# Fusion of Global Shape and Local Features Using Boosting for Object Class Recognition

Noridayu Manshor, Amir Rizaan Abdul Rahiman, Mandava Rajeswari, and Dhanesh Ramachandram

*Abstract*—**In object class recognition, the state-of-the-art works shows using combination varies local features may produce a good performance in recognition. These local features may have a different performance on one category to other category which it depends on the richness of local features. Due to that limitation, the shape features of objects are taken into consideration to be combined with local features. In this paper, we use Fourier Descriptor (FD), Elliptical Fourier Descriptors (EFD) and Moment Invariant (MI) as a global shape feature and Scale Invariant Feature Transform (SIFT) as local features. For learning technique, boosting is used in improving the recognition objects. This approach identifies the correct and misclassified dataset iteratively. Experimental results indicate that the recognition model outperform improved the accuracy of classification by up to 10% that is comparable to or better than that of state-of-the-art approaches.**

*Index Terms*—**Boosting, classification, global features, local features**

## I. Introduction

Object class recognition has recently received attention from the vision research community. It is a challenging problem in computer vision especially in the presence of intra-class variation, clutter, occlusion, and pose changes. Compared to the recognition of specific individual objects from images (e.g. different images of the same car), object class recognition involves classification of objects belonging to a class such as car, motorbike, or human face with different instances of the object, (e.g. images of different cars).The difficulty of recognizing classes of objects requires methods of comparing images that capture the variation within the class while discriminating against objects from different classes. At a higher level of human understanding, it is sufficient to identify the category or class of the object. The object class recognition problem is also termed as generic object recognition [1] and object categorization [2].

Lots of studies in object class recognition previously used local features for recognizing object classes. Local features ref er to the features that are extracted based on the interest points detected on the object. The features are extracted around the interest points in an object patch. What makes local features appealing is the ability to examine the variability of object classes, which consists of different scales, sizes, poses, etc. However, many object classes such as "cups", "horses" or "cows" are better described by shape features compared to local features. For example, "cups" objects have limited local features, which makes them difficult to discrimination with others classes. Consequently, local features may give poor recognition results. Local features focus on the local information of objects without considering other properties such as shape. This causes a problem for the computer to recognize objects that have limited or plain local features [3]. Thus, shape features are often used as a replacement of, or complement to local features in several works, such as [1], [4], [5].

For learning method, many researchers reported that Boosting [1], [6], [7] has shown improvement on many recognition problems which iteratively learning classifiers by reweighting the data. Boosting combine all different types of features in one feature pool and homogenous classifiers are used to train those features iteratively in Boosting. Recently, most object class recognition approaches exists in the literature use Boosting approach. However, the low level features used on those works are different. [1] combine three interest point detectors together with four types of local features, namely subsampled gray values, basic intensity moments, moment invariants, and SIFT. [6] used the others local features Gradient Location-Orientation Histogram (GLOH) and opponent angle color descriptor. Another object class recognition approach using different features from PCA-SIFT, shape context and spatial features is presented by [8]. Their model is a multi layer boosting system which the first boosting layer chooses the most important features from a pool of PCA-SIFT descriptors and shape-context descriptors. To improve the performance of the first boosting layer, the spatial relationships between the selected features are computed in the second layer of boosting. With this technique, the most authors only focused on local features without taking into account the shape of objects.

Thus, in this study, we combine two different types of low level features, global shape features and local features for the purpose of object class recognition. We intend to use these two types of low level features due to the important of shape and local features in recognizing unrestricted poses of object class. Furthermore, increasing the number of visual features will increase the recognition performance. The boosting is used to learn the combination of proposed features.

## II. Methodology

### A. Dataset

The objects in an image generally have many variations of

Manuscript received July 13, 2012; revised August 13, 2012.This study has been supported by the funds from Universiti Putra Malaysia's Research University Grant Scheme: (Project Number: 05-01-11-1259RU/F1, Cost Center: 9199868).

Noridayu Manshor and Amir Rizaan Abdul Rahiman are with Faculty of Computer Science and Information Technology, Universiti Putra Malaysia (e-mail: ayu@ fsktm.upm.edu.my, amir@fsktm.upm.edu.my).

Mandava Rajeswari and Dhanesh Ramachadram are with School of Computer Science, Universiti Sains Malaysia (e-mail: {mandava, dhaneshr}@cs.usm.my)

poses and orientation in different scales. Therefore, Graz02 dataset [1], [6], [9] is used to give more realistic and less restrictive object condition. Among the datasets that are available for computer vision and image processing research, the Graz02 dataset is one of the most challenging dataset due to the range of image variability, such as different scales and views (Fig. 1). This dataset prepared at Graz University of Technology [1].

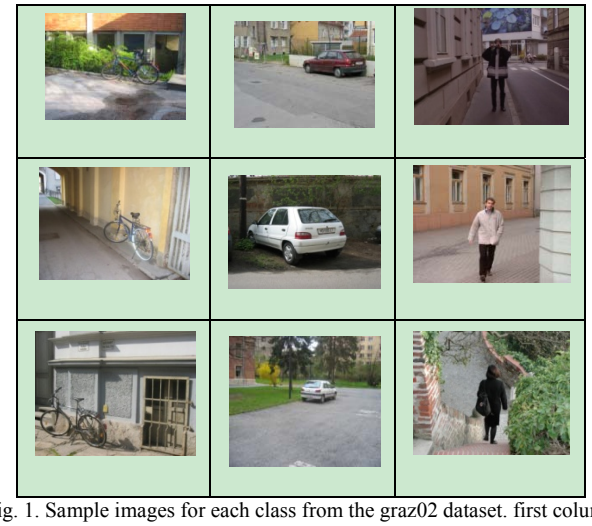

Fig. 1. Sample images for each class from the graz02 dataset. first column presents 'bikes' class, followed by 'cars' and 'persons' classes.

### B. Global Shape Features

Shape is an important part of the semantic content of images and it should be the main feature in recognizing object classes [10], [11]. This research focuses the boundary based shape features which describe the whole contour of object class. To get more generalized shapes, it depends heavily on the segmentation process, or based on the detection of shape contours. The major advantage of global shape features is that they can be extracted and matched with minimal computational time [21]. They are insensitive to surface features such as texture, color features and also invariant to lighting conditions. Furthermore the shape of objects is easily to encode. Good recognition accuracy requires an effective shape features to look similar to the interpretation to human perceptual [12]. The boundary based features that are used in this study are discussed below to understand the theory of those features.

*1) Fourier descriptor (FD)*

The boundary (outline) of the object is treated as lying in a complex plane [13] which the row and column co-ordinates of each point on the boundary (outline), $B(k)=[x(k),y(k)], k=0,1,...,K-1$ can be expressed as a complex number as denoted

$$b(k)=x(k)+jy(k) \tag{1}$$

where $j$ is $sqrt(-1)$. The boundary point is started at an arbitrary point, $(x_0,y_0)$ and tracing once around in the counterclockwise direction at a constant speed yields a sequence of coordinates that represented by complex numbers. Dealing with discrete images, the Discrete Fourier Transform (DFT) is applied. The DFT of $b(k)$ is defined as

$$DFT(u)=\sum_{k=0}^{K-1} b(k)e^{-j2\pi uk/K} \tag{2}$$

For $u=0,1,2,...,K-1$ . The complex coefficient $DFT(u)$ are called the Fourier Descriptors of the boundary which gives the shape of an object. The inverse of Fourier transform of these coefficients restores $b(k)$ where $k=0,1,2,...,K-1$ as following:

$$b(k)=\frac{1}{K}\sum_{u=0}^{K-1} DFT(u)e^{j2\pi uk/K} \tag{3}$$

The inverse Fourier Descriptors is computed by specifying number of descriptors, to yield a closed spatial curve.

*2) Elliptical fourier descriptors (EFD)*

EFD is apply to the closed contour of object by defining with differential chain code, represented as a point coordinate of closed contour.

Length $(dt_i)$ of element $(v_i)$ of the chain code is given by the equation 4.

$$dt_i=1+\left(\frac{\sqrt{2}-1}{2}\right)(1-(-1)^{v_i}) \tag{4}$$

Therefore, for the whole number of element in a contour, the length is,

$$t_n=\sum_{i=1}^{n} dt_i \tag{5}$$

Following equation presents the projection of each $v_i$ , on the $X$ and $Y-$axis, respectively,

$$dx_i=sign(6-v_i)*sign(2-v_i),$$
$$dy_i=sign(4-v_i)sign(v_i) \tag{6}$$

For all element of the chain, $p$ , the projection on the $X$ and $Y-$axis will be,

$$x_p=\sum_{i=1}^{p} dx_i,$$
$$y_p=\sum_{i=1}^{p} dy_i \tag{7}$$

The EFD is calculated from the sum of elliptical harmonics. In identifying the closed contour points, $K,N$ harmonics are considered. [14] use for each harmonic, four Fourier coefficients $a_n,b_n,c_n$ and $d_n$. Equation 8 presents these four coefficients. These harmonics and their corresponding coefficients are used to produce coordinates that define ellipses that fit within the object's outline to represent the object's shape.

$$a_n=\frac{T}{2n^2\pi^2}\sum_{i=1}^{k}\frac{dx_i}{dt_i}\left[\cos\frac{2n\pi t_i}{T}-\cos\frac{2n\pi t_{i-1}}{T}\right]$$

$$b_n = \frac{T}{2n^2\pi^2}\sum_{i=1}^{k}\frac{dx_i}{dt_i}\left[\sin\frac{2n\pi t_i}{T} - \sin\frac{2n\pi t_{i-1}}{T}\right]$$

$$c_n = \frac{T}{2n^2\pi^2}\sum_{i=1}^{k}\frac{dy_i}{dt_i}\left[\cos\frac{2n\pi t_i}{T} - \cos\frac{2n\pi t_{i-1}}{T}\right]$$

$$d_n = \frac{T}{2n^2\pi^2}\sum_{i=1}^{k}\frac{dy_i}{dt_i}\left[\sin\frac{2n\pi t_i}{T} - \sin\frac{2n\pi t_{i-1}}{T}\right] \quad (8)$$

*3) Moment invariants (MI)*

This features are extracted from boundary and interior region of an object. In this study, the moment invariants are extracted from segmented objects based on boundary points. The moment invariant from Hu [15] proposed seven expression calculated from normalized central moments that are invariant to object scales, translations and rotations.

This feature is used in this research because it can represent different geometrical features in the objects. It also can be applied for disjoint shapes that cannot be supported by FD [16]. Discrete central moments of an image, $f(x,y)$ are defined by the following equation:

$$\mu_{P,q} = \sum_{x=1}^{M}\sum_{y=1}^{N}(x-\bar{x})^p(y-\bar{y})^q f(x,y) \quad (9)$$

The normalized cental moments are given by:

$$\eta'_{p,q} = \frac{\mu_{p,q}}{\mu^{\gamma}{}_{0,0}}, \quad (10)$$

where $\gamma = \frac{p+q}{2}+1$ and $p+q = 2,3,...$

The seven moment invariants introduced by Hu are derived from aforementioned equations are:

$$\begin{aligned}
\phi_1 &= \eta'_{2,0} + \eta'_{0,2},\\
\phi_2 &= (\eta'_{2,0} + \eta'_{0,2})^2 + 4\eta'^2{}_{1,1},\\
\phi_3 &= (\eta'_{3,0} - 3\eta'_{1,2})^2 + (3\eta'_{2,1} - \eta'_{0,3})^2,\\
\phi_4 &= (\eta'_{3,0} - \eta'_{1,2})^2 + (\eta'_{2,1} + \eta'_{0,3})^2,\\
\phi_5 &= (\eta'_{3,0} - 3\eta'_{1,2})(\eta'_{3,0} + \eta'_{1,2})[(\eta'_{3,0} + \eta'_{1,2})^2 -\\
&\quad 3(\eta'_{2,1} + \eta'_{0,3})^2] + (3\eta'_{2,1} - \eta'_{0,3})(\eta'_{2,1} + \eta''_{0,3})\\
&\quad [3(\eta'_{3,0} + \eta'_{1,2})^2 - (\eta'_{2,1} + \eta'_{0,3})^2]\\
\phi_6 &= (\eta'_{2,0} - \eta'_{0,2})[(\eta'_{3,0} + \eta'_{1,2})^2 - (\eta'_{2,1} + \eta'_{0,3})^2] +\\
&\quad 4\eta'_{1,1}(\eta'_{3,0} + \eta'_{1,2})(\eta'_{2,1} + \eta'_{0,3})\\
\phi_7 &= (3\eta'_{1,2} - \eta'_{3,0})(\eta'_{3,0} + \eta'_{1,2})[(\eta'_{3,0} + \eta'_{1,2})^2 -\\
&\quad 3(\eta'_{2,1} + \eta'_{0,3})^2] + (3\eta'_{2,1} - \eta'_{0,3})(\eta'_{2,1} + \eta'_{0,3})\\
&\quad [3(\eta'_{3,0} + \eta'_{1,2})^2 - (\eta'_{2,1} + \eta'_{0,3})^2]
\end{aligned} \quad (11)$$

### C. Local Feature

Harris-Affine detector is one of the best feature detectors is chosen to detect the interest points of the object [17], [18] in this study. The aspect that consider in this selection according to this detector can support affine transformation of objects such as from rotation, illumination, and deformation perspectives. Once the regions of interest are extracted we have to use an appropriate local feature for them. SIFT features is a local features which robust to some image variations such as viewpoints, orientation, illumination and scales [19]. By this characteristic, SIFT features is appropriate for the task of object class recognition. It gives high probability of accurate matches across wide range of image variations. Thus, this study combines this feature with global shape features for improving the result in object class recognition. The extracted SIFT features from all objects in earlier steps are used to construct the visual vocabulary. The procedure involves in this stage is clustering all the features of the same class based on the number of cluster determined. This mean that visual vocabulary will contain the entire of acquired cluster centers which correspond to the prototype object features.

The new object features can be defined based on the visual vocabulary constructed in previous step. It is obtained by counting how many times each visual word occurs within the object. From this process, the feature histogram of each object can be generated. Each object will have distinct feature histogram, but the assumption that, each object shows the similar patterns for object in the same category and dissimilar patterns for object in different categories.

In this study, empirically, 40 descriptors of FD and 28 EFDs are used in this study. This number accurately describes the shape of objects. Therefore, FD consists of 40-dimension, EFD 28-dimension and MI is 7-dimension. For SIFT features, each class is clustered using K-Means algorithm with *K= 100* [1] and end up for 300 as a total of vocabulary, *V* size (three concepts).

### D. Learning Method

Adaboost algorithm is used to learn the combination of those features in improving the classification accuracy for object class recognition. The intuitive idea behind Adaboost [20] is to train a series of weak classifiers and to iteratively improve performance of those classifiers. The algorithm relies on continuously changing the weights of training set so that those that are frequently misclassified get higher weights: this way, new classifiers that are added to the set are more likely to classify those hard examples correctly. The equal weight is initialized to all instances in the training data. The weak classifier is trained iteratively and each instance of training data is reweighted according to the weak classifier's output. The weight is decreased when the instances are correctly classified; whereas misclassified ones are increased hence the weak learner will focus mostly on hard examples. Once all the weak classifiers have been trained, their predictions are then combined through weighted majority voting scheme. The boosting algorithm is summarized in Fig. 2.

Given training data $T = \{(x_1, y_1),...,(x_n, y_n)\}$

Initialize data weighting coefficients, $w_n^{(1)} = \frac{1}{N}$ .

For $m = 1,...,M$ :

Fit a classifier $h_m(x)$ to training data by using

distribution $w_n^{(m)}$

Evaluate the weighted training error of classifier $m$:

$$\varepsilon_m = \sum_{n=1}^{N} w_n^{(m)} I\{h_m(x) \neq y_n\}$$

Then use $\varepsilon_m$ to evaluate

$$\alpha_m = \frac{1}{2}\ln\{\frac{1-\varepsilon_m}{\varepsilon_m}\}$$

Update data weighting coefficients

$$w_n^{(m+1)} = \frac{w_n^{(m)} \exp\{-\alpha_m y_n h_m(x_n)}{Z_m}$$

where $Z_m$ is a normalization factor.

Make prediction using the final classifier

$$f(x) = sign(\sum_{m=1}^{M} \alpha_m h_m(x))$$

Fig. 2. Adaboost algorithm

## III. Result and Discussion

The classification performance of our proposed methods is compared to state-of-the-art researches in object class recognition using the similar dataset. Besides that, the performance of proposed features in our studies is compared to those previous works using similar Boosting approach. This to evaluate the effectiveness of combination those features may improve the classification accuracy for object class recognition. The parameters of Boosting is followed from [1], [6]. The weight of training dataset are initialized to *1* and the number of iteration of training a weak learner is *T=150*. The size of training data and testing data from Graz02 dataset is adopted from [1]. For training, 150 positive samples and 150 negative samples are used. The total of testing sample is 150, where 75 positive and 75 negative samples. Negative samples consist of the remaining two concepts. For instance 75 positive 'bikes' object and 75 negative consists of 'cars' and 'persons' class.

Table I shows ROC-equal error rates result for all classes compare to other works. As shown in Table I, we observe that combining global shape and local features improve the classification state-of-art features about 10-15% for 'bikes' and 'persons' class and have less performance of [6] in the 'cars' dataset. From this result, both features play an important role to improve the accuracy of object class recognition rather than only focusing on local features as proposed by [1], [6].

TABLE I: Comparison of Roc-Equal Rates with Other Works Using Boosting Approach

| | Our | [6] | [1] |
|---|---|---|---|
| **Bikes** | 0.891 | 0.747 | 0.778 |
| **Cars** | 0.737 | 0.813 | 0.705 |
| **Persons** | 0.920 | 0.813 | 0.812 |

Besides that, the recognition result for single feature is shown in Table II for each object class. Generally, the result shows that single feature do not give the good result as Table II. By performing recognition using single feature, the global shape features giving better results than the local feature.

TABLE II: Comparison of Roc- equal Rates Using Single Feature (Boosting approach)

| | FD | EFD | MI | SIFT |
|---|---|---|---|---|
| **Bikes** | 0.870 | 0.780 | 0.836 | 0.660 |
| **Cars** | 0.734 | 0.887 | 0.507 | 0.410 |
| **Persons** | 0.910 | 0.566 | 0.74 | 0.660 |

## IV. Conclusion

The work in this paper has presented the effects of combination of global shape and local features for object class recognition. The boosting learning algorithm is performed to train those combinations. It is clear that the both different types of feature shows high performance rather than when using them separately in recognition. The performance of our object class recognition model exceeds many of the object class recognition state-of-the-art approaches using benchmark datasets. Also, comparison using single feature is performed which reveals that global shape features has higher performance than local feature. Thus, both features need to be taken into account in recognizing the complex datasets.

For future studies, we plan investigate other decision fusion techniques. This will mainly involve learning algorithms where each feature, global shape features and local feature is independently trained by individual classifiers.

## References


[1] A. Opelt and A. Pinz, et al, "Generic Object Recognition with Boosting," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 3, pp. 416-431, 2006.

[2] G. Csurka, C. Dance, L. Fan, J. Willamowski, and C. Bray, "Visual Categorization With Bags of Keypoints," *Pattern Recognition and Machine Learning in Computer Vision Workshop*. Grenoble, France, 2004.

[3] A. Mansur and K. Yoshinori, " Integration of Multiple Methods for Robust Object Recognition," *International Symposium on Visual Computing*, 2006.

[4] X. Yu and Y. Li, et al, Object Detection Using Shape Codebook. *In Proc.* British *Machine Vision Conference (BMVC'07)*, pp. 1-10, 2007.

[5] J. Shotton and A. Blake, et al, "Multi-Scale Categorical Object Recognition Using Contour Fragments." *IEEE Trans. Pattern Anal. Mach. Intell*, vol. 30, no. 7, pp. 1270-1281, 2008.

[6] D. Hegazy and J. Denzler, "Generic Object Recognition using Boosted Combined Features," *Second International Workshop*, *Rob Vi*, Auckland, New Zealand, 2008.

[7] P. Negri and X. Clady, et al, "A cascade of boosted generative and discriminative classifiers for vehicle detection," *Eurasip Journal on Advances in Signal Processing*, 2008, pp. 1-12, 2008.

[8] W. Zhang and B. Yu, et al, "Object Class Recognition Using Multiple Layer Boosting with Heterogeneous Features," *2005 IEEE Computer Society Conference on Computer Vision And Pattern Recognition (CVPR'05*, vol. 2, pp. 323-330.

[9] M. Marszałek and C. Schmid, "Accurate Object Localization with Shape Masks, *IEEE Conference on Computer Vision and Pattern Recognition*, 2007.

[10] D. Zhang and G. Lu, "A comparative study on shape retrieval using Fourier descriptors with different shape signatures," *in Int. Conf. on Intelligent Multimedia and Distance Education*, Citeseer, 2002.

[11] J. P. Eakins, "Towards intelligent image retrieval," *Pattern Recognition*, vol. 35, no. 1, 2002.

[12] M. Yang and K. Kpalma, et al, "A survey of shape feature extraction techniques," *Pattern recognition techniques, technology, and applications*, pp. 626, 2006.
[13] C. T. Zahn and R. Z. Roskies, "Fourier Descriptors for Plane Closed Curves." *IEEE Transactions on Computers*, vol. 21, pp. 269-281, 1972.
[14] F. P. Kuhl and C. R. Giardina, "Elliptic Fourier features of a closed contour* 1," *Computer graphics and image processing*, vol. 18, no. 3, pp. 236-258, 1992.
[15] M. K. Hu, "Visual pattern recognition by moment invariants." *IRE Transactions on Information Theory*, vol. 8, pp. 179-187, 1962.
[16] Q. Chen, "Evaluation of OCR algorithms for images with different spatial resolutions and noises," *Citeseer*, 2003.
[17] K. Mikolajczyk and C. Schmid, "Scale & affine invariant interest point detectors." *International Journal of Computer Vision*, vol. 60, no. 1, pp. 63-86, 2004.
[18] K. Mikolajczyk and T. Tuytelaars, et al, "A comparison of affine region detectors," *International Journal of Computer Vision*, vol. 65, no. 1, pp. 43-72, 2005.
[19] D. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," *Int. Journal on Computer Vision*, vol. 60, 2004, pp. 91-110.
[20] Y. Freund and R. E. Schapire, "Experiments With A New Boosting Algorithm, " *Morgan Kaufmann Publishers*, Inc, 1996.
[21] T. Glatard and J. Montagnat, et al, Texture based medical image indexing and retrieval: application to cardiac imaging, *Workshop on Multimedia Information Retrieval (MIR)*, NY, USA, 2004.

**Noridayu Manshor** received the B.S degree in Computer Science from Universiti Putra Malaysia (UPM) in 2000 and the M.Sc in 2004 from Universiti Teknologi Malaysia (UTM), and currently pursuing her Ph.D degree in Computer Science at Universiti Sains Malaysia (USM), Malaysia. Currently, she is a lecturer at Faculty of Computer Science and Information Technology, UPM. Her current research interests include image processing, computer vision and pattern recognition.

**Amir Rizaan Abdul Rahiman** received the B.S degree in Computer Science from University Putra Malaysia (UPM) in 2000 and the M.Sc and Ph. D degrees in Computer Science from University Teknologi Malaysia (UTM), and University Sains Malaysia (USM), Malaysia in 2004 and 2011, respectively. Currently, he is lecturer at Faculty of Computer Science and Information Technology, UPM. His current research interests include multimedia applications, e-learning solution, flash-based storage systems, and multimedia storage systems.

**Mandava Rajeswari** received the M.Tech in 1980 from Indian Institute of Technology, Kanpur and Ph.D degrees in 1995 from University of Wales Swansea. Join USM in 1982. Her main research interest is to process, analyze and to extract contents and information from the images; derive knowledge from the extracted information; to represent the knowledge and use the knowledge in various applications in addition to using it to guide the information extraction from the images. In the early stages of this research the focus was to extract information from the images and put into several applications that include automated visual inspection, and real time process control in industry; robot vision for intelligent assembly; image database retrieval and image segmentation. The major domain of research is in medical images and natural images.

**Dhanesh Ramachandram** received the Ph.D degree in 2003 from USM. *C*urrently, he is a lecturer at School of Computer Sciences, Universiti Sains Malaysia. His main research includes Segmentation of Multiple Sclerosis Lesions in MRI, Context in Object Class Recognition, Kernel Methods in Machine Learning and Ensemble Classifiers for Object Class Recognition