Research Journal of Applied Sciences, Engineering and Technology 10(10): 1089-1101, 2015
DOI: 10.19026/rjaset.10.1878
ISSN: 2040-7459; e-ISSN: 2040-7467

Submitted: January 13, 2015 Accepted: March 7, 2015 Published: August 05, 2015

Research Article

Descriptor Trends in Texture Classification for Material Recognition

[1]Hayder Ayad, [2]Mohammed Hasan Abdulameer, [3]Loay E. George and [1]Nidaa F. Hassan
[1]Department of Computer Science, University of Technology
[2]Department of computer science, Faculty of education for women, University of kufa,
[3]Department Computer Science, College of Science, Al-Jaderia Campus, University of Baghdad,
Baghdad, Iraq

Abstract: Recent rapid growth in the demand for technology and image investigation in many applications, such as image retrieval systems and Visual Object Categorization (VOC), effective management of these applications has become crucial. Computer vision and its various applications are a primary focus of research. Content-based image retrieval is considered an extremely challenging issue and has remained an open research area. Obviously, the main challenge associated with this kind of research is the gap between the low-level features and the richness of the semantic concept of the human mind. This problem is called the semantic gap. Several methods have been proposed to increase the performance of the system and reduce the semantic gap. These proposed techniques make use of either global or local features or a combination of both global and local features on one side and the visual content and keyword-based retrieval on the other side. However, the aim of this study is to provide a constructive critique of the algorithms used in extracting the low-level features, either globally or locally or as a combination of both. In addition, it identifies the factors that can affect the low-level features that lead to the semantic gap. As well as, proposed a new framework to improve the Gabor filter and the edge histogram limitations. Finally, recommendations are made for the choice of the descriptors used to describe the low-level features, both locally and globally, depending on the area of limitations or drawbacks of the previous state-of-the-art research.

Keywords: Combination feature, content-based image retrieval, global and local descriptors, SIFT descriptor, similarity measure

INTRODUCTION

An image retrieval system makes use of three methods of retrieving an image from the whole database: the use of keywords, features or concepts (Alemu *et al*., 2009). In fact, Content-Based Image Retrieval (CBIR) has been utilized since the 1980s. The primary concept behind the creation of this kind of system is to retrieve relevant images and overcome the limitations of Text-Based Image Retrieval (TBIR). In fact, TBIR uses an annotation to describe an image and performs the retrieval based on that annotation. The text-based image retrieval system has two apparent disadvantages. First, it requires humans to create the annotation to the image; this is inefficient because it takes time for the annotation to be made. Second, the description of the image may not be accurate because it may be influenced by the user's own subjectivity. Therefore, TBIR is not an efficient method of retrieving images from a large collection in a dataset (Liua *et al*., 2007). Using the CBIR system, the other researchers suggested using the visual content of the image and storing the image corresponding to this feature (Abdullah and Wiering, 2007). A great deal of research has focused on how to define and use low-level features in the correct way to achieve better accuracy. This attempt was made in order to reduce the semantic gap between the visual features and the richness of the human semantic system (Liua *et al*., 2007).

Much of the research performed in this area has focused on using the low-level features, namely color, texture and shape (Hiremath and Pujari, 2007). In fact, the color feature is the most common feature used in CBIR systems because it is the easiest to extract from the image. Many researchers have employed the color space of the image because it is much closer to human perception (Liua *et al*., 2007). The texture feature is also commonly used in image retrieval, especially on images with significant texture features (Tamura *et al*., 1987; Bin Adam *et al*., 2011). The shape feature shown in several CBIR systems is robust and provides a real description of the image in different states (Yap *et al*., 2003; Arif *et al*., 2009). However, a problem still exists with these kinds of features that are related to the descriptor algorithm used in extracting the features (Wu and Wu, 2009). In general, global and local features are important parts of image retrieval systems. In a global feature, the primitive feature is taken from the entire

Corresponding Author: Hayder Ayad, Department of Computer Science, University of Technology, Baghdad, Iraq

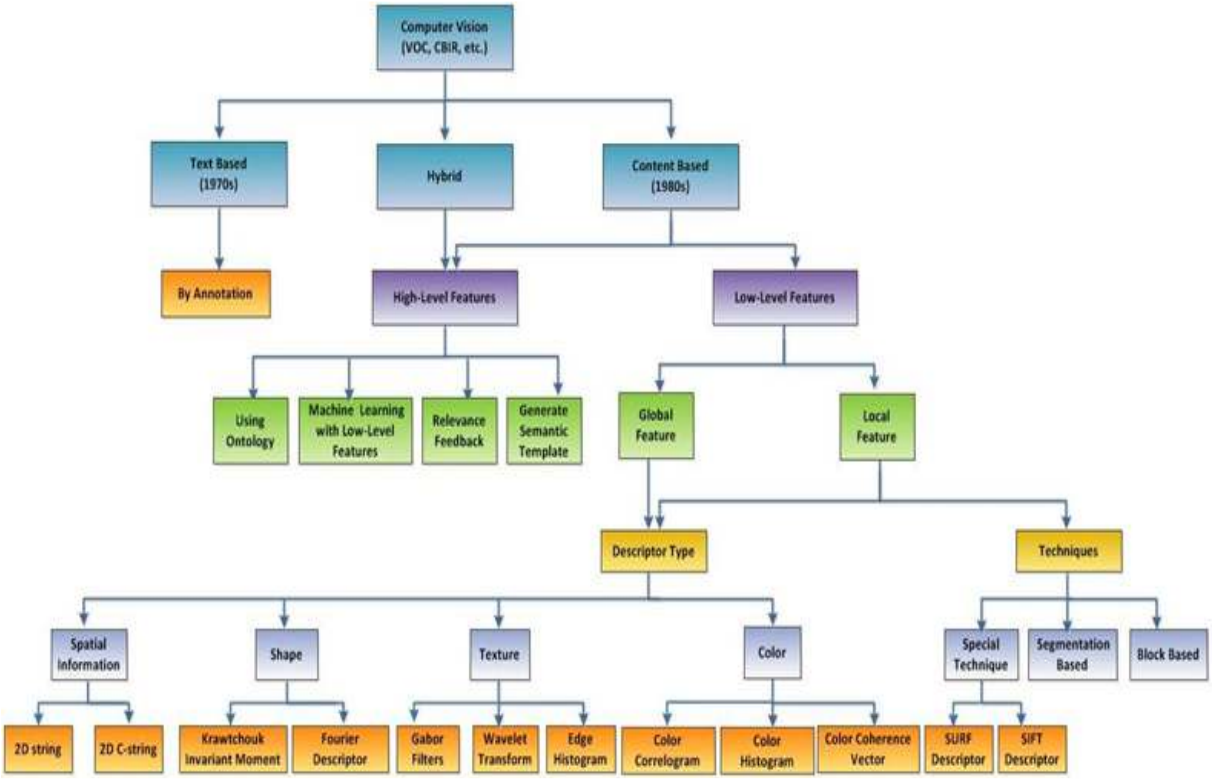


Fig. 1: Illustration of the techniques used in computer vision applications (CBIR, VOC)

image. In contrast, in a local feature, the feature is taken from a specific area; this appears to be more effective than the global feature (Chih-Fong and Wei-Chao, 2009). However, in terms of the similarity measure, there are several ways of obtaining a similar image, either using a distance measure, such as the Euclidean distance, or with a machine learning algorithm, such as Support Vector Machine (SVM), Neural Network and the Bayesian classifier (Liua *et al*., 2007; Abdullah *et al*., 2010). Figure 1 shows how the image retrieval system and other computer vision applications grow in order to improve system accuracy.

This study focuses on CBIR system and it provides a constructive criticism to feature extraction methods of the low level features. As well as, shows the effectiveness of using combination features from one side and the using local and global features from other side. In addition, a new framework based on unsupervised clustering method for improving the Gabor filters has been proposed.

METHODOLOGY

Feature extraction: To perform CBIR, the features must be extracted from the image. In fact, feature extraction is divided into two parts: the low-level feature and the high-level feature. The low-level feature involves the extraction to be made either globally or locally. On the other hand, the high-level feature refers to the semantic concept (Chih-Fong and Wei-Chao, 2009; Abdullah *et al*., 2010).

Low-level feature: In general, there are two types of features in image content: the visual and semantic contents. The visual content consists of either general visual content, such as color, texture, shape and spatial relationship, or domain-specific visual content, such as human faces (Long *et al*., 2003). Researchers have found that there is no direct link between the low-level features and the high-level features; therefore, there is a gap between them. Several researchers have addressed this problem and tried to solve it using either global, local or a combination between the global and local features (Deb, 2008; Jaswal and Kaul, 2009). In the sections below, the four types of low-level features will briefly be discussed.

Color: The color feature is one of the most important low-level features because extracting this type of feature is considered easy. Furthermore, the color feature is much closer to human perception and is commonly used in image retrieval systems. In fact, this type of feature can be used to identify the location and number of objects in an image (Long *et al*., 2003; Kurniawati *et al*., 2009). Several researchers have introduced algorithms to describe the distribution of color, such as color histogram, color correlogram, Color Coherence Vector (CCV) and color moment. In fact,

the color histogram is usually used in image retrieval systems. However, there is a drawback in the descriptor because it does not take the spatial information in the image into account. Therefore, color correlogram and color coherence vector were introduced to overcome this drawback by considering the spatial information (Long *et al*., 2003; Lukac and Plataniotis, 2007; Abdullah *et al*., 2010). The literature review clearly shows that CCV works better than color histogram, especially with images with significant texture features or uniform colors. Additionally, both of them work better in the HSV color space than the CIE L*a*b* color space (Long *et al*., 2003; Kurniawati *et al*., 2009). Color moment is also widely used in image retrieval systems; it gives a better performance with CIE L*a*b* color space and CIE L*u*v* than the other color spaces because using the third stage of color moment produces features that are sensitive to noise. Therefore, this kind of descriptor is used as the first stage to narrow down the search space (Xiaoyin, 2010). Lastly, the SIFT color descriptor with diverse types is studied in Van de Sande *et al*. (2010). It works better than color moment and color histogram under certain conditions, such as changes in illumination and viewpoint.

However, other researchers have used the color space with the descriptors above because it is much closer to human perception. Obviously, several color spaces are provided in different applications, namely, RGB, YUV, YCrCb, CIE L*u*v*, HSX (HSV and HSI), CIE L*a*b* and the Hue-Min-Max-Difference (HMMD) (Liua *et al*., 2007). In addition, MPEG-7 introduced some basic techniques for extracting color features: dominant color, color structure, scalable color and color layout as color features (Manjunath *et al*., 2001).

Texture: Texture is one of the most important low-level features because it provides meaningful information that supports the classification part and retrieval performance (Bataineh *et al*., 2011). This feature is not as commonly used as the color feature, but it is useful when used to describe real-world images. In addition, the use of the texture feature can help achieve the high-level feature necessary to increase the performance of Content-Based Image Retrieval (CBIR) (Liua *et al*., 2007). The methods used to extract the texture feature are divided into two parts. First, the structural method tries to extract and characterize the texture features in the image by determining the structural primitives and their placement rules. Second, there are statistical methods, such as co-occurrence matrices, Shift-invariant Principal Component Analysis (SPCA), Fourier power spectra, the Tamura feature, the Markov random field, Wold decomposition and multi-resolution filtering, such as the Gabor filters, Wavelet transform and the fractal model (Wang *et al*., 2001; Lui *et al*., 2007; Bin Adam *et al*., 2011).

Several algorithms for the extraction of texture features have been proposed for the image retrieval system. The Tamura feature consists of six features for extracting the texture features, namely, coarseness, contrast, directionality, line likeness, regularity and roughness. The first three features are more effective than the remaining three features. In fact, adding the latter three features to the extraction makes no change in system performance. The main drawback of this descriptor can be observed when applying it with multiple resolutions to account for scale (Long *et al*., 2003). On the other hand, Wold texture provides a good description when it is applied to Brodatz textures. However, the drawback of this descriptor is that it becomes less effective when transformation, orientation and distortion affects occur in the images. It is also less effective when applied to natural images, which are not as structured and homogeneous (Leow and Lai, 2000). In addition, MPEG-7 introduced edge histogram for extracting local texture feature. This descriptor capturing the spatial distortion of the edge and is invariant to most image effects, such as transformation. Among the different texture descriptors, there are two descriptors commonly used in retrieval systems, computer vision and object matching, namely, the Gabor filter and wavelet transform. These two descriptors show that they are more robust with different types of image effects; both of them have been proposed for rectangle images (Wang *et al*., 2001; Chih-Fong and Wei-Chao, 2009). Figure 2 provides an example of the Gabor filters with two filters.

Shape: The shape feature is not as widely used as the color and texture features. This is related to the difficulty of finding a rigorous segmentation algorithm (Sarfraz, 2006; Kurniawati *et al*., 2009). However, this feature has a good discriminative capacity to distinguish the images and gives a real description of the object or region when it is easily available. When the shape feature is used in an application, it must be invariant to most image effects, such as transformation (translation, rotation, scaling), noise and other effects (Mohan *et al*., 2008).

From the state-of-the-art techniques, we found that the shape descriptor can be divided into two parts: boundary-based and region-based. In the boundary-based part, the exterior boundary features of the shape are used to extract the features, while in the region-based part, the inner shape region is used to extract the features (Pabboju and Reddy, 2009). However, several descriptors are introduced to serve as boundary or region feature descriptors; for example, the MPEG-7 descriptors introduced three shape descriptors. The first is for 3D objects, the second is for region and the last one is for boundary (Liua *et al*., 2007).

However, several descriptors have been presented for extracting both types of shape features, such as the

(a) (b) (c) (d) (e)

Fig. 2: An example of one category of a caltech 101 dataset; (a) Source image, gabor wavelet filter at, (b) Scale 0 and orientation of 315°, (c) Scale 1 and orientation of 45°, image primitive using after using gabor wavelet filter at, (d) Scale 0 and orientation 315°, (e) Scale 1 and orientation of 45°

Fig. 3: Edge map for the dinosaur category of the corel dataset using canny edge detection, the parameters used are standard deviation (σ = 1.4, low threshold = 0.5, high threshold = 2.5, sigma = 1)

Fourier descriptor, moment invariant family, rectilinear shapes, chain code, polygonal approximation, finite element models, turning angles, grey level, region area, compactness, signature and others (Long *et al*., 2003; Liua *et al*., 2007; Mohan *et al*., 2008; Wu and Wu, 2009; Pabboju and Reddy, 2009; Jaswal and Kaul, 2009). Most of these descriptors are used after segmenting the image into homogeneous regions; these descriptors serve the function of either global or local feature descriptors. Furthermore, several researchers have presented algorithms for shape detection, such as the multi-resolution algorithm, distance potential and pressure forces. All of these algorithms have problems, namely the initial boundary of the object and the converge connectivity. Later, Gradient Vector Flow field (GFV) was introduced to overcome the shortcomings of the previous method (Xu and Prince, 1998). The edge detection technique provides meaningful information that can be used in the segmentation algorithm, shape detection method and object localization. Examples of this technique are the Roberts operator, the Sobel operator, the Laplacian of Gaussian and Canny (1986) edge detection (Nadernejad *et al*., 2008; Abdullah *et al*., 2007a). Figure 3 describes shape edge detection with canny edge detection technique.

On the other hand, the signature and compactness descriptors shown are usually used to describe the object. The signature is considered a one-dimensional function that serves as a boundary descriptor and is invariant to translation effects (Mingqiang *et al*., 2008). Apart from that, the scaling invariant can be achieved by performing a scale of the signature function (Mohan *et al*., 2008). Compactness is widely used to describe the shape features because it is easy to compute and has invariant properties against transformation effects. In contrast, the drawbacks of this descriptor are that it describes only approximate shapes and is unstable for curved shapes. Chain coding (Freeman code) has good characteristics: it is compact, invariant to translation effects and indexable. The drawback of this descriptor is that it is not independent on rotation or scaling and is inefficient with complex shapes. In addition, turning angles are used to extract the shape features and it has invariant properties against transformation effects. The problem with this kind of descriptor is the reference point; if it is rotated on the counter shape, the feature value will change (Mingqiang *et al*., 2008).

The literature has specified that the Fourier descriptor and the Moment family are commonly used as shape feature descriptors (Wu and Wu, 2009; Jaswal and Kaul, 2009). The former produces a feature that has invariant properties against transformation effects and can represent the image uniquely. In fact, this descriptor is used in several applications, such as object recognition and classification as well as remote sensing. In contrast, there are two main drawbacks to this descriptor. First is the occlusion effect, which, when it occurs to the shape of the object, will lead to the change of the feature value from one image to another. Second, the computational complexity is also high (Sarfraz, 2006).

The moment family is also widely used in pattern recognition to describe the geometric features of various objects (Abdullah *et al*., 2007b). In fact, the first type of moment family presented by Hu (1962) was called the Hu Moment. At present, there are seven invariant moments of 3rd order; they are invariant only to rotation. After some time, the moment invariant came into common use in object recognition and image retrieval systems. In contrast, the drawbacks of this moment are dependence, incompleteness and sensitivity to noise (Wu and Wu, 2009). In addition, the central moment was introduced as an invariant to the translation effect and can yield an invariant to the scaling effect through normalization of the central moment. After that, complex moment was introduced as an invariant to the rotation effect (Mohan *et al*., 2008). Finally, several kinds of orthogonal invariant moments were introduced and put into wide use as

shape descriptors. These kinds of moments are robust to noise and close to the zero redundancy measure in any feature set. There are a variety of types of this moment, such as the Legendre, Zernike, Fourier-Mellin, Czebyshev, Tchebichef, Krawtchouk and Hahn moments (Yap *et al*., 2003; Mohan *et al*., 2008; Mingqiang *et al*., 2008; Arif *et al*., 2009; Wu and Wu, 2009; Zolkifli *et al*., 2011). The Krawtchouk moment invariant, Tchebichef moment, Legendre moment and Zernike moment have been more widely used than the others. The main point of creating the Zernike moment is to surmount the failing in the geometric moment; this descriptor is invariant to rotation effects and robust to noise and expressiveness. However, the Zernike moment has three drawbacks. First, it is limited when applied to different datasets; second, small, local disturbances can affect the moment values; and finally, it is not invariant to all transformation effects (Mingqiang *et al*., 2008). The Legendre moment is invariant to only the translation and rotation effects (Arif *et al*., 2009). Lastly, the Krawtchouk moment invariant is invariant to most image effects, such as the transformation effect and noise. The motivation to create this kind of orthogonal moment is to overcome the limitations of other moments. This means a discretization error may exist when computing continuous orthogonal moments, such as the Zernike moment and the Legendre moment. Additionally, this descriptor does not need to use normalization because it consists of a set of discrete orthogonal moments. These characteristics make this descriptor widely useful in pattern recognition and image retrieval systems (Yap *et al*., 2003).

Spatial information: Spatial location is an important feature in an image retrieval system. It identifies the right position of the object in the image and can achieve high-level semantic concepts (Liua *et al*., 2007). For example, if there were two images, one of a sea scene and the other of a sky scene, both would produce the same histogram feature. This problem can be solved using spatial location (Long *et al*., 2003). In fact, the spatial location can clarify the positions as top, bottom or upper depending on the object or the region. In short, the relative spatial relationship between the objects shown is more important than the other relationships, such as the relative spatial location, in driving the semantic meaning (Alemu *et al*., 2009).

The literature presents several ways of obtaining spatial information. It appears that the 2D string and variety types are the most common methods used to represent spatial information and improve the representation of the relationships between the objects (Shi-Kuo *et al*., 1987). In the 2D G-string, the spatial relationship is separated into two groups. The first serves as a local spatial relationship and the second one serves as a global relationship. The 2D C-string was introduced to reduce the number of cutting objects and the 2D B-string was used to represent the objects using two symbols. In conclusion, in an image retrieval system, especially in content-based image retrieval, the use of this kind of feature is still unattainable because there is no precise segmentation algorithm that can provide a real object or region, especially for real-world images (Long *et al*., 2003; Liua *et al*., 2007).

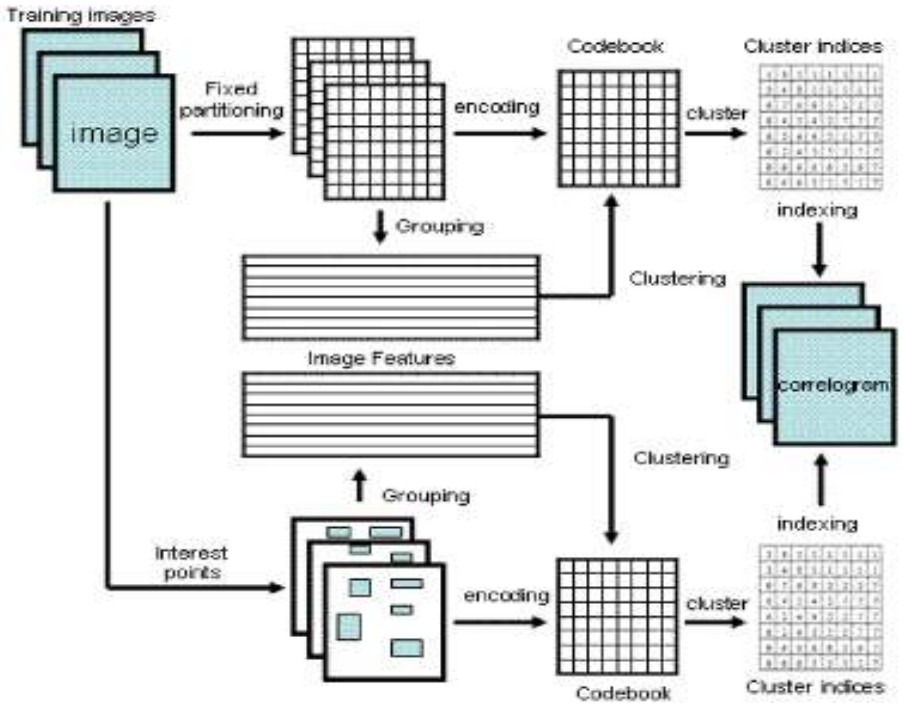


Fig. 4: The structure of the clustering correlogram system (Abdullah *et al*., 2010)

High-level feature: At present, high-level features are considered the main challenge in CBIR. Achieving this kind of feature can reduce the semantic gap and researchers have tried to use different methods to obtain such features. To achieve a high-level feature, either a keyword or a combination of the visual content and the keyword (Alemu *et al*., 2009) can be used. However, several techniques based on high-level features, such as ontology and the machine learning algorithm, provide relevant feedback information to develop a semantic learning template except (Liua *et al*., 2007).

From the systems above, we have clarified that object ontology supplies meaningful information to define the high-level concept. The machine learning algorithms with both supervised and unsupervised algorithms are proposed to drive the semantic meaning, such as the Support Vector Machine (SVM), neural network, Bayesian classifier and K-means algorithm (Liua *et al*., 2007; Xiaoyin, 2010; Abdullah *et al*., 2009). The SVM shown below is precise and is frequently used in object categorization and image retrieval systems (Abdullah *et al*., 2010). By comparing relevant feedback among the other offline systems, it provides better accuracy when producing a high-level feature (Liua *et al*., 2007). Abdullah *et al*. (2010) presented an intelligent system that used a combination of features with a clustering correlogram to produce the high-level feature. The experiment shows that this system outperforms the other state-of-the-art models. Figure 4 describes the system's structure.

IMAGE FEATURE REPRESENTATION

The image features are represented in three types: global, local and a combination of global and local features. A brief description of the three types is provided below.

Global feature: At present, due to the increase in new technology and the investigation of images in many applications, efficient mechanisms to manage these images are required. In fact, searching for the target image is considered the main issue in many studies (Pabboju and Reddy, 2009). Therefore, the global feature has become an important feature over the last decade. The global feature extracts the primitive feature from the entire image. Each pixel in the image is taken into account (Chih-Fong and Wei-Chao, 2009; Bataineh *et al*., 2011). The advantages of this feature are its compact representation and faster computational rate. However, the disadvantage of this feature is that geometric distortion could change the values of the feature. Thus, the use of this feature alone is not very efficient and could lead to a semantic gap (Pujari *et al*., 2010).

Many researchers have proposed algorithms for extracting the global features to improve the performance level of the retrieval system. The moment invariant was proposed for the extraction of the shape feature (Hiremath and Pujari, 2007). However, this descriptor is not invariant to transformation effects and is thus considered a drawback. The use of color and texture features with a weight assignment operator algorithm could only be performed on an image with a number of texture features (Lui *et al*., 2007; Bataineh *et al*., 2011). Different kinds of global descriptors, such as the color average as a color feature, are used; however, this will be insufficient if the segmentation provides inhomogeneous color regions (Pabboju and Reddy, 2009). To summarize, the use of the global feature alone in retrieving an image is inefficient. Therefore, most researchers have suggested and recommended combining the global feature with other descriptors, such as the local descriptor, to overcome the limitation of using the global feature individually (Hiremath and Pujari, 2007; Lui *et al*., 2007; Pabboju and Reddy, 2009; Chih-Fong and Wei-Chao, 2009; Wu and Wu, 2009; Alemu *et al*., 2009; Pujari *et al*., 2010).

Local feature: The local feature is one of the most commonly used features in many areas of image processing, such as image matching and a variety of computer vision techniques. This is due to the characteristics of this feature, such as its invariance to transformation effects, illumination change, viewpoint and others (Yang and Wang, 2008). Based on our review of the literature, two methods of obtaining the local feature have been specified with either segmentation or special techniques, such as the Scale-Invariant Feature descriptor (SIFT). A brief description of both methods is provided in the section below.

Based segmentation method: The use of the segmentation method in detecting an object or interest region correctly and in the manner of human perception remains an open area of research, especially with real-world images (Yang and Wang, 2008; Wu and Wu, 2009). In fact, several segmentation algorithms have been introduced in the last decade for the purpose of image retrieval and object recognition, such as curve evaluation, graph partition and energy diffusion (Liua *et al*., 2007). These algorithms present problems when they are applied to images with mixed color and texture features. Additionally, the texture feature is considered a difficult feature used to segment the image.

Later, several techniques were introduced to overcome the limitation of these algorithms, such as the JSEG segmentation, Blobworld segmentation and K-Means with Connectivity Constraint (KMCC). The JSEG segmentation method was introduced to segment the images with a mixture of both texture and color features (Deng *et al*., 1999). Figure 5 illustrates the JSEG segmentation methods.

(a) (b)

Fig. 5: JSEG segmentation methods; (a) The original image, (b) Image segment with 27 regions (Serra, 2003)

The Blobworld method is a well-known segmentation method in which the algorithm extracts the object or region of interest by clustering the pixels into a joint color-texture-position feature space (Liua *et al*., 2007). The KMCC algorithm provides the object or region of interest, inspired by the K-means clustering algorithm (Deng *et al*., 1999; Serra, 2003). Hence, the selection or choice of a segmentation algorithm is based on image type. In the object or region, descriptors have been classified into two parts: boundary-based and region-based (Yap *et al*., 2003). As previously mentioned, many descriptors have been introduced for this purpose, such as Fourier descriptors, moment with variety types, signature, compactness, chine code and others (Mingqiang *et al*., 2008). The moment with variety types, such as the Zernike moment (Zolkifli *et al*., 2011) and the Legendre moment, are commonly used to extract the regional features. Lastly, the Krawtchouk moment invariant was introduced to extract both types of feature (global and local) (Yap *et al*., 2003; Mingqiang *et al*., 2008; Arif *et al*., 2009).

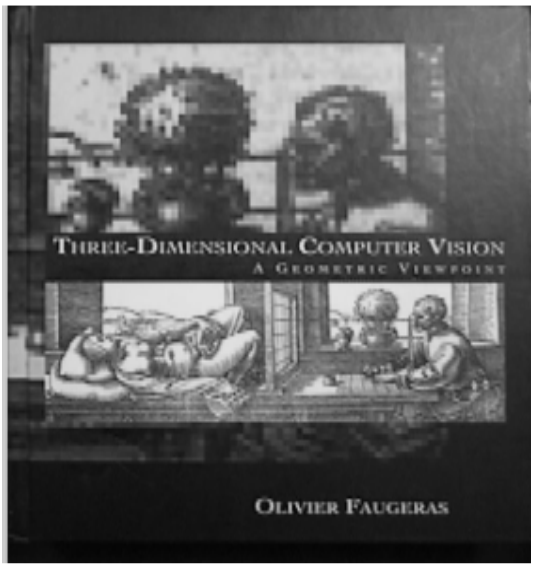
Based special technique: Several techniques, such as SIFT and the Speed up Robust Feature (SURF), have been introduced to overcome the limitation of the segmentation methods and to provide a local feature. The characteristics of these descriptors are distinctive, do not require segmentation and are robust to most image effects, including the occlusion effect. These techniques require two stages in obtaining the local feature: the interest point detector, such as the Harris detector and the feature descriptors, such as the moment invariant. These descriptors should have invariant properties, such as invariance to the geometric distortion and photometric changes (Ke and Sukthankar, 2004; Bay *et al*., 2008).

Interest point detector: Since the 1980s, several detectors have been introduced to detect the interest point. The first detector, namely, the corner detector, was introduced by Moravec (Mikolajczyk and Schmid, 2005). Later, the Harris corner detector was introduced to overcome the limitations of the corner detector; it is used widely in a variety of applications, such as object recognition. This detector is not invariant to all transformation effects (Harris and Mike, 1988).

Next, Harris Laplace and Hessian Laplace were introduced as invariants to rotation and scaling. The Harris affine region, Hessian affine region and others were introduced in the last decade to overcome the limitations of the previous methods and to improve the local feature (Mikolajczyk and Schmid, 2005). Other detectors, such as the Kirsch, Canny, Laplacian, Sobel and Prewitt edge detector, have also become popular for extracting interesting local features (Abdullah *et al*., 2007a). In summary, the previous researchers found that the Kirsch edge detector produced better results when discriminating character or local object classes.

Feature descriptor: Due to the need for invariant feature descriptors in extracting local features, many descriptors have been presented, such as moment invariants, differential invariants, phase-based local features, Gaussian derivative, complex features, steerable filter and others (Mikolajczyk and Schmid, 2005; Bay *et al*., 2008). Most of these descriptors are not invariant to all image effects. Therefore, the SIFT local descriptor was introduced to be invariant to the transformation effects, noise and occlusion and partially affine to the illumination change and 3D viewpoint (Lowe, 2004). Principal Component Analysis-Scale Invariant Feature Transform (PCA-SIFT) was introduced to reduce the feature taken by SIFT and to improve the time and scalability in matching schema. However, the drawback of this descriptor is that any application that depends on the PCA technique will be limited because this technique requires an offline stage for training and to approximate the covariance matrix (Ke and Sukthankar, 2004). Later, Mikolajczyk and Schmid (2005) used the PCA with the Gradient Location and Orientation Histogram (GLOH) descriptor to improve the performance of the local descriptor. Furthermore, the SURF descriptor is able to perform like SIFT, but faster; this is achieved by reducing the number of features taken by SIFT (Bay *et al*., 2008). Figure 6 describes the SIFT feature descriptor.

Combination feature: Content-based image retrieval is still an open area for research for the purpose of finding several techniques to perform the CBIR and to increase its performance. In fact, using a combination feature seems to be more effective than the other methods (Alemu *et al*., 2009; Zolkifli *et al*., 2011). The use of a combination feature is more effective in image retrieval compared to the traditional CBIR system, which uses either global or local features. The advantage of the combination feature is that each feature complements the other, improving the system's retrieval accuracy and making the system's retrieval more intelligent (Liu *et al*., 2007; Wu and Wu, 2009; Chih-Fong and Wei-Chao, 2009; Pujari *et al*., 2010).



(a)

(b)

Fig. 6: Illustration of the SIFT feature descriptor; (a) Original grey image, (b) SIFT feature assigned with 882 key point

Several studies have shown that the use of the combination feature is better than using any descriptor individually. Many works have introduced different techniques for combining features to increase the system's efficiency. Some have used the combination of color, texture and shape features (Zhang and Yang, 2008; Pujari *et al*., 2010), while others used the combination of color and texture features with weight operator assignment to assign weight to the features (Lui *et al*., 2007). Still others used a combination of global and local shape features (Wu and Wu, 2009; Zolkifli *et al*., 2011). On the other hand, Mohan *et al*. (2008) used both types of local features, namely, boundary and region. As a result, this combination achieved a robust feature that supports the CBIR system and improves the accuracy. Finally, spatial information plays an important part in achieving the high-level feature.

In fact, Abdullah *et al*. (2010) presented an intelligent approach by combining the MPEG-7 descriptor with spatial information. This operation is achieved with cluster correlogram. This system works well with different types of datasets but performs less well with Pascal 2007 because there are many small objects that the system does not recognize.

SIMILARITY MEASURE

Similarity measure is one of the most important parts of many applications, especially CBIR and object recognition, because these applications need to make decisions. As a result, efficient methods have been proposed in the last decade to obtain accurate results. Based on the literature, similarity measure methods can be classified into two parts. First, the traditional methods are the distance measures, such as Euclidean distance, Manhattan distance and Tanimoto distance (Abdullah *et al*., 2010). Second, both supervised and unsupervised machine learning algorithms have been introduced to obtain similarity measures, such as SVM and K-nearest neighbor (Zhu *et al*., 2007). Both of these methods have advantages and disadvantages. The traditional method is easy to implement and produces good accuracy. The machine learning algorithm produces precise results and is commonly used in object recognition, handwriting and CBIR, but is not easy to implement (Chih-Fong and Wei-Chao, 2009). In fact, the traditional methods, which include the Euclidean distance measure, the log sum square and percentage error and the Canberra distance measure, have been used by Hiremath and Pujari (2007), Zhang and Yang (2008), Wu and Wu (2009) and Pujari *et al*. (2010). The computation of these methods has been characterized as simple and easier to implement. In addition, Chih-Fong and Wei-Chao (2009), Abdullah *et al*. (2010) and Van de Sande *et al*. (2010) have used machine learning algorithms to perform the similarity measures, namely, the SVM classifier and K-nearest neighbor. The literature shows that these algorithms offer precise classification results. However, there are drawbacks to the machine learning algorithms. Because they are not always clear, humans are required to investigate to carry out the training and the sample number used to train the classifier is unfair. In contrast, the benefits of using the machine learning algorithms are that they do not require human resources to make rules and that they have consistent classification (Abdullah *et al*., 2010).

Conceptual framework for improving the gabor filters and edge histogram descriptor for object categorization: The state-of-the-art researches show that using filter-based descriptor is much effective in achieving the high-level feature. Among the different filter based-descriptors, there are two filters commonly used in computer vision applications such as object categorization, image retrieval systems and object matching, namely, the Gabor filter and wavelet transform. These two descriptors show that they are more robust with different types of image effects; both of them have been proposed for rectangle images (Wang *et al*., 2001; Chih-Fong and Wei-Chao, 2009).

Basically, the Gabor filters supply's meaningful information by using diverse magnitude and orientation. Furthermore, several salient features can be captured by this filter such as spatial frequency characteristic, spatial location and orientation selective (Daugman, 1988; Moghadam *et al*., 2012). Generally, the characteristics that make this filter commonly used in computer vision applications summarized as follow:

- The shapes of Gabor filters are similar to the interested fields of simple cells in the primary visual cortex.
- The Gabor filters are best for measuring local spatial frequencies.
- Gabor filters have been found to yield distortion tolerant feature spaces for other pattern recognition tasks, including texture segmentation, handwritten numeral recognition and fingerprint recognition.
- The Gabor function gets the best time-frequency resolution for signal analysis.

Manily, Choi *et al*. (2008) verifies the efficacious of the Gabor filter in object detection, recognition and tracking. Shen and Ji (2009) proposed a framework that used the Gabor filter with SVM in categorizing objects, the result showed that the proposed framework is working well under certine conditions.

However, in reviewing the state-of-the-art-researches, we found that this filter construct a redundant and incompact filters that may extract sufficient features and affect on the system accuracy and decrease the whole system performance.

A few studies has been introduced in the literature review to overcome the target problem of Gabor filter such as Zhu *et al*. (2004) proposed a framework to overcome the limitation of the Gabor filter through reducing the Gabor filter dimensionality and produce

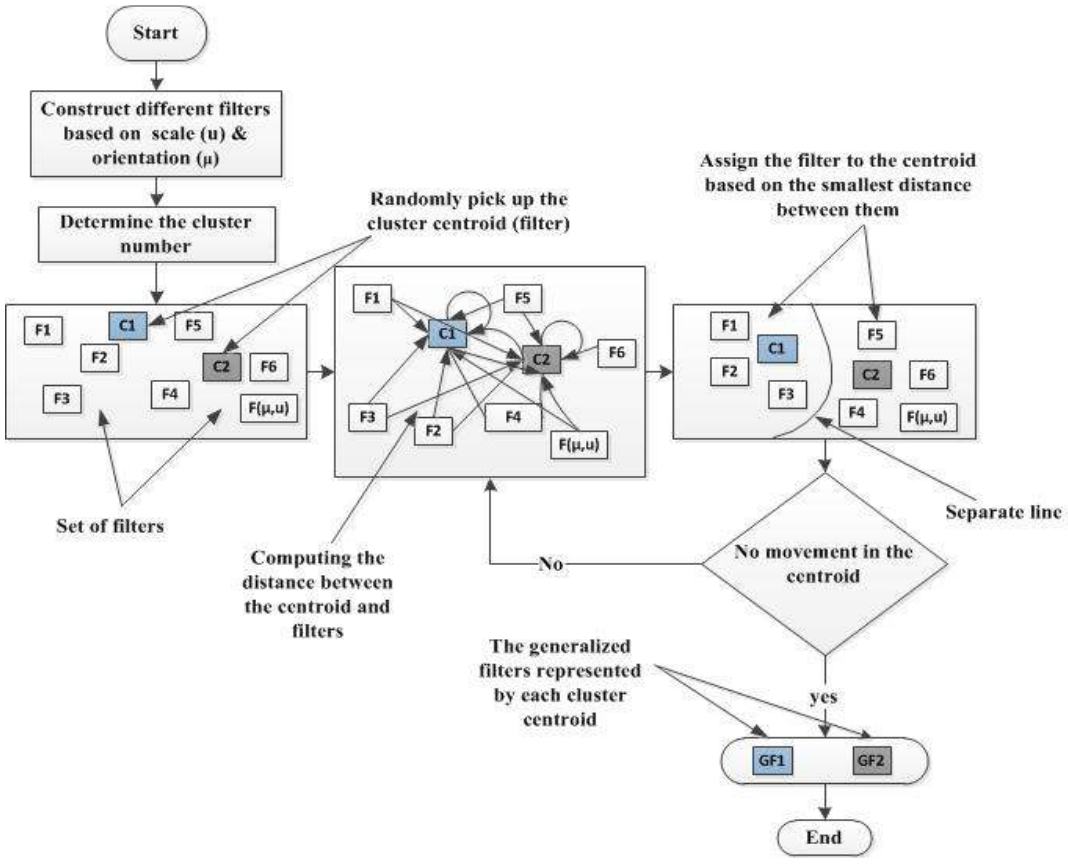


Fig. 7: The proposed framework of the generalize gabor filter

optimal filters by using Principle Component Analysis (PCA) for a face detection purpose. In addition, the Fast Fourier Transform algorithm (FFT) has also been proposed to improve the target problem through reducing the Gabor filter dimensionality in the frequency domain (Zhu *et al*., 2004).

Several researches show that using clustering technique has an advantage in grouping data into similar group (cluster). Each group has members are similar between themselves and dissimilar with the other groups. To be clearer, the clustering provides an accurate characterization of unobserved samples generated from the same probability distribution.

Based on the literature, many researches were used the clustering technique for different purpose. Basically, Bag of word technique used the K-means clustering technique to minimize and generalize the features of Scale Invariant Feature Transform (SIFT) descriptor. This technique ensures the removable of unwanted features and produce generalize features that give accurate representation to the object. Furthermore, the K-means shows higher potential results when it used for segmentation purpose. An example, the K-Means with Connectivity Constraint (KMCC) segmentation method, this method provide the object or region of interest, inspired by the K-means clustering algorithm.

Therefore, we propose the K-means clustering technique as a future work to improve the Gabor filter performance and generalizing set of Gabor filters. This

Table 1: Brief summary to the well-known descriptors that used in VOC and CBIR

Descriptor type	Descriptor name	Strength	Weakness
Color	Color histogram	It is easy to compute, can be used for both global and local descriptions of the color, and is robust to translation, rotation, occlusion and viewing angle.	The color histogram does not take the spatial information of the pixel, lack of robustness against scaling, and noise effects into account.
	Color Coherence Vector (CCV)	Same as color histogram with the addition of spatial information.	The computation is expensive due to the high dimensions, is not invariant to the scale, and illumination effects vary.
	Color correlogram	Same as above with the addition of the spatial correlation of each pair of colors.	Computational cost, not invariant against the reset of effect, such as scale, and illumination effects vary.
Texture	Tamura features	It has some invariant properties in addition; first three features of this descriptor are more efficient than the last three, especially with image retrieval.	The main drawback occurs when applying it with multiple resolutions to account for scale.
	Wold texture feature	It has a strong harmonic component, highly directional textures have a strong evanescent component, and less structured textures tend to have a stronger in deterministic component.	The main drawback of this descriptor is its reduced effectiveness when the transformation effect, orientation and distortion occur in the image. Does not work well if the images are not structural and not homogenous.
	Edge histogram	This descriptor obtains the local feature (capturing the spatial distortion of the edge) and is invariant to most image effects, such as transformation.	This descriptor does not work well with the existing non discriminative feature map (very sensitive to the object and scene distortion); scale dependent.
	Gabor filters	It used in multi-resolution scheme for both texture feature and classification parts because it provides filters based on different size, orientation and scaling.	It constructs redundant filters; not all of these filters can provide meaningful information. The computational cost is high.
Shape	Curvature Scale Space (CSS)	It is invariant to translation, rotation and scaling.	It is primarily effective when the object is taken from different viewpoint.
	Fourier descriptor	It is invariant to some image effects such as transformation, is robust against noise, compact and easy to compute. This descriptor gives a unique description of the shape.	It has high computational cost, does not work well when occlusion occurs to the shape.
	Moment invariant	It is invariant to scale, rotation, translation.	It is dependence, incompleteness, sensitive to the noise and low discriminatory power.
	Zernike Moment (ZM)	It is invariant to the rotation effect; robust to noise and to the mirror variation that can occur to the shape and expressiveness, which means that a little information will be redundant.	It has high computational cost; not being affine to all transformation effect means it is scale and translation dependent; any small, local disturbance can affect the descriptor value.
	Legendre moment	It is robust against most image effects such as translation and rotation, robust to noise.	The main drawback is the discretization error, which accrues when the order of moments increases; information redundancy, noise sensitivity and computation cost are high.
	Krawtchouk moment invariant	The error does not exist when computing this moment, no need for spatial normalization because the discretization error does not exist, affine to the transformation effect, can be used for both global and local features.	It is partially affine to the noise effect, not affine against occlusion and deformation effects.
	SIFT	It is distinctive; does not require segmentation to obtain the local feature and it has invariant properties against transformation effects, occlusion, and clutter; partially affine against illumination change and viewpoint.	The main drawback of this descriptor is the high-dimensional feature vectors it constructs.

process ensure in removable the redundant filters and produce compact and optimal filters that have new characteristics, can capture the most important salient features and representing the object efficiently. Figure 7 illustrate the conceptual framework of generalization of Gabor filter.

On the other hand, MPEG-7 introduced edge histogram for extracting local texture feature. This descriptor capturing the spatial distortion of the edge and is invariant to most image effects, such as transformation. Conversely, this descriptor is considered as an orientation descriptor and it has a benefit in capturing the most important features that are used in describing the object (Ayad *et al*., 2012). In fact, Abdullah *et al*. (2010) used this descriptor and its report shows that the main issue that facing the orientation descriptor in specific the edge histogram is the using of single orientation in extracting the local texture feature. This limitation makes this descriptor fail in capturing the salient feature that is important in describing and classifying the objects correctly.

Therefore, we proposed as an advance filter namely Gabor filter to overcome the overcome the target problem and improving the performance of this descriptor by providing different feature maps based on different scale and orientation. The proposed method will allow this descriptor to produce excellent results and a distinctive texture feature (Ayad *et al*., 2012).

DISCUSSION

In this section, we provide a summary of the strengths and weaknesses of some of the basic descriptors used in computer vision applications, mainly in image retrieval and categorization systems (Table 1).

CONCLUSION

Based on this review of the content-based image retrieval system, the main problem associated with research in the area, the semantic gap, is specified. The literature has identified several factors affecting the low-level feature that leads to the semantic gap. These factors include image effects, such as transformation, illumination, different viewpoints, noise and occlusion on one side and the segmentation methods on the other side. Clearly, the dataset used to demonstrate the performance of the system plays an important part in reducing the semantic gap. This is because interference between the object's foreground and background, if any, could cause confusion in the descriptor's ability to describe the image clearly, especially with shape descriptors. Examples of these datasets containing real-world images are the Pascal, Caltech 101 and Corel datasets. As a result, the dataset is also considered as another factor that can affect the semantic gap. Therefore, choosing the dataset is extremely important in reducing this gap. Several descriptors have been presented, together with their image effects. Most of them have invariant properties to these effects. In addition, the sources of weakness and the strengths of each descriptor have been specified and constructive critiques of most of these descriptors have been made. Therefore, this study's recommendations for color descriptors are the color correlogram and the SIFT color descriptor with diverse types because they give precise results under certain conditions. For the texture descriptor, the Gabor filter and wavelet transformation are recommended because they give better accuracy and provide a multi-resolution texture feature and classification schema. In addition, the edge histogram is recommended because it gives precise results in applications in which the feature map is almost clear. Furthermore, we clarify the main issue in Gabor filter and edge histogram and then we proposed a future framework to overcome these limitations. Apart from that, the Fourier descriptor and the Krawtchouk moment invariant outperform than the other descriptors, such as the Hu moment, Zernike moment and Legendre moment. Therefore, they are recommended for the shape feature. The special techniques used for obtaining the local feature, such as SIFT and SURF, are recommended for use as local descriptors due to the advantages that these descriptors possess. From the state-of-the-art research, two methods of obtaining the local feature have been identified, using segmentation or special techniques, such as SIFT, SURF and others. The problem of how to reach a clear object still arises in the segmentation methods. Therefore, the effect that occurs after the segmentation process such as object deformation can be remedied using the combination features. In addition, this review found that the features produced by the combination features are robust against several effects and could increase the system's performance. Finally, the similarity measures of both types have advantages and disadvantages. Therefore, selecting one of the methods depends on several factors, such as the precise similarity, the difficulty of the computational task and the time spent on deciding the matching.

REFERENCES

Abdullah, A., R.C. Veltkamp and M.A. Wiering, 2010. Fixed partitioning and salient points with MPEG-7 cluster correlograms for image categorization. Pattern Recogn., 43: 650-662.

Abdullah, A. and Wiering, M.A. 2007. CIREC: Cluster Correlogram Image Retrieval and Categorization using MPEG-7 Descriptors. IEEE Symposium on Computational Intelligence in Image and Signal Processing.

Abdullah, S.N.H.S., M. Khalid, R. Yusof and K. Omar, 2007a. Comparison of feature extractors in license plate recognition. Proceedings of 1st Asia International Conference on Modelling and Simulation (AMS2007). Phuket, Thailand, pp: 502-506.

Abdullah, S.N.H.S., M. Khalid, R. Yusof and K. Omar, 2007b. License plate recognition based on geometrical features topology analysis and support vector machine. Proceeding of the Malaysia-Japan International Symposium on Advanced Technology (MJISAT'2007). Kuala Lumpur, Malaysia.

Abdullah, S.N.H.S., K. Omar, S. Sahran and M. Khalid, 2009. License plate recognition based on support vector machine. Proceeding of the International Conference on Electrical Engineering and Informatics (ICEEI'09), 1: 78-82.

Alemu, Y., J.B. Koh, M. Ikram and D.K. Kim, 2009. Image retrieval in multimedia databases: A survey. Proceeding of the 5th International Conference on Intelligent Information Hiding and Multimedia Signal Processing.

Arif, T., Z. Shaaban, L. Krekor and S. Baba, 2009. Object classification via geometrical, zernike and legendre moments. J. Theor. Appl. Inf. Technol., 7: 31-37.

Ayad, H., S.N.H.S. Abdullah and A. Abdullah, 2012. Visual object categorization based on orientation descriptor. Proceeding of the 6th Asia International Conference on Mathematical Modeling and Computer Simulation (ASM'2012). Bali, Indonesia.

Bataineh, B., S.N.H. Abdullah and K. Omar, 2011. A statistical global feature extraction method for optical font recognition. In: Nguyen, N.T., C.G. Kim and A. Janiak (Eds.), ACIIDS, 2011. LNAI 6591, Springer-Verlag, Berlin, Heidelberg, pp: 257-267.

Bay, H., T. Tuytelaars and L.V. Gool, 2008. SURF: Speeded up robust features. Comput. Vis. Image Und., 110(3).

Bin Adam, H., M. Fauzi bin Hassan, M. Jaisbin Gimin and A. Bin Abdullah, 2011. Material surface analysis for robot labeling. Proceeding of the International Conference on Pattern Analysis and Intelligent Robotics (ICPAIR), 1: 136-138.

Canny, J., 1986. A computational approach to edge detection. IEEE T. Pattern Anal., 8: 679-698.

Chih-Fong, T. and L. Wei-Chao, 2009. A Comparative study of global and local feature representations in image database categorization. Proceeding of the 5th International Joint Conference on INC, IMS and IDC (NCM'09), pp: 1563-1566.

Choi, W.P., S.H. Tse, K.W. Wang and K.M. Lam, 2008. Simplified gabor wavelets for human face recognition. Pattern Recogn., 41(3): 1186-1199.

Daugman, J.G., 1988. Complete discrete 2-D gabor transforms by neural networks for image analysis and compression. IEEE T. Acoust. Speech, 36(7): 1169-1179.

Deb, S., 2008. Overview of image segmentation techniques and searching for future directions of research in content-based image retrieval. Proceeding of the 1st IEEE International Conference on Ubi-Media Computing, pp: 184-189.

Deng, Y., B.S. Manjunath and H. Shin, 1999. Color image segmentation. Proceeding of the CVPR, pp: 2446-2451.

Harris, C. and S. Mike, 1988. A combined corner and edge detector. Proceeding of the Alvey Vision Conference, pp: 147-151.

Hiremath, P.S. and J. Pujari, 2007. Content based image retrieval using color, texture and shape features. Proceeding of the International Conference on Advanced Computing and Communications (ADCOM'2007), pp: 780-784.

Hu, M.K., 1962. Visual pattern recognition by moment invariants. IRE T. Inform. Theor., 8: 179-187.

Jaswal, G. and A. Kaul, 2009. Content based image retrieval: A literature review. Proceeding of the National Conference on Computing, Communication and Control (CCC-09), pp: 198-201.

Ke, Y. and R. Sukthankar, 2004. PCA-SIFT: A more distinctive representation for local image descriptors. Proceeding of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR, 2004), 2: 506-513.

Kurniawati, N.N., S.N.H.S. Abdullah and S. Abdullah, 2009. Investigation on image processing techniques for diagnosing paddy diseases. Proceeding of the International Conference of Soft Computing and Pattern Recognition (SOCPAR'09), pp: 272-277.

Leow, W.K. and S.Y. Lai, 2000. Scale and Orientation-invariant Texture Matching for Image Retrieval. In: Pietikainen, M.K. (Ed.), Texture Analysis in Machine Vision. World Scientific, Singapore.

Liua, Y., D. Zhanga, G. Lua and W.Y. Mab, 2007. Asurvey of content-based image retrieval with high-level semantics. Pattern Recogn., 40: 262-282.

Long, F., H. Zhang and D.D. Feng, 2003. Multimedia Information Retrieval and Management: Technological Fundamentals and Applications. Springer, New York.

Lowe, D.G., 2004. Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vision, 60: 91-110.

Lui, P., K. Jia and Z. Wang, 2007. An effective image retrieval method based on color and texture combined features. Proceeding of the 3rd International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIHMSP'2007), pp: 169-172.

Lukac, R. and K.N. Plataniotis, 2007. Color Image Processing: Methods and Applications. Taylor and Francis Group, LLC., Boca Raton.

Manjunath, B.S., J.R. Ohm, V.V. Vasudevan and A. Yamada, 2001. Color and texture descriptors. IEEE T. Circ. Syst. Vid., 11: 703-715.

Mikolajczyk, K. and C. Schmid, 2005. A performance evaluation of local descriptors. IEEE T. Pattern Anal., 27: 1615-1630.

Mingqiang, Y., K. Kidiyo and R. Joseph, 2008. A Survey of Shape Feature Extraction Techniques. In: Peng-Yeng Yin (Ed.), Pattern Recognition Techniques, Technology and Applications. I-Tech, Vienna, Austria.

Moghadam, P., J.A. Starzyk and W.S. Wijesoma, 2012. Fast vanishing-point detection in unstructured environments. IEEE T. Image Process., 21(1): 425-430.

Mohan, V., P. Shanmugapriya and Y. Venkataramani, 2008. Object recognition using image descriptors. Proceeding of the International Conference on Communication and Networking Computing (ICCCn 2008), pp: 1-4.

Nadernejad, E., S. Sharifzadeh and H. Hassanpour, 2008. Edge detection techniques: Evaluations and comparisons. Appl. Math. Sci., 2: 1507-1520.

Pabboju, S. and A.V.G. Reddy, 2009. Novel approach for content-based image indexing and retrieval system using global and region features. Int. J. Comput. Sci. Netw. Secur., 9(2): 119-130.

Pujari, J., S.N. Pushpalatha and P.D. Desai, 2010. Content-based image retrieval using color and shape descriptors. Proceeding of the International Conference on Signal and Image Processing (ICSIP), pp: 239-242.

Sarfraz, M., 2006. Object recognition using fourier descriptors: Some experiments and observations. Proceedings of the International Conference on Computer Graphics, Imaging and Visualisation (CGIV, 2006), pp: 281-286.

Serra, J., 2003. Image segmentation. Proceeding of the IEEE International Conference on Image Processing (ICIP).

Shen, L.L. and Z. Ji, 2009. Gabor wavelet selection and SVM classification for object recognition. Acta Automatica Sinica, 35(4): 350-355.

Shi-Kuo, C., S. Qing-Yun and Y. Cheng-Wen, 1987. Iconic indexing by 2-D strings. IEEE T. Pattern Anal., 9: 413-428.

Tamura, H., S. Mori and T. Yamawaki, 1987. Textural features corresponding to visual perception. IEEE T. Syst. Man Cyb., 8: 460-473.

Van de Sande, K.E.A., T. Gevers and C.G.M. Snoek, 2010. Evaluating color descriptors for object and scene recognition. IEEE T. Pattern Anal., 32: 1582-1596.

Wang, J.Z., J. Li and G. Wiederhold, 2001. SIMPLIcity: Semantics-sensitive integrated matching for picture libraries. IEEE T. Pattern Anal., 23(9): 947-963.

Wu, Y. and Y. Wu, 2009. Shape-based image retrieval using combining global and local shape features. Proceeding of the 2nd International Congress on Image and Signal Processing (CISP'09), pp: 1-5.

Xiaoyin, D., 2010. Image retrieval using color moment invariant. Proceeding of the 7th International Conference on Information Technology: New Generations (ITNG), pp: 200-203.

Xu, C. and J.L. Prince, 1998. Snakes, shapes and gradient vector flow. IEEE T. Image Process., 7: 359-369.

Yang, H. and Q. Wang, 2008. A novel local feature descriptor for image matching. Proceeding of the IEEE International Conference on Multimedia and Expo, pp: 239-242.

Yap, P.T., R. Paramesran and O. Seng-Huat, 2003. Image analysis by krawtchouk moments. IEEE T. Image Process., 12: 1367-1377.

Zhang, Y. and J. Yang, 2008. An object based image retrieval. Proceeding of the 2nd International Symposium on Intelligent Information Technology Application (IITA'08), pp: 385-388.

Zhu, J., M.I. Vai and P.U. Mak, 2004. A new Enhanced Nearest Feature Space (ENFS) classifier for gabor wavelets features-based face recognition. In: Zhang, D. and A.K. Jain (Eds.), ICBA, 2004. LNCS 3072, Springer-Verlag, Berlin, Heidelberg, pp: 124-131.

Zhu, Y., X. Liu and W. Mio, 2007. Content-based image categorization and retrieval using neural networks. Proceeding of the IEEE International Conference on Multimedia and Expo, pp: 528-531.

Zolkifli, Z.F.M., M. Farif Jemili, F. Hashim and S.N.H.S. Abdullah, 2011. Optimal features and classes for estimating mobile robot orientation based on support vector machine. Proceeding of the 14th FIRA RoboWorld Congress. Kaohsiung, Taiwan.