FUSION OF GLOBAL SHAPE AND LOCAL FEATURES USING MULTI CLASSIFIER FRAMEWORK FOR OBJECT CLASS RECOGNITION

by

NORIDAYU MANSHOR

Thesis submitted in fulfillment of the requirements
for the degree of
Doctor of Philosophy

October 2013

Acknowledgements

"In the name of God, most Gracious, most Compassionate".

All praise belongs to God for the countless blessings He has bestowed upon me.

First and foremost, I would like to express my deepest gratitude towards my supervisor, Prof. Dr. Mandava Rajeswari, for her invaluable inspiration, guidance, support, and dedication through the challenging journey of pursuing Ph.D. I am forever grateful for the chance to work in her research group. I also thank to my co-supervisor Associate Professor Dr. Dhanesh Ramachandram for all the useful ideas and advice throughout my research.

I'm also sincerely thankful to my employer, University Putra Malaysia (UPM) and the Ministry of Higher Education (MOHE) Malaysia for awarding a full scholarship support during my studies. To all my colleagues in the Faculty of Computer Science and Information Technology (FSKTM), UPM, thanks for your very useful suggestions, comments, and motivations in improving the research.

Not forgetting my friends at the Computer Vision Research Group (CVRG) at Universiti Sains Malaysia. Thank you very much Alfian, Mozaher, Anusha, Adel, Ehsan, Osama and Jawarneh for many interesting and inspiring discussions, sharing the memorable moments that we had at the lab. I will forever remember our moments together.

Last but not least, forever gratitude towards my beloved husband, Amir Rizaan, my sons, Danial and Danish and my sweet twins Arissa and Aneesa for their understanding, prayer, love and support. Special thanks to my parents, papa, mama and abah, mak and my siblings for invaluable advice, assistance and love.

Thank you…

Tables of Contents

List of Tables

List of Figures

List of Symbols and Abbreviations

BC	Base Classifier
BF	Boundary Fragment
BoK	Bag of Keypoints
CBIR	Content-based Image Retrieval
CFS	Correlation-based Feature Selection
DFT	Discrete Fourier Transform
EFD	Elliptical FD
ER	Error Rate
FD	Fourier Descriptors
FN	False Negative
FP	False Positive
GLOH	Gradient Location and Orientation Histogram
HOG	Histogram of Oriented Gradient
ICA	Independent Component Analysis
LDA	Linear Discriminant Analysis
LSE	Least Square Estimation
MC	Meta-Classifier
MI	Moment Invariants
MLP	Multi Layer Perceptron
NB	Naïve Bayes
PCA	Principal Component Analysis
RBF	Radial Basis Function
ROC	Receiver-Operating-Characteristic
ROI	Regions of interest
SIFT	Scale-invariant Feature Transform
SOM	Self Organizing Maps
SVM	Support Vector Machine
TN	True Negative
TP	True Positive
UIUC	University of Illinois Urbana Campaign

GABUNGAN BENTUK GLOBAL DAN CIRI TEMPATAN MENGGUNAKAN RANGKA KERJA PELBAGAI PENGELAS UNTUK PENGECAMAN KELAS OBJEK

ABSTRAK

Pengecaman kelas objek berurusan dengan klasifikasi objek individu untuk kelas tertentu. Dalam imej yang semula jadi, objek muncul dalam pelbagai gaya dan skala, dengan atau tanpa oklusi. Objek pengiktirafan kelas biasanya melibatkan pengekstrakan, pemprosesan dan analisis sifat visual seperti warna, bentuk, atau tekstur dari objek, dan kemudian mengaitkan label kelas kepadanya. Dalam tesis ini, sifat rupa bentuk global dan tempatan dianggap sebagai ciri-ciri yang diskriminatif untuk pengiktirafan kelas objek. Bagi sifat tempatan, masalah klasifikasi berlaku jika objek itu adalah terlalu kecil dan mempunyai sifat tempatan yang lemah. Selain itu, sifat tempatan tidak memberi kepentingan tersirat kepada bentuk objek, yang merupakan salah satu sifat penting untuk penglihatan manusia. Mengecam objek adalah sukar jika terdapat perubahan gaya. Oleh itu, perubahan gaya akan mengakibatkan perubahan dalam sifat bentuk bagi sesuatu objek di dalam kelas yang sama. Oleh itu, kedua sifat, tempatan dan rupa bentuk digabungkan untuk mendapatkan klasifikasi prestasi yang lebih baik bagi setiap kelas objek. Kesudahannya, satu rangka kerja meta-pengelas dicadangkan sebagai model untuk pengecaman kelas objek. Meta-pengelas digunakan untuk mempelajari satu meta-pengelas yang optimum bagi meramalkan ketepatan pengelasan pengelas asas bagi setiap objek. Dalam rangka kerja ini, setiap individu pengelas dilatih menggunakan sifat tempatan dan rupa bentuk global. Kemudian, keputusan daripada individu pengelas digabungkan sebagai input kepada meta-pengelas. Keputusan eksperimen

menunjukkan model ini setanding, atau lebih tinggi prestasinya dengan kerja-kerja yang sedia ada bagi pengecaman kelas objek.

FUSION OF GLOBAL SHAPE AND LOCAL FEATURES USING MULTI CLASSIFIER FRAMEWORK FOR OBJECT CLASS RECOGNITION

ABSTRACT

Object class recognition deals with the classification of individual objects to a certain class. In images of natural scenes, objects appear in a variety of poses and scales, with or without occlusion. Object class recognition typically involves the extraction, processing and analysis of visual features such as color, shape, or texture from an object, and then associating a class label to it. In this thesis, global shape and local features are considered as discriminative features for object class recognition. For local features, misclassification problems occur if the object is too small and possess weak local features. Besides that, local features do not give implicit importance to the shape of the object, which is one of important features to human vision. Detecting objects is difficult if the pose changes. Consequently, pose changes will result in changes in shape features for an object in the same class. Hence, both local and shape features are combined in order to obtain better classification performance for each object class. Ultimately, a meta-classifier framework is proposed as a model for object class recognition. Meta-classifier is used to learn a meta-classifier that optimally predicts the correctness of classification of base classifier for each object. In this framework, individual classifiers are trained using the local and global shape features, respectively. Then, these classifiers results are combined as input to the meta-classifier. Experimental results have shown to be comparable, or superior to existing state-of- the-art works for object class recognition.

CHAPTER 1 : INTRODUCTION

1.1 Introduction

Due to the recent developments in technology, huge amounts of images are easily generated using relatively affordable devices such as digital cameras, video camcorders and mobile phones. The Internet has also allowed easy and ubiquitous access, indirectly contributing to the massive consumption of images data. Due to these, and also due to the wide availability of mass storage devices, the amount of images data is growing to colossal proportions.

In order for effective and intuitive retrieval, these images should be annotated. Annotation is the process of assigning meaningful labels to data, mostly via a set of keywords. For digital images for example, most image databases employ manual annotation (Gong, Zhang et al. 1994), which entails labeling an image using descriptive keywords that best explains it. Since the annotation process is done by experts, descriptions for an image are very detailed. However, such an annotation practice is time consuming and laborious, and tedious task for entering the description of images manually.

In order to circumvent manual annotation, automated techniques for annotating images are required. Automatic annotation is the process of assigning labels to images according to their visual content. This can be done using two approaches: 1) Global annotation – where an overall description of an image is given, and 2) Region or Object labeling – where annotations are done on individual image

components. Current research on automatic image annotation however, is more inclined towards the second approach (Sumathi, 2011).

The major drawback faced in first approach, is the lack of insufficient integration of human knowledge on images. In order to perform object-level annotation, regions of interest (ROI) or the object itself has to be extracted firstly from the image and its spatial relationships identified. The task of object classifications is known as Object Class Recognition. Other synonyms for Object Class Recognition are such as generic object recognition (Opelt et al. 2006a) and object categorization (Csurka, Dance et al. 2004). The main task in Object Class Recognition is to discriminate between objects of one class and those of other classes. The challenges of Object Class Recognition are to find class models that are invariant to changes in appearance within a class, while being discriminative enough to distinguish between objects from different classes.

Specifically, Object Class Recognition involves extracting features from an identified object, and then associating a label to it representing the object's class. An object class furthermore can contain various objects of the same genre. For example, the object class "flower" may consist of a variety of flowers and an object class "cars" may consist of cars of different brands and models with a variety of shapes and sizes. For instances, Figure 1.1 shows some examples of images where the car class appears. It is straightforward to perceive that these three cars are very different in terms of visual appearance, but all must be classified within the same class.

Figure 1.1: The ‘car’ class.

Objects can have a variety of poses, scales, with or without occlusion, depending on the viewing direction, angle, and distance. The object class can be understood by a computer based on its visual features such as shape, color and texture. The challenge is to map or relate these visual features to a higher level conceptual representation that is closer to human understanding. The discrepancy in understanding between machines and humans is known as the “Semantic Gap”. At this juncture, most related research efforts work on mapping an object within an image to a suitable concept (Opelt et al. 2006a, 2006b, 2006c; Shotton et al. 2009) . With Object Class Recognition, the annotations given to an object can be consistent with the meaning it has to convey.

1.2 Motivation

The visual features used in Object Class Recognition are normally local features (Mansur et al. 2007; Zhang et al. 2007; Hare et al. 2011). Local features are computed at multiple points in the image. Local features are those that are a representation of a group of pixels within a small local region. They have no bearing on the concept of a semantically meaningful region, either that of an object or the complete image. Local features are preferred since they have invariant properties that are robust to viewpoints, translation, rotation, etc. The popular local feature is Scale-invariant Feature Transform (SIFT) proposed by Lowe (2004), which use

local maxima of the difference-of-Gaussians function as interest points and histograms of gradient orientations computed around the points as the descriptors. The distinctiveness of SIFT features, as well as their abundance over a large range of image scales; makes them suitable for object recognition in cluttered images. A problem however arises when objects are too small, and do not have strong local features (Murphy, 2006). Sometimes, local similarity may not generate correct results. Therefore, SIFT features cannot be properly generated in such cases.

Global features can also be used for Object Class Recognition (Lisin et al. 2005; Oliveira at al. 2007). These features extracted to represent the whole object thereby capturing the appearance of an object (Kragic et al. 2009). For instance, color histogram represents the distribution of object colors. Another example are shape features which is one of the most prominent features humans base their recognition on. Psychological experiments have shown that natural objects are primarily recognized by their shape characteristics (Biederman 1987). In recent years however, shape has been ignored for Object Class Recognition. This trend though, is changing as works such as Ferrari and Schmid (2008), Leibe, Leonardis et al. (2008), and Shotton, Winn et al. (2009) have started to incorporate shape features into Object Class Recognition. These shape features are mostly shape-fragments, rather than the entire shape given by the full boundary/contour of the object. Object can be segmented based on color or texture similarities, to obtain an accurate shape representation. However, the results are normally suboptimal where objects tend to be under or over-segmented. Thus, shape-based Object Class Recognition is greatly dependent on the segmentation process. Moreover, shape features can also be ambiguous, especially when objects are only captured from one viewpoint. This

problem however, can sometimes be solved by capturing the object from various poses. Although, they have different poses, these objects can be categorized into similar class. To overcome this limitation, several papers take the advantageous from local features in combining with shape features to contribute to the improvement of Object Class Recognition (Mansur and Yoshinori, 2007; Opelt et al. 2006b; Zhang et al., 2005).

The problem of Object Class Recognition is not only related to the features point of view but also depending on the classifier design. In the past, fusing different classifiers has managed to improve classification accuracy (Sannen et al. 2010; Hegazy and Denzler 2008; Opelt et al. 2006a) The main idea is that, by combining different classifier outputs, higher accuracy can be achieved as opposed to using just one classifier. To improve classification accuracy, a suitable fusion method and selection of appropriate classifiers have to be taken into account. The fusion of classifiers can be done at the feature-level and decision-level. In the past, Content-based Image Retrieval (CBIR) researches fused several features into a single feature vector (Oliveira and Nunes 2008; Veltkamp and Tanase 2002). However, this has its limitations such as increased computational time due to the curse of dimensionality (Mangai et al. 2010; Faundez-Zanuy, 2009). To overcome this, fusing at the decision-level is more promising by constructing a multiple classifier for each image feature (Mangai at al. 2010; Antenreiter et al. 2009; Murphy et al. 2006). The final decision is identified based on combination outputs from each classifier.

In the case of this work, the classifier fusion is adopted due to the diversity of information from the local and global features. The final predicted object class result

is produced through the integration of outputs obtained from the discriminant function of different classifiers. The computational burden of the base classifier also motivates us to adopt classifier combination. Since the different classifier may produce different results, thus, classifier fusion can be used to balance the performance of a set of classifiers in order to increase the classification accuracy.

In this thesis, two challenges are addressed in two separate phases. In the first phase, the formulation of the specific global features and local features for identifying specific objects, which best represent objects in multiple views, rotation and scale. In the second phase, the development of an efficient algorithm for combining both features in the first phase is undertaken. This is to provide accurate classification for objects into their respective class or category. A set of individual classifiers are trained using local and global features of objects, and the outputs of all individual classifiers are combined to predict the final class of object.

1.3 Objective

The main objective of this research is to improve Object Class Recognition using classifier fusion. To achieve this objective, the following sub-objectives have to be accomplished:

- To propose and investigate the role of shape features and local features in improving Object Class Recognition.
- To investigate optimal fusion strategy to improve Object Class Recognition.

- To formulate a framework of classifier fusion for Object Class Recognition.

1.4 Scope of the Research

The scope of this work is as follows:

- This thesis deals with Object Class Recognition, and not with object segmentation. Thus, pre-segmented datasets are used where objects have already been fully segmented from their background to investigate the role of shape. The segmentation process is done by using manual or automatic segmentation to obtain a complete contour. Occluded objects are not considered. It may be noted that this is a common practice and is consistent with the popular Visual Object Classes Challenge (Everingham et al. 2010);

- The datasets are categorized into restricted views and unrestricted views based on benchmark datasets from Caltech (Fergus et al. 2003; Ponce et al. 2006; Opelt et al. 2006a; Hegazy and Denzler 2008) and Graz02 (Hegazy and Denzler 2008; Opelt et al. 2006a) to provide direct comparisons with other related works;

- The proposed method is tested on three concepts of rigid objects, namely ‘cars’, ‘airplanes’, ‘motorbikes’ and special categories ‘faces’ from the

Caltech[1] dataset and three concepts of rigid object from the Graz02 dataset, 'bikes', 'cars' and 'persons' (Opelt et al. 2006a).

1.5 Contribution

- Demonstrate the first use of full contour shape features - Fourier Descriptor, Elliptical Fourier Descriptor and Moment Invariant (global shape feature) and combining with SIFT (local feature) for Object Class Recognition.

Previous researchers mostly used a combination of different local features to classify the objects (Opelt et al. 2006a; Hegazy and Denzler 2008). Few researches had used global shape features combined with local features (Zhang et al. 2005; Oliveira et al. 2007). In recognizing invariant object classes, sometimes shape seems to be the more powerful feature and sometimes local features (Stark and Schiele 2007). Hence, one of the contributions of this thesis is to present how the shape-based features approach improves Object Class Recognition if these features are combined with the local feature.

- Develop an efficient meta-classifier model for Object Class Recognition.

Another important aspect is to combine both global shape and local features via classifier fusion. Classifier fusion can be used as a way to balance the

[1] http://www.vision.caltech.edu/html-files/archive.html

results from a set of classifiers in order to achieve improvement of recognition performance. This work proposed a meta-classifier approach where it may reduce the bias and error of the base classifiers. Previously, meta-classifier approach is used to text classification problem (Bennett 2006; Morariu et al. 2010; Kim et al. (2003) and fewer researchers performed it on image classification problems such as pedestrian attitude recognition (Borca-Muresan and Nedevschi 2008). By using a meta-classifier approach, both features (global shape and local features) from the dataset and the base classifier outputs are taken into account rather than solely relying on the base classifier outputs alone. The intuitiveness of this approach is to improve the base learner's prediction performance by producing a new set of hypotheses from the base learners' outputs. This set of the hypotheses serves as input to the meta-classifiers. In order to do so, this study needs to exploit the different learning algorithms for improving the performance of object class recognition. As stated in Opelt, et al. (2006a), Hatami and Ebrahimpour (2007), Hegazy and Denzler (2008) and Shotton, Winn et al. (2009), boosting approach has improved the accuracy of object class recognition. Thus, in this study, boosting technique is applied to revolve around the construction and development of the meta-classifier approach with intention to combine global shape and local features.

- Improvement of classification accuracy for Object Class Recognition using the proposed meta-classifier framework specifically demonstrated using

Caltech and Graz02 datasets between different object features, fusion method and different machine learning techniques.

1.6 Thesis Organization

The thesis is organized into seven chapters. This chapter introduced the background of this study and major challenges faced in bridging the semantic gap between low level images features and high level human understanding. Furthermore, this chapter outlines the problem and objectives of this research.

Chapter 2 presents related works of automatic image annotation. This chapter reviews two approaches for automatic images annotation, which are global and region/block annotation. The important issues in building a model for image annotation from the image features and classifier point of views are discussed in detail.

Chapter 3 describes the theoretical background of this study. This includes theoretical foundation for global shape and local features, learning algorithms and training parameters used. The techniques of performance evaluation used also discussed in this chapter.

Chapter 4 evaluates the recognition performance using a different type of features. This chapter explains the dataset, process of features extraction and compares the performance of Object Class Recognition using global shape and local features as mentioned in Chapter 3.

Chapter 5 performs a feature fusion framework for automatically recognizing unlabelled objects using a single classifier technique. The single classifier is tested on different combination features in term of evaluating the recognition accuracy and running cost to build the model for each class.

Chapter 6 overcomes the 'curse of dimensionality' problem in feature fusion. The adopted feature selection methods reduce the computation cost in building the model for each object class while maintaining Object Class Recognition accuracy.

Chapter 7 presents the proposed decision fusion model using different learning algorithms, where two different feature types are combined. It explains the step-by-step architecture of the proposed approach. The combination rules and meta-classifier approach is used as a combined method. The evaluation of the proposed approach is done using the Graz02 database. The comparison with state-of-the-art works is also provided in this chapter.

Chapter 8 summarizes the research contributions and achievements in the field of Object Class Recognition. The limitation of this research and future research work are also suggested in this chapter.

CHAPTER 2 : LITERATURE REVIEW

2.1 Introduction

This chapter presents a review of works pertaining to automatic image annotation and object class recognition. The diagram in Figure 2.1 shows the organization of the research on Object Class Recognition. This taxonomy is arranged based on the perspective of Automatic Image Annotation domain. This diagram represents the classification process of the object class to achieve the consistent and efficient image annotation system. In this chapter, the categories of automatic image annotation with respect to Object Class Recognition are discussed, followed by a section on the various image features and classification methods used in image classification and object class recognition. The final section provides conclusions based on the reviewed literature, which defines the direction taken and ideas proposed in this research.

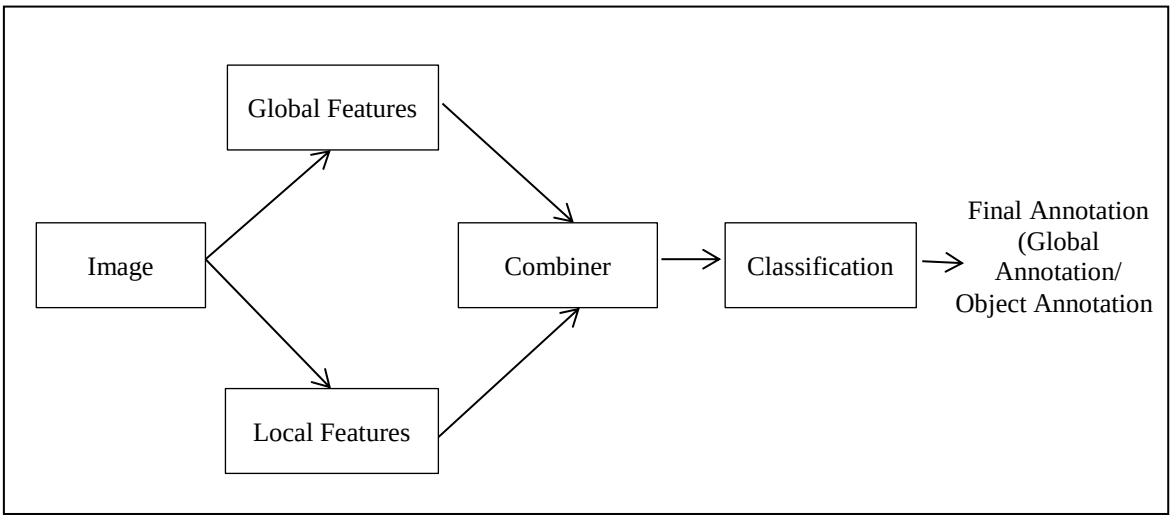

Figure 2.1: Taxonomy of the contents of Chapter 2.

2.2 Automatic Image Annotation

Solving the semantic gap between human and machine understanding is still an active research area in computer vision. Computers are able to efficiently process queries based on extracted low-level features. This however, is not the case for humans since image search and interpretation cannot be intuitively performed based on arrays of numerical values, which is commonly how such features are represented. Therefore, insightful automatic image annotation is needed to enable naïve users to specify conceptual queries through the use of relevant keywords. A variety of approaches have been introduced for automatic image annotation such as co-occurrence model (Mori et al. 1999), machine translation model (Duygulu et al. 2002) and latent space approaches (Nakayama, 2008) and classification approaches.

Image classification is one promosing approach towards automatic image annotation (Zhang et al. 2008; Zhang et al. 2011). The success of image classification primarily depends on two inter-related factors; 1) suitable visual features in representing the variability of image content in terms of poses, sizes, color, illumination and translation and 2) effective learning algorithms to finally perform image classification based on the selected visual features.

Annotation of images can be performed using two approaches (i) global or entire image labeling (ii) region or object labeling. In the context of global labeling, Oliva and Torralba (2001, 2002) explored scene oriented approaches to annotate entire images using basic scene labels such as `street', `buildings' or `highways', obtained through relevant low level Gabor filters. Yavlinsky et al. (2005)

introduced simple global features; the distribution of pixel color in CIE space and Tamura texture feature using non-parametric models.

For region or object labeling, an image is divided into separate regions that are homogenous with respect to chosen properties such as brightness, color, texture, etc. One of the first attempts of region or object labeling was reported in Mori et al. (1999). In this paper, images are tiled into grids of rectangular regions and co-occurrence model of words and low-level features are applied on the regions. Although this approach is computationally less costly, it is unable to identify the concepts accurately. The example of these tiled image regions is shown in Figure 2.2. It can be concluded that, this approach also produces the similar problem faced by global image annotation methods.

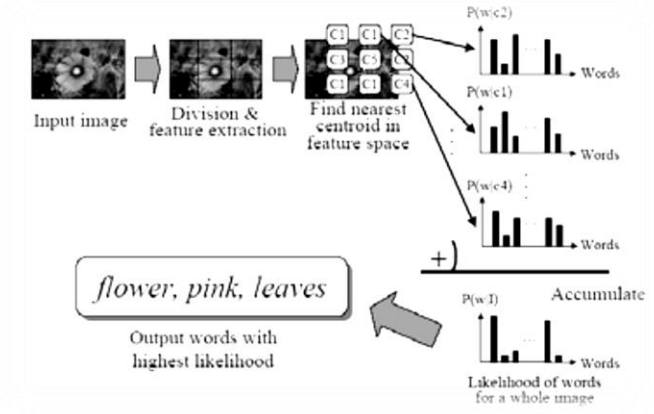

Figure 2.2: Co-occurrence model (Mori et al. (1999).

Several researchers have introduced automatic image annotation methodologies that are based on regions/blobs in order to improve the method proposed by Mori et al. (1999). Duygulu et al. (2002) created a discrete vocabulary of clusters of these blobs across an image collection, and a model inspired by machine translation is applied to translate between the set of blobs comprising an

image and annotation keywords. Jeon et al. (2003) improved the results obtained by Duygulu et al. (2002) through recasting image annotation in cross-lingual information retrieval and applying a cross-media relevance model in order to perform annotation. Lavrenko et al. (2003) adapted the model proposed by Jeon et al. (2003) to build continuous probability density functions to describe the process of generating blob features. Li and James. (2003) used a classification approach to annotate images automatically. In this approach, each annotated word is treated as an independent class and a different image classification model for every word is created.

The advantage of global image annotation does not require segmentation process. The feature extraction process is done directly from the entire image or image partitions without considering the problem of segmentation. For object annotation, it can produce a more precise concept than global image annotation (Kuettel et al. 2012). This is probably because, the classification can be done based on extracted feature in each object. But, it has the disadvantage of requiring object segmentation and automatic segmentation process may affect the performance of object annotation since automatic image segmentation may not be completely reliable. For example, edge-based segmentation approaches detect the changes of image intensity between two dissimilar regions in order to partition an image. The problem with such approaches occurs when the image has noise such as broken edges and/or overlapping regions (Sonka et al. 1999).

Alternatively, another approach, region-based segmentation, segments regions based on pixel homogeneity properties such as color, textures and intensity.

Normalized cuts and watershed are examples of region based segmentation approach. Deng et al. (2001) proposed a region-based segmentation approach based on color and texture homogeneity. Segmenting an object using region based segmentation will create multiple segments in one object. This is because an object will consist of two dissimilar groups of pixels which may create two partitions in an object. Consequently, the current automatic segmentation algorithms cannot produce accurate enough shape representations as expected by the user as shown in Figure 2.3 (Campilho et al. 2006; Chen et al. 2012). For instance, it is difficult to separate an object from its background, if its boundary's color is similar to the background color.

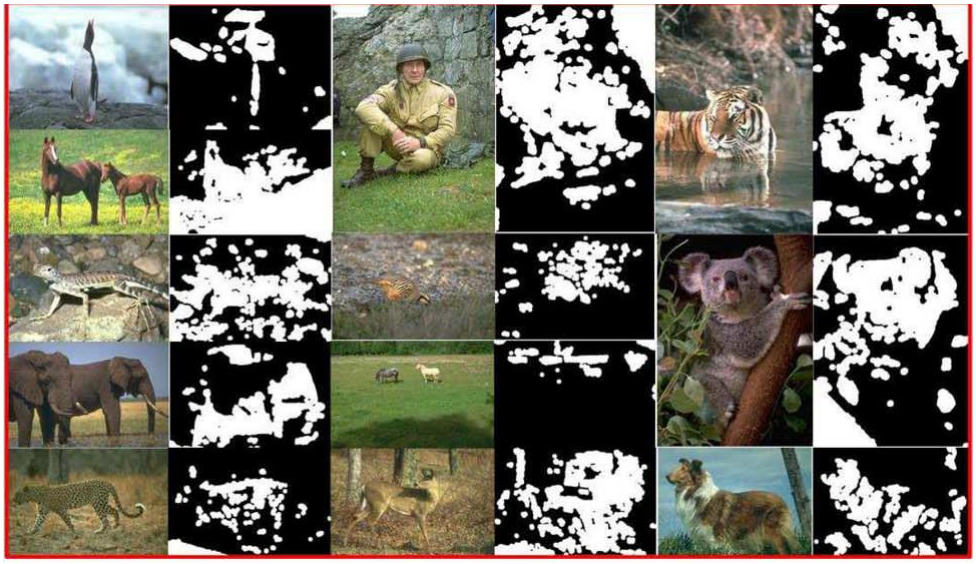

Figure 2.3: Segmentation of objects with complex background (Chen et al. 2012).

As stated, object annotation best reflects the contents of the image by describing its objects. However, as mentioned earlier, this is highly dependent on the ability to segment objects in the image. This area has been receiving much attention lately and several satisfactory approaches are being proposed (Brox, et al. 2011;

Carreira and Sminchisescu 2012). In our research, the focus is on selecting representative features and classifier design assuming image is segmented into objects. Thus, proposed work in this thesis uses either publicly available presegmented datasets or manually performs segmentation where presegmentation is not available.

This study focuses on the selection of features from segmented objects and design of classifier fusion for object class labeling (recognition). To recognize the object class, discriminative features have to be identified to ensure the objects can be grouped into their respective classes. The architecture of the classifier model has to be considered so that the fed features can produce good recognition performance as stated in one of the main objective for this study.

2.3 Feature Extraction

The first important issue in recognizing objects is to identify the most discriminative features. The process of feature extraction is the main task of many applications such as face recognition, hand-written character recognition, video event detection and object class recognition. Therefore, various features have been proposed to improve the performance of object class recognition. Since the object class may appear in various pose, scales, and illumination, the selection of the most discriminative features is very important before embarking into the task of object class recognition. Generally, the features that can be extracted from objects are categorized into two types, global and local features.

2.3.1 Global Features

Global features are used to represent an entire image, and are widely used in many existing CBIR researches. Good examples of global features are color histograms and shape features. The distribution of color in images can be calculated by using color histograms. The advantages of global features are their ability to generalize the whole image (Lisin et al. 2005) and require lower time computational cost (i.e. time) to extract (Glatard et al. 2004).

Shape

As decided in the beginning of this study, shape feature is chosen as one of the important features to recognize the object. Shape is used for retrieval and/or recognition of shape-based objects. Shape describes the geometry information of an object. It is invariant to lighting conditions with variations in object color and texture, and varies smoothly with object pose change (Shotton et al. 2008). In the real world, objects are easily to recognize based on their shape because it is consistent with the human experience and intuition, where heavy reliance is put on the integral shape of an object. Therefore, shapes are frequently used as a vital discriminative feature for object recognition. Some previous shape representations most frequently used in CBIR focused on extraction of whole shapes, such as compactness, area, perimeter and eccentricity. The major advantage of such global shape features is that they can be extracted and matched with minimal computational time (Glatard et al. 2004). For getting more generalized shapes, it depends heavily on the segmentation process, or based on the detection of shape contours.

Shape can be categorized into region-based and contour-based. The former describes the entire shape region using homogenous criterion such as color and texture. The latter category describes the silhouette of the object based on its boundary information. More recently, Ferrari et al. (2008) combined groups of adjacent segments of contour into invariant descriptors and used sliding windows of localized histograms for object detection. Several works used contour fragments to recognize objects based on local contour features in any scale by building a class specific codebook (Opelt et al. 2006b; Yu et al. 2007; Shotton et al. 2008). They used Canny edge detector to find the edges in images. A linked edge is considered as a candidate boundary fragments in a training set. The geometrical relationship between the shape codewords and characteristics of a particular object category are stored (called a grammar of shape codebook). These shape models are used for object detection by providing the location and sizes of objects.

2.3.2 Local Features

Local features refer 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. Local features are computed at multiple points in the object and are consequently more robust to occlusion and clutter. The local features are most widely used to overcome object class recognition accuracy (Mikolajczyk et al. 2005). This is because of that feature is robust to the translation, rotation, views, scales and can recognize partially occluded object (Lowe 2004). Hegerath et al. (2006) extracted local features from image patches of different sizes. For instance, they considered the patch sizes (in pixels) 7 x 7, 11 x 11, 21 x 21 and 31 x 31. The reason to do that extraction approach is to represent object parts of different sizes

and to handle the scale changes. Note that segmentation is not performed before the extraction of the local features. Examples of local features are Scale-Invariant Feature transform (SIFT) and Gradient Location and Orientation Histogram (GLOH).

Scale-Invariant Feature transform (SIFT)

SIFT is a very good local feature for objects with different view, scale, image blur, light change and translation (Mikolajczyk and Schmid 2005). It was introduced by Lowe (1999) to solve the problem of 2D-object recognition. The difference-of-Gaussian is applied to identify the interest points of an object. 128 features are extracted around multiple interest points of object patches. This produces multi-dimensional features for a single object. To produce a single feature vector, (Csurka et al. 2004) proposed the Bag of Keypoints (BoK) approach. Leibe et. al (2006) used Harris-Laplace and Hessian-Laplace detectors to produce the SIFT features for 'pedestrians', 'cars', 'motorbikes', 'faces' and 'cows' classes (Leibe et al. 2006). Opelt et. al (2006a) used SIFT with different local features such as sub-sampled gray values, basic intensity moments and moment invariants as an input to the classifier to recognize object class. The authors conclude that classification performance using a combination of many local features produces higher accuracy result compared to the SIFT feature alone.

2.3.3 Combination of Global and Local Features

Previously, most prior researches in object class recognition focused on one type of feature for discriminating between objects of different classes. The

recognition of object class creates a problem if the classes cannot be discriminated by using only one feature (Csurka et al. 2004; Opelt et al. 2006a; Hegazy and Denzler 2008). For example, 'horse' and 'cow' classes may have similar local information. Thus, these objects cannot be properly distinguished by using local information alone. Moreover, local features have several limitations. Firstly, if the object does not have enough local information such as for 'bikes' and 'glass' classes, the SIFT features cannot provide a discriminative feature for those objects (Mansur and Yoshinori 2007). Secondly, the local features do not consider the shape of the objects. Due to these limitations, several researches in object class recognition combined local with global features to give stronger discriminative power for categorizing the objects into their respective classes.

Opelt et al. (2006b, 2006c) combined features from image patches and edge boundaries for recognizing object categories. In this study, Boundary Fragment (BF) models were used. BF consists of a set of curve fragments, which represents the edge of objects and its centroid using a codebook. Zhang et al. (2005) introduced spatial features to combine with local feature, PCA-SIFT and global shape context. Shape context features were computed based on points detected in the edge image. The points are represented from an internal and external contour of an image. The similarity of shape is obtained by calculating the shape histogram distance between two shapes. In this work, the limitation of shape context is sensitive to object occlusion and hard to extract the shape's contour for every complex background (Zhang, et al. 2005).

The other global shape features such as area, perimeter, compactness, and local binary pattern as a texture features are used together with SIFT descriptors proposed by Lisin et al. (2005). These features are used to recognize images of multi-cellular organisms in marine science. Oliveira et al. (2007) applied a Haar-like feature as a global shape feature and Histogram of Oriented Gradient (HOG) features as a local feature to recognize cars and pedestrians in outdoor environments. Another global feature called 'gist of image' introduced by Murphy et al. (2006), captures coarse texture and spatial layout of an image. The authors combined this feature with image fragment from the filter image outputs.

The combination of global and local features as discussed earlier improves the performance of object class recognition. The selection of suitable features not only gives influence to the performance of the recognition engine, the combination also has to be focused. The techniques for combining different features are explained in subsection 2.4.3. Table 2.1 presents the summary of previous studies on object class recognition that used more than one feature to categorize the variation of objects belonging to the similar category in different scales, poses and appearance.

Based on the literature that has been reviewed, more recent works tend to use a variety of features in addition to local features to classify object class (Jeong et al. 2009; Mansur and Yoshinori 2007; Oliveira et al. 2007; Zhang et al. 2005) . This is because the challenges in recognizing objects are considered as a very difficult task especially when it involves different view, location, position, scaling and etc. Most authors used incomplete contour or boundary of the shape of objects by building a class specific codebook (Yuan and Hui 2008; Shotton et al. 2008; Ferrari et al. 2008;

Yu et al. 2007). These approaches do not give very detailed descriptions of the shape of the object.

Table 2.1: The combination of features used in object class recognition researches.

Previous works	Global	Local	Dataset
(Fergus et al. 2003)	Shape	PCA-SIFT	Caltech
(Opelt 2006b ; Opelt et al. 2006c)	boundary fragments	SIFT	UIUC cars, Caltech
(Zhang et al. 2005)	Shape context	PCA-SIFT, spatial features	Caltech and GRAZ
(Mansur and Yoshinori 2007)	Gabor filter	SIFT	Caltech
(Opelt, et al. 2006a)		Subsampled gray values, basic intensity moment, moment invariant, SIT, intensity distribution	Caltech and Graz
(Lisin et al. 2005)	Shape: area, perimeter, compactness Texture: local binary patterns, shape index	SIFT	Plankton database
(Oliveira et al. 2007)	Haar-like feature	Histogram of Oriented Gradient	INRIA, Caltech
(Jiang et al. 2007)	Color moment and wavelet texture	SIFT	PASCAL 2005 & TRECVID-2006
(Meng et al. 2005)		Moment invariant, SIFT	Caltech (motorbikes, airplanes, faces), Graz02 (bikes, persons)
(Marszalek and Schmid 2007)	Hue	SIFT	PASCAL 2006
(Hegazy and Denzler 2008)		GLOH, color (opponent angle)	Caltech, Graz02
(Jeong et al. 2009)	Color histogram, edge histogram, radon transform	SIFT	Not mentioned
(Oliveira and Nunes 2008)		Histogram of Oriented Gradient, Local Receptive Field	Caltech, Graz, INRIA
(Antenreiter et al. 2009)		Texture statistics of segments, subsampled grayvalues, basic moment, moment invariants, SIFT, PCA-SIFT,	PASCAL 2006 & 2007
(Murphy et al. 2006)	Gist	Image fragment	MIT-CSAIL, UIUC

2.4 Building a classifier

To build an optimal image annotation system, it is necessary to be able to properly distinguish objects from different classes (i.e. object class recognition). This means that, an object class should be generalized to such a degree that two objects of the same class are labeled similarly. The most prevalent challenge is to find the most discriminative features and to design classifier models that are robust to the changes of appearance within a class and capability to discriminate between objects from different classes. Compared to the recognition of specific objects from images (e.g. different images of object, for example a 'cars'), object class recognition involves classification of objects belonging to a class such as 'cars', 'motorbikes', or 'human face' with different instances of the object, (e.g. images of different 'cars'). Figure 2.4 shows examples of specific object and object class recognition. In the following subsection, the issues of object class recognition approaches are summarized and the significance of these approaches to this research are discussed.

2.4.1 Approach to Object Class Recognition

Several approaches for object class recognition exist in the literature. Some methods differ in the types of features and approaches. Contour/shape-based models (Leibe et al. 2005; Shotton et al. 2005; Ferrari et al. 2008; Shotton et al. 2008), constellation models (Weber 2000; Scalzo and Piater 2007; Fergus et al. 2003), and keypoint-based or appearance-based models have proven to be successful to categorize object classes.