

## ABSTRACT

As the number of images to be stored in personal collections, public data sets and the internet is becoming huge and ever growing, it becomes crucial to develop computationally efficient methods for organizing and searching images. Therefore, the ability to classify images into semantic categories and objects is essential in order to manage and organize the collection of images on a database. The problem is challenging because the appearance of object instances varies substantially owing to changes in pose, imaging and lighting conditions, occlusions and within-class shape variations. Ideally, the representation should be flexible enough to cover a wide range of visually different classes, each with large within-category variations, while retaining good discriminative power between the classes.

In this research, the problem of improving the classification performance of various classes of images like objects, semantic scenes, fine grained, textile designs and fabrics are explored from a supervised learning perspective. Hence this thesis is a result of the study and comparative evaluation of different methodologies for automatic image classification in terms of its degree of recognition, using performance measures like accuracy rate, Macro Precision, Micro Precision/Mean Average Precision (MAP) and F1 measure.

The most popular method for image classification that uses features is Bag of Words (BoW). This method works, by selecting or sampling patches from the image and characterizing them by vectors of local visual descriptors. The descriptors are quantized to form a visual word dictionary called codebook with the help of different clustering algorithms. The occurrences of each label called the “visual word” are then counted to build a global

histogram called “Bag of Words”, summarizing the image contents. The histogram is fed to a classifier to estimate the image’s category label. Visual words are not intrinsic entities and different quantization methods or clustering methods can lead to very different performances. The parameters that affect performance of BoW are dictionary generation or Codebook formation method, dictionary size, histogram weighting, normalization, and distance function. Essential tasks that influence the performance of Codebook formation method are, parameters like choice of local features, sampling strategies, quantization method and number of visual words.

The performance of Bag of visual words approach need to be investigated, to understand and study the effect of variation of parameters like local features, quantization methods like different hard clustering techniques and variation of number of visual words. In the initial phase of this thesis work, the performance of Bag of Words approach is studied by varying the feature descriptor and dictionary generation method.

Traditional Bag of Words approach uses hard clustering for codebook generation. A given feature may be nearly the same distance from two cluster centers. For a typical hard clustering method, only the slightly nearer neighbor is selected to represent that feature. So, the ambiguous features are not well represented by the visual vocabulary. To address this problem, soft clustering model based Bag of Visual words for image classification is implemented with features being clustered using different soft clustering algorithms.

Fuzzy clustering algorithms are based on minimization of an objective function. When determining the distance between two features, as required by clustering and term assignment, common choices are the Manhattan ( $L_1$ ), Euclidean ( $L_2$ ), or Mahalanobis distances. The objective

function of Fuzzy C Means (FCM) proposed by Bezdek is based on the Euclidean norm or  $L_2$  norm. When  $R_1$ -norm is applied to K-means clustering, it is seen that  $L_1$ -norm K-means leads to poor results while  $R_1$ -K-means outperforms standard K-means. By applying the same concept, objective function of FCM and GK clustering algorithm is redefined in terms  $R_1$ -norm. The consequences of fuzzy clustering for codebook generation in Bag of Words approach for object recognition is investigated by varying the power of the membership function.

To increase the performance of traditional Bag of Words approach further, Multiple Dictionaries BoW (MDBoW) method that uses more visual words from different independent dictionaries instead of adding more words to the same dictionary is implemented using Soft clustering method. The performance of Fuzzy based Multiple Dictionary Bag of Words is analysed with existing Fuzzy C means, modified FCM with  $R_1$ -norm in the objective function, and they are compared with the baseline method.

The amount of information managed by computer is increasing with the development of Internet and multimedia technologies that grow exponentially. Different algorithms have been developed to reduce the amount of memory required for the digital documents of images multimedia transmission. Spatial domain image retrieval of compressed images using Bag of Words model is implemented since the local features in Bag of Words model can be extracted from images only in the spatial domain. By varying the compression ratio, the performance of the model is studied by compressing and decompressing the images using different lossy and lossless compression schemes.

Indoor scene recognition is a challenging open problem in high-level vision. For unmanned Miniature Aerial Vehicle (MAV) to fly autonomously in indoor environment, it has to identify the type of the

environment in which it is navigating and then follow different rules to navigate. They are generally scenes with the most similar object distributions and categorised as semantic scenes. Not much research work has been done on indoor scenes that are semantic. The main difficulty is that while some indoor scenes (e.g. corridors) can be well characterized by global spatial properties, the objects they contain better characterize others (e.g. bookstores). Indoor scenes have more synthetic objects than outdoor scenes. Synthetic objects tend to have more contours and edges. There are various edge types like material edges, shadow or shading edges, specular edges, and inter reflection edges. Therefore, an attempt is made to classify images of semantic indoor environments commonly found in home and office buildings for MAV to navigate based on Gist features extracted from various edge maps of images.

Fine-grained classification demands an algorithm to discriminate among highly similar object classes that are often differentiated by only subtle differences. Thus, an automated visual system for this task could be valuable in many applications. Flowers, birds and butterflies have more contours and edges. Since the Gist features extracted from the original image and edge map of images represent the spatial textural features, shape and color features, an attempt is made to classify fine-grained images like flowers, birds and butterflies based on this method. The performance of the algorithm for synergy of different features is analysed.

The peculiar textural features differentiate fabrics from each other. Therefore, an automatic image recognition or classification system for fabric design, stylized fabric patterns, fabric and woven fabric for the process of automation of textile and clothing manufacturing units is also implemented using the same classification scheme.