if a test image belongs to the object or background class was based on the number of salient points of the test image that matched with the reference image. The recognition rate obtained for single class objects is shown in Table **1**.

In case of Airplane vs. Background 250 Airplane images are trained and 824 Airplane images and 25 Background images were tested. The recognition rate obtained in case of testing airplane images is 98.66% and background images is 100% which shows that none of the background images are classified as Airplane.

In case of Motorbike dataset 200 images are trained and 626 images are tested. The recognition rate obtained for Motorbike dataset is 97.12% and for background images, it is about 92%. The reduced recognition rate is due to the rotation in the wheels of the Motorbike and due to partial occlusion.

In case of Face dataset 100 images are trained and 345 images are tested. The recognition rate obtained in case of Face dataset is about 98.20% and for background images, it is about 96%. The recognition rates get reduced since the patches extracted from the background of the Face image gets matched with the Background image.

In Car images, 30 Car images are trained and 96 Car images are tested. The recognition rate obtained in testing Car images are 95.83%. In case of testing 25 negative images 84% recognition rate is obtained. This is due to the fact that the dataset consists of car images that show the rear part and further the salient points are more at the background of positive images

3.3. Multi-class Object Recognition

Multi-Class object recognition is carried out with (i) Air Vehicles, such as, Airplanes and Helicopters, (ii) Road vehicles, such as, Bicycles, Motorbikes and Cars and (iii) Fruits, such as Mango, Jackfruit and Banana.

In the case of multi-class objects: Airplane and Helicopter, 50 Airplanes and 25 Helicopters are jointly trained and the test image includes 100 Airplanes and 50 Helicopters from Caltech database. The recognition rate obtained is 94% for Airplane and 86% for Helicopter. Table **2** shows the recognition rate obtained for multiclass object task: Airplane and Helicopter.

Table 2. Recognition Rate for Multiclass Objects: Airplane and Helicopter

Object	Number of Images Used for		Recognition Rate (%)
	Training	Testing	
Airplane	50	100	94
Helicopter	25	50	86

In the case of Multiclass objects: Bicycle, Motorbike and Car, 25 images each of Bicycle and Car and 50 images of Motorbike are jointly trained and double the number of trained images in each class are tested. The recognition rate obtained is 88% for Bicycle, 82% for Motorbike and 86% for Car.

In Multi-class tasks, misrecognition is said to occur when one class of object is recognized as object from another class for example, Bicycle recognized as Motorbike, Car recognized as Bicycle etc. This is due to the fact that, the test images contain many different viewpoints and partial occlusion. The recognition rate obtained for multiclass object: Bicycle, Motorbike and Car are shown in Table **3**.

Table 3. Recognition Rate for Multiclass Object: Bicycle, Motorbike and Car

Object	Number of Images Used for		Recognition Rate (%)
	Training	Testing	
Bicycle	25	50	88
Motorbike	50	100	82
Car	25	50	86

4. CONCLUSION

In this work, the proposed method focuses on object recognition in “complex” images which proves to be a challeng-

ing task, as several factors complicate a successful recognition, among them clutter, occlusion and object transformations like translation and scaling.

This method uses image patches as features for recognition. Concerning the patch locations, wavelet-based salient points are used for better performance. Salient points are detected using wavelet transform and the zerotree representation of the wavelet coefficients. This approach can be used in emerging multimedia applications. Here compactly supported orthogonal wavelets are considered and the salient point reflects the local maximum of the wavelet coefficients. Salient points detecting method based on wavelet transform yields optimum results when compared with traditional methods of interest point detection. This gives the points both in homogenous regions as well as in regions of high variance.

Several patch sizes are tested to represent small, middle-sized and large object parts and found out that 11×11 pixels outperforms than other patch sizes. Regarding the PCA transformation, 40 PCA coefficients are encountered to be sufficient. In case of brightness normalization it is discovered that it does not improve the results.

Here, the proposed algorithm is used for recognizing objects which are represented by patches. The results on the four datasets of the Caltech in case of single class objects and also the two tasks in case of multiclass objects proved that the proposed method is able to successfully recognize objects under challenging conditions.

REFERENCES

[1] Hegerath, "*Patch-based Object Recognition*", Diploma thesis, Human Language Technology and Pattern Recognition Group, RWTH Aachen University, Aachen, Germany, 2006.

[2] M. C. Burl, M. Weber, and P. Perona, "A probabilistic approach to object recognition using local photometry and global geometry", in *European Conference on Computer Vision*, 1998, Vol. 2, pp. 628-641.

[3] M. Weber, M. Welling, and P. Perona, "Unsupervised learning of models for recognition", in *Sixth European Conference of Computer Vision*, Dublin, Ireland, 2000, Vol. 1, pp. 18-32.

[4] S. Agarwal and D. Roth, "Learning a sparse representation for object detection", in *Seventh European Conference on Computer Vision*, 2002, Vol. 4, pp. 113-130.

[5] S. Agarwal, A. Awan, and D. Roth, "Learning to detect objects in images *via* a sparse, part-based representation". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 26, No. 11, pp. 1475-1490, 2004.

[6] G. Dorko, and C. Schmid, "Selection of scale-invariant parts for object class recognition", in *Ninth IEEE International Conference on Computer Vision*, 2003, vol. 1, pp. 634-639.

[7] R. Fergus, P. Perona, and A. Zisserman, "Object class recognition by unsupervised scale-invariant learning", in *IEEE Conference on Computer Vision and Pattern Recognition*, 2003, Vol. 2, pp. 264-271.

[8] B. Leibe, and B. Schiele, "Interleaved object categorization and segmentation", in *British Machine Vision Conference*, Norwich, UK, 2003, pp. 759-768.

[9] P. Carbonetto, G. Dorko, and C. Schmid, "*Bayesian learning for weakly supervised object classification*", Technical report, INRIA Rhone-Alpes, Grenoble, France, 2003.

[10] G. Csurka, L. Dance, J. Willamowski, and C. Bray, "Visual categorization with bags of key points" in *ECCV Workshop on statistical Learning in computer vision*, 2004, pp. 59-74.

[11] E. B. Sudderth, A. Torralba, W. T. Freeman, and A. S. Willsky, "Learning hierarchical models of scenes, objects, and parts", in *Tenth IEEE International Conference on Computer Vision*, 2005, Vol. 2, pp. 1331-1338.

[12] T. Deselaers, D. Keysers, and H. Ney, "Discriminative training for object recognition using image patches", in *International Conference on Computer Vision and Pattern Recognition*, 2005, Vol. 2, pp. 157-162.

[13] T. Deselaers, D. Keysers, and H. Ney, "Improving a discriminative approach to object recognition using image patches", in *27th DAGM Symposium*, 2005, pp. 326-333.

[14] D. Keysers, T. Deselaers, and T. M. Breuel, "Optimal geometric matching for patch-based object detection*"*, *Electronic Letters on Computer Vision and Image Analysis*, Vol. 6, No. 1, pp. 44-54, 2007.

[15] C. Schmid, and R. Mohr, "Local gray value invariants for image retrieval", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 19, No. 5, pp. 530-535, 1997.

[16] C. Harris, and M. J. Stephens, "A combined corner and edge detector", in *Alvey Vision Conference*, 1988, pp. 147-152.

[17] C. H. Chen, J. S. Lee, and Y. N. Sun, "Wavelet transformation for gray-level corner detection", *Pattern Recognition*, Vol. 28, No. 6, pp. 853-861, 1995.

[18] E. Loupias, N. Sebe, S. Bres, and J-M. Jolion, "Wavelet-based salient points for image retrieval", in *International Conference on Image Processing*, 2000, Vol. 2, pp. 518-521.

[19] Shapiro, J. M, "Embedded image coding using zero trees of wavelet coefficients" *IEEE Transactions on Signal Processing*, Vol. 41, No. 12, pp. 3445-3462, 1993.

[20] R. Vidal, and Yi Ma Sastry, "Generalized principal component analysis", in *International Conference on Computer Vision and Pattern Recognition*, 2003, Vol. 1, pp. 621- 628.

[21] K. Tapas, D. M. Mount, N. S. Netanyahu, C. D. Piatk, R. Silverman, and A. Y. Wu, "An efficient k-means clustering algorithm: analysis and implementation*"*, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 24, No. 7, 2002.

[22] J. Illingworth, and J. Kittler, "A survey of the hough transform", *Computer Vision, Graphics and Image Processing*, Vol. 44, pp. 87-116, 1998.

[23] H. J. Wolfson, and I. Rigoutsos, "Geometric hashing: An overview", *IEEE Computational Science and Engineering*, Vol. 4, No. 4, pp. 10-21, 1997.

[24] D. P. Huttenlocher, G. A. Klanderman, and W. J. Rucklidge, "Comparing images using the Hausdorff distance", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 15, No. 9, pp. 850-863, 1993.

[25] M. Ni, and S. E. Reichenbach, "A statistics-guided progressive RAST algorithm for peak template matching in GCxGC", in *IEEE Workshop on Statistical Signal Processing*, 2003, pp. 383-386.

[26] T. M. Breuel, "*Geometric Aspects of Visual Object Recognition*", Ph. D. thesis, Massachusetts Institute of Technology, USA, 1992.

[27] T. M. Breuel, "Recognition by Adaptive Subdivision of Transformation Space: practical experiences and comparison with the Hough transform", *IEE Colloquium on 'Hough Transforms'*, Vol. 7, (Digest No. 106), pp. 1-4, 1993.

[28] T. M. Breuel, "Fast recognition using adaptive subdivision of transformation space", in *International Conference on Computer Vision and Pattern Recognition*, Champaign, 1992, pp. 445-451.

[29] Visual Geometry Group. Available: www. robots. ox. ac. uk/~vgg/data/data-cats. html

Received: June 29, 2009 Revised: August 07, 2009 Accepted: August 07, 2009