المجلة الإلكترونية الشاملة متعددة التخصصات
العدد الرابع والثلاثون شهر (٣) ٢٠٢١

An Approach To Image Classification Based On SURF Descriptors And Colour Histograms

Rawabi Alwanin

College of Computer and Information Science, King Saud University

Riyadh, Saudi Arabia

E-mail: ralwaneen@ksu.edu.sa

Abstract

Image matching and classification is one of the challenging problems in the field of computer vision and image processing. It is based on classifying images into semantic classes using low level features such as colour and orientation. However, the task of image matching and classification plays an important role in a variety of applications including object recognition, content based image retrieval systems, image indexing and many more.

A variety of image classification techniques have been introduced. SURF (Speeded Up Robust Features) is one of the important invariant feature descriptors which is mainly applied to grayscale images. However, colour information plays an important role in matching and classifying images. Thus, we propose a novel approach that balances between geometrical characteristics and colour information by combining different image classification techniques including SURF, colour histograms and spatial colour histograms.

In this project, we show how our classification scheme performs on an image dataset consists of landscapes, faces and buildings and retrieved automatically from Google Images using keyword searches. For image comparison, we construct three kinds of models: the most typical images, image clusters and composite images, which are formed from image clusters.

Preliminary results show that image clusters and composite images have better performance when compared with query images (accuracy 60.8% and 81.67% respectively). They also show that colour features have more discrimination power than geometrical features for the classification problem considered in this study. Extensive experimental evaluations show that our approach results in an accuracy of 85.83% and has better accuracy than the original SURF. The number of correct classifications is increased by about 31.67%.

Keywords: Images, Processing, Matching, Diagrams, Colours , SURF.

ملخص البحث

تعد مطابقة الصور وتصنيفها إحدى المشكلات الصعبة في مجال رؤية الكمبيوتر ومعالجة الصور. يعتمد على تصنيف الصور إلى فئات دلالية باستخدام ميزات منخفضة المستوى مثل اللون والاتجاه. ومع ذلك ، فإن مهمة مطابقة الصور وتصنيفها تلعب دورًا مهمًا في مجموعة متنوعة من التطبيقات بما في ذلك التعرف على الكائنات وأنظمة استرجاع الصور القائمة على المحتوى وفهرسة الصور وغيرها الكثير.

تم تقديم مجموعة متنوعة من تقنيات تصنيف الصور. تعد (SURF) الميزات القوية المعجلة واحدة من واصفات الميزات الثابتة المهمة والتي يتم تطبيقها بشكل أساسي على الصور ذات التدرج الرمادي. ومع ذلك ، تلعب معلومات الألوان دورًا مهمًا في مطابقة الصور وتصنيفها. وبالتالي ، فإننا نقترح نهجًا جديدًا يوازن بين الخصائص الهندسية ومعلومات الألوان من خلال الجمع بين تقنيات تصنيف الصور المختلفة بما في ذلك SURF والمخططات الملونة ومخططات الألوان المكانية.

في هذا المشروع ، نعرض كيفية أداء مخطط التصنيف الخاص بنا على مجموعة بيانات الصور التي تتكون من مناظر طبيعية ووجوه ومباني ويتم استردادها تلقائيًا من صور Google باستخدام عمليات البحث عن الكلمات الرئيسية. لمقارنة الصور ، نقوم ببناء ثلاثة أنواع من النماذج: الصور الأكثر نموذجية ومجموعات الصور والصور المركبة ، والتي تتكون من مجموعات الصور.

تظهر النتائج الأولية أن مجموعات الصور والصور المركبة لديها أداء أفضل بالمقارنة مع صور الاستعلام (دقة 60.8٪ و 81.67٪ على التوالي). كما أنها توضح أن ميزات اللون تتمتع بقوة تمييز أكبر من السمات الهندسية لمشكلة التصنيف التي تم تناولها في هذه الدراسة. تُظهر التقييمات التجريبية المكثفة أن نهجنا ينتج عنه دقة تصل إلى 85.83٪ ولديه دقة أفضل من SURF الأصلي. تم زيادة عدد التصنيفات الصحيحة بحوالي 31.67٪.

الكلمات المفتاحية : الصور ،المعالجة ،المطابقة ،المخططات ، الألوان ، SURF .

1. Introduction

In recent years, the World Wide Web has played a key role in the rapid development of digital photography leading to large image and video libraries (including programs, news, games, and art) available online in digital format. Real-time browsing and retrieval has resulted in a growing need for effective techniques to index these libraries and organise them into categories in order to make them more useful. This need has in turn resulted in the emergence of image classification techniques.

Image classification is one of the major tasks in computer vision and image processing and is the core of many applications. It can be defined as grouping images into semantic classes based on image features. It is an emerging technology that is used to tackle the problem of many computer vision applications including object recognition, image indexing and content based image retrieval. Content based image retrieval has become an increasingly important area in computer vision and multimedia computing. Successful classification of images results in filtering out irrelevant images which improves the performance of such systems. However, image classification is a challenging problem that is based on finding reliable similarities between images that belong to the same class or represent the same object.

1.1 Aim

The main aim of this project is to improve image classification accuracy by introducing a new approach that balances between colour information and geometrical characteristics through combining three different classification techniques. Thus, the project investigates the problem of image classification by analysing the performance of different image classification techniques.

1.2 Objectives

To achieve the aim of this project, the following objectives should be attained:

- Gain a clear understanding of current classification techniques and identify their weaknesses to take corrective actions. Furthermore, review recent research into efforts that have been made so far to improve image classification
- Construct different kinds of models which are then used for image comparison and classification.
- Design, implement and test the classification system.
- Evaluate the overall performance of the classification system using a set of images that belong to different classes.

2. Literature Review

In recent years, sharing of digital photos has become widespread due to the availability of Internet access. As a result, hundreds and hundreds of image and video libraries are globally available on the Internet and are easily accessed. Thus, developing efficient image retrieval and indexing systems is becoming increasingly important and a major area of interest. This necessitates the need to classify digital images into semantically different classes (Shukla, Mishra, & Sharma, 2013).

When people search a database for images, they either know exactly what they are asking for, such as images for people, animals or buildings, or they have an abstract idea of what they are looking for, such as images for planning a vacation. For this type of queries, there is a need for classifying these images into classes based on abstract concepts so that only target images are retrieved and explored (Vailaya, Jain, & Zhang, On Image Classification: City Images vs. Landscapes, 1998).

Basically, the task of image classification consists of forming an appropriate representation of images and then comparing these representations in order to find correspondences (Chapelle, Haffner, & Vapnik , 1999). Image classification is a challenging problem that lies on reliably finding similarities among images that represent the same object based on objects' descriptors or in other words describing an image based on the semantic scene it represents (Bay, Tuytelaars, & Gool, 2006) (Szummer & Picard, 1998).

In computer vision and image processing applications, the task of finding similarity between images that represent the same object is increasingly becoming a challenging problem. Image classification is based on image features including colour, orientation and edge. In order for image classification to be more accurate, these features should be invariant to different image transformations such as rotation, illumination, scale, viewpoint, noise, etc. This is because similar images that have different viewing conditions are sometimes considered different which should be avoided. Thus, selecting invariant features is one of the important steps that affect classification performance (Khan, McCane, & Wyvill, 2011).

Another problem that image classification brings is turning these low level features into semantic classes. A case in point is content-based image retrieval systems, which take advantage of image classification. Users often search an image using semantic queries (or what they are called high level features), such as "show me a sunset image" instead of "show me a predominantly red and orange image". However, low level features, as opposed to high level features, are all that can be reliably detected and extracted, for example colour histograms are reliably calculated from colour images but trees, faces or buildings cannot be easily detected. This leads image retrieval systems to bad performance when some semantic queries are used. So, the major problem arising from difficulty to semantically classify images into meaningful groups by turning these semantic queries into low level features (Vailaya & Zhang, Image classification for content-based indexing, 2001).

In computer vision, a fair amount of literature has been published on image classification and indexing. Due to the importance of image retrieval especially for Internet image search engines, (Fan, Men, Chen, & Yang, 2009) published a paper in which they investigated whether combining colour histograms along with SURF descriptors could increase descriptors' distinctiveness, as colour and geometric features are combined in a single feature vector. Since SURF algorithm works on grayscale images, they are motivated by the fact that colour is one of the important attributes of digital images that gain more attention in a variety of image processing applications as it provides useful information for the task of image classification and matching.

They have evaluated their algorithm on different colour images under different viewing conditions such as: scale, rotation, etc. Basically, what they did is constructing a two-part vector as follows. First, keypoints are detected and SURF descriptors are constructed using SURF algorithm. Next, a square window is formed around keypoints that are not matched by SURF and colour histograms are calculated for each window. To match two images, two different distance measures are used. 64-SURF descriptors are first compared using Euclidean distance to measure similarity between descriptors. If the distance ratio of the first best match and the second-best match is greater than 70% then it is considered a good match. Then, colour histograms for unmatched descriptors are compared for similarity using Bhattacharyya distance. So, what they are doing is closely related to what we are going to do as combining colour features and geometry features is the main task.

Experimental results have shown that combining colour features with SURF descriptors is more robust and distinctive compared to original SURF descriptors. In the evaluated dataset, the accuracy of the matching is increased by 8.9%.

Another study by Vailaya, et al. (1998) involved an approach to image classification of cities versus landscapes. What they are trying to do is bridging the gap between low level features and high level features of specific classes. Moreover, five image features (colour histogram, colour coherence vector, DCT coefficient, edge direction histogram, and edge direction coherence vector) have been evaluated for their distinction ability between city and landscape classes, where cities are identified by man-made objects such as buildings, cars and roads while landscapes do not have these structures.

The approach have been evaluated by comparing input set of 2716 city and landscape images to an existing training set of human-labelled images using K-Nearest Neighbour classifier, it results in 93.9% accuracy in classifying input images to city and landscape classes. Additionally, 528 landscape images have further been classified into subcategories: forests, mountains, and sunset/sunrise in accuracy of 91.7%.

So, what they have done is focussing on a particular classification problems (cities vs. landscapes) instead of learning all concepts. As a result, they have investigated the distinction power of each feature to find out which of them are appropriate to discriminate these kinds of images. Then, instead of classifying images based on a single feature, pairwise classification, based on most distinctive features, is performed. They have found that edge direction features (histograms and coherence vectors) are discriminative enough to classify city and landscape images ignoring reject option. This means that the classifier does not reject images that do not belong to either class.

Using low level histogram features for colour image classification was studied by (Sergyan, 2008). The aim of this study is to use simple image features such as colour histogram vectors that can be easily generated and compared. The main advantage of such features is that they have sufficient robustness and are efficiently generated and compared.

The approach was evaluated using 200 images of different classes: landscapes, buildings, faces and indoor images with one object with homogenous background. The results of this study show that 87% of images were accurately classified into their corresponding classes.

3. System Design and Implementation

The aim of this section is to give an overview of the image classification system developed in this project. The system architecture and components are explained in this section. The implementation methodologies of major parts of the system are also demonstrated.

3.1 System Design

System design is one of the key aspects of system development. It can be defined as the process of identifying system architecture, components and data in order to meet specified requirements. In this section, we indicate the process of developing our image classification system. This process helps to build the system starting from major components to small details.

3.1.1 System Definition and Scope

In our project, we develop an image classification system that combines three techniques: Speeded Up Robust Features, colour histograms and spatial colour histograms. By combining these different techniques, we expect to take advantage of both geometric features and colour features since SURF algorithm works only on grayscale images. A large part of the methodological framework involves conducting experiments and analysing results.

Before describing the general picture of the system, it is worth to mention that in our project, we construct three models for a given class. These models are taken to be representatives of that class. Once models are constructed, they will be examined on a set of images and the model that is more efficient and, at the same time, gives better accuracy will be used in the system as a representative of a class. Different approaches of constructing a model will be described later in this section.

So, the general picture of the system is as follows. For a given class, the top 40 images are retrieved from Google Images using keyword search. Then, a model is constructed for each class. The process of model construction is repeated for each technique yielding N * 3 different models (3 models for each class), where N indicates the number of classes. Image classification is performed on the same set of images using the three image classification techniques and the models that have been constructed for each technique. Since classification techniques are applied independently, classification results are combined using different methods which will be investigated in detail in section 3.3.3.

3.1.2 System Architecture and Components

The system architecture represents major functionalities of the system and illustrates the interactions that take place between different components. Building the system architecture gives a better understanding of the system design and how different components integrate. In our project, the system is designed to have a number of components. These components include: model construction, image comparison and decision making. System architecture is shown in figure 3.1.

- **Model Construction**

Model construction is a key component in our classification system and plays a key role in building different models for comparing images. In our approach, we define three ways of making a model: finding the most typical image, detecting image clusters and forming a composite image. However, we are going to select only one model for each technique and the selection is based on the accuracy that each model achieves. The different approaches to constructing models will be explained later in this section.

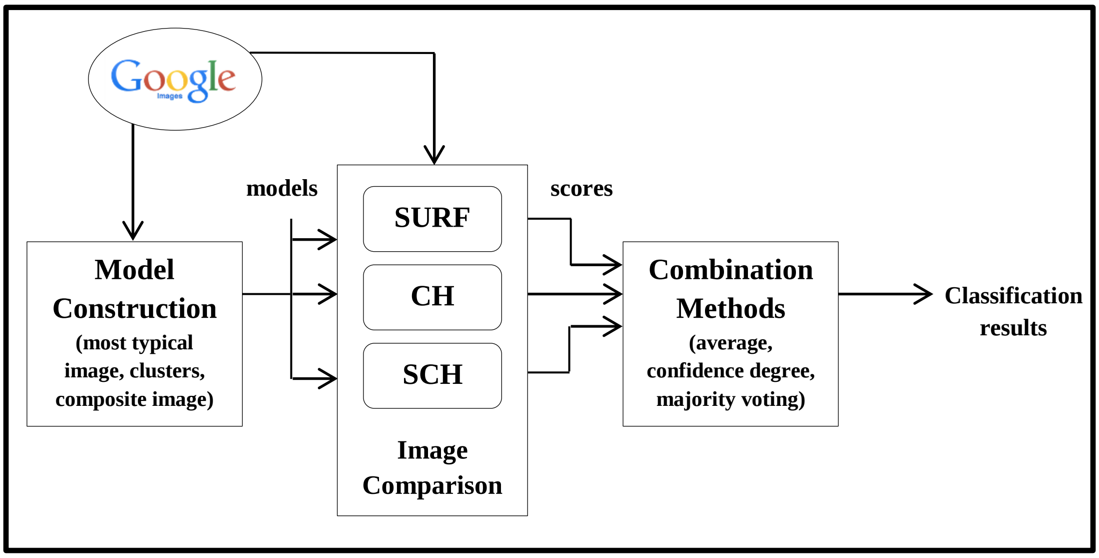
For a given keyword, the top 40 images are automatically retrieved from Google Images using Google Image Search API. Then, a model is simply constructed for that dataset and the process is repeated for each classification technique. The output of this component consists of three models for each keyword and these models are believed to be representative of the class they belong to. As mentioned earlier, a single kind of model for each technique will then be used as an input for the next component, which is image comparison and classification.

- **Image Comparison and Classification**

Once the top 40 images are retrieved from Google Images (using a certain number of keywords) and models are constructed for each keyword (i.e. for each class), the same set of images will then be used for image classification. For a given technique, every query image is classified into either one of the predefined classes according to the distance between the query image and a model of a class. Suppose that C_1, C_2, C_3 denote the three classes that are used and $\{M_{11}, M_{12}, M_{13}\}$, $\{M_{21}, M_{22}, M_{23}\}$, $\{M_{31}, M_{32}, M_{33}\}$ denote the three models that are constructed for C_1, C_2, C_3 respectively. To perform image classification, the query image QI is compared to every model of each class i.e. it is compared to M_{1j}, M_{2j} and M_{3j} and the distances are calculated. Mainly, QI is said to belong to the class C_i if the distance between QI and M_{ij} is the shortest. The output of this component consists of three different classification scores related to each technique.

- **Combination Methods**

In our project, we propose a novel approach that integrates different image features using three classification techniques. Thus, there is a need to combine different scores resulted from these techniques. The output of this component is an overall score based on the three classification techniques that have been used. Three approaches of combining classification results are developed including: average, confidence degree and majority voting. These methods will be described in detail in section 3.2.2.



* CH = Colour Histograms
SCH = Spatial Colour Histograms

Figure 3.1: System Architecture

3.2 Implementation

The algorithm is coded in Python as it is simple and powerful programming language that provides functionalities for manipulating images in conjunction with OpenCV. In this section, the implementation of the major functions in the system is explained in detail.

3.2.1 Image Classification Techniques

A variety of image classification techniques are available for extracting reliable features that will be then used for image matching. SURF, colour histograms and spatial colour histograms are three techniques that have been used in our algorithm.

3.2.1.1 SURF

As reported earlier in the previous section, SIFT (Scale Invariant Feature Transform) is computationally slow compared to SURF. Research has shown that SURF is three times faster than SIFT even though the accuracy is closely comparable (OpenCV, Introduction to SURF (Speeded-Up Robust Features), 2014). Thus, SURF will be used in this project. SURF (Speeded-Up Robust Features) is one of the image classification methods that have been proven to be the most discriminative feature descriptors among other invariant descriptors. Moreover, it has been reported to give good classification results.

SURF functionalities for detecting keypoints and extracting descriptors are provided by OpenCV. OpenCV (Open Source Computer Vision) is an open source library for computer vision and machine learning software. It is developed to provide an infrastructure for computer vision applications. The library has over 2500 algorithms including both computer vision and machine learning algorithms.

Furthermore, it supports different interfaces such as C, C++, java and python. In this project, we will be using SURF from OpenCV library along with some algorithms including, feature matching and histogram calculation algorithms (OpenCV, About, 2014).

Generally, finding point correspondences between two images using SURF consists of three phases. First, invariant points are detected under different image transformations such as: rotation, scale and noise. Next, every neighbour region around each point is represented by a feature vector which should be as discriminative as possible. Finally, two feature descriptors are matched based on some distance calculations.

Given an image, SURF detector is applied in order to detect interest points (or keypoints) based on the approximate Hessian matrix. Figure 3.2 shows detected key points within an image. The number of detected keypoints is 1357. For clarity, it has been reduced to some 50. It can be seen that SURF detects white blobs on a butterfly wings.

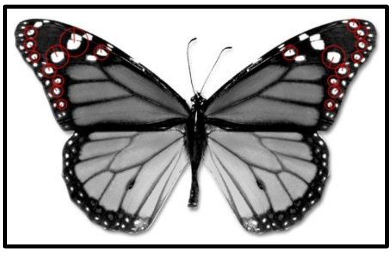


Figure 3.2: Key point Detection

The second step is representing the neighbourhood of every interest point by a feature vector. The feature vector that has been used in the project is a 64-dimensional vector which represents the relationship between the keypoints and the neighbouring regions. Furthermore, it demonstrates the distribution of pixel intensities within the neighbourhood of the keypoint it describes. So, for every detected keypoint, a descriptor is attached. SURF descriptors are extracted by determining the orientation based on information obtained from the region around the keypoint.

The process of descriptors extraction goes through two steps. First, every keypoint is assigned an orientation where all descriptor calculations are based on. The keypoint orientation is determined by calculating Haar wavelet responses for a set of pixels (Haar wavelets are filters used to find gradients in the x and y directions). The second step is constructing a square window around the keypoint and orienting it along the keypoint orientation. Then, the window is split up into equal 4 × 4 square subregions and Haar wavelet responses are computed for 5 × 5 sample grids in each subregion as shown in figure 3.3 (for illustrative purposes, we only show 2 × 2 sample grids in the left image.

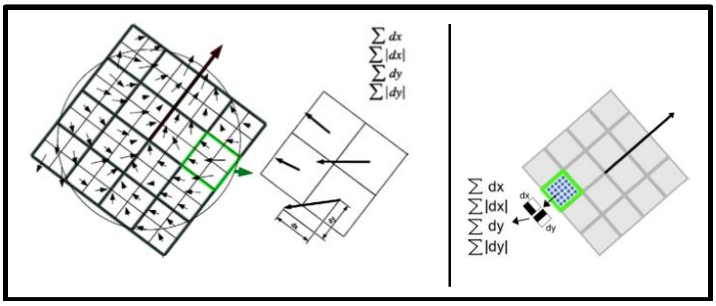


Figure 3.3: Descriptor Components

For each sub region, the sums dx, |dx|, dy and |dy| are computed based on the orientation of the sample grid, where dx is the x Haar wavelet response in horizontal direction and dy is the y Haar wavelet response in vertical direction and these directions are based on the keypoint orientation. So, a 4-dimensional vector (V) is calculated from each subregion as follows:

$$V_{subregion} = \left[\sum dx, \sum dy, \sum |dx|, \sum |dy|\right]$$

The above process is repeated for all 4 × 4 subregions and vectors are concatenated resulting in an overall descriptor vector of length 64.

In order to compare two images, a matching algorithm needs to be applied. In our project, K-Nearest Neighbour classifier is used with k = 2. Given two images A and B, each descriptor in A is compared with every descriptor in B and the closest k matches are returned using some distance calculations. So, the matcher returns matches (M_i, M_{j1}, M_{j2}) such that M_i (the i^{th} descriptor in A) has M_{j1} and M_{j2} (the two j^{th} descriptors in B) as the closest matches. The distance measurement that is used between descriptors is cv2.NORM_L2 which is based on the Euclidean distance between two feature descriptors. It calculates the absolute or relative difference norm as follows:

$$\mathbf{norm} = \|\, \mathbf{src_1} - \mathbf{src_2} \,\|_{\mathbf{L_2}} = \sqrt{\sum_{\mathbf{i}} (\mathbf{src_1(i)} - \mathbf{src_2(i)})^2}$$

Figure 3.4 shows a simple example on feature matching between two images.

Figure 3.4: Feature Matching

To select the correct matches and discard false ones, a ratio test, proposed by (Lowe, 1999), is applied. The correct matches are determined by calculating the ratio of distance from the closest match to the distance of the second closest match. If the ratio is below some threshold, the match is discarded. In our algorithm, the match is considered as a good match if the ratio of distance from the closest match to the distance of the second closest match is less than 0.75.

3.2.1.2 Single Colour Histogram

The second image classification technique that has been adopted in the project is colour histograms. As previous figures show, images are first turned into grayscale images before applying SURF algorithm and colour information is discarded. However, colour provides key information in image classification and matching task and when is considered, it is believed that more distinctions could be recognized than grey level versions of the same images. For example, landscapes tend to have similar colour distribution such as blue sky on the top and green grass on the bottom and buildings tend to have a grey colour. This leads us to employ colour histograms in our approach.

Figure 3.5: Sample Image

Taking the above image represented in figure 3.5, colour features are identified in terms of histograms in the HSV (Hue, Saturation, Value) colour space. First, the image is split into its H, S and V planes. Then, the histogram is calculated for each channel plane. In our approach, the histogram is calculated for hue channel.

As mentioned earlier, colour histograms are counts of number of pixels that have a particular intensity value organised into predefined bins. These bins are segmented according to the range of the value to be measured. Hue can take 181 values ranging from 0 to 180. In our algorithm, the range has been segmented into 181 bins as follows:

$$[\mathbf{0}, \mathbf{180}] = [\mathbf{0}] \cup [\mathbf{1}] \cup [\mathbf{2}] \cup [\mathbf{3}] \cup \ldots \cup [\mathbf{180}]$$

$$\mathbf{range} = \mathbf{bin_1} \cup \mathbf{bin_2} \cup \mathbf{bin_3} \cup \ldots \cup \mathbf{bin_{n=180}}$$

Then the number of pixels that belongs to each bin is counted. Figure 3.6 shows the histogram plot of the hue channel for the image represented in figure 3.5.

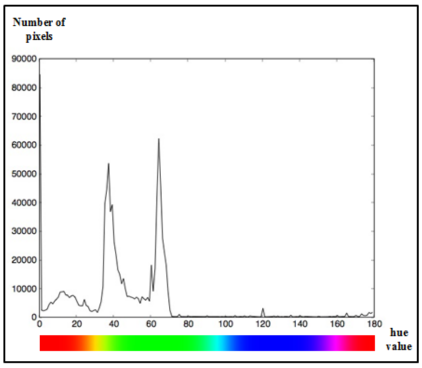


Figure 3.6: Histogram Plot

Two images can be compared for similarity by comparing their colour histograms. In order to compare two colour histograms H_1 and H_2, a measure D (H_1, H_2) should be selected to express the matching. The distance measure that has been used in the algorithm is Bhattacharyya distance, which is implemented by OpenCV:

$$\mathbf{D}_{\mathrm{Bhattacharyya}}(\mathbf{H_1}, \mathbf{H_2}) = \sqrt{\mathbf{1} - \sum_{\mathbf{i}} \frac{\sqrt{\mathbf{H_1(i)} \,.\, \mathbf{H_2(i)}}}{\sqrt{\sum_{\mathbf{i}} \mathbf{H_1(i)} . \sum_{\mathbf{i}} \mathbf{H_2(i)}}}}$$

3.2.1.3 Spatial Colour Histograms

The last image classification technique that has been employed is using spatial colour histograms. Spatial colour histograms can be defined as colour histograms that are extracted from local patches in the image in order to preserve spatial information. Spatial colour histograms have been reported as efficient and effective in object recognition. They have the ability to detect objects within an image and identify large number of background sub images as non-objects. Another advantage of spatial colour histograms is the low computation cost required to calculate local histograms (Zhang, Gao, Chen, & Zhao, 2006).

So, what we have done in our procedure is basically splitting an image into 25 equal blocks and extracting a local colour histogram from each block. Thereby, for a given image, 25 colour histograms are extracted. Two images are compared for similarity by comparing the pairing of histograms extracted from correspondent blocks.

To explain the exact implementation of this mechanism in more detail, let us have a look at how it works. First, an input image is equally split into 5 × 5 blocks named B_1, B_2, B_3, …., B_{25} as shown in figure 3.7.

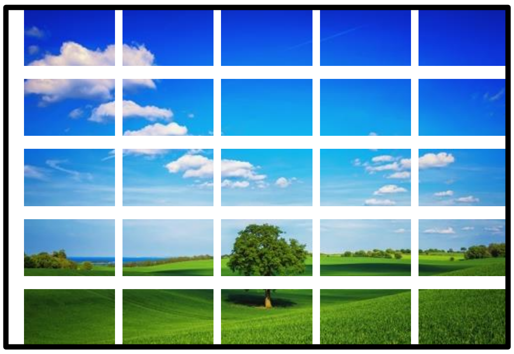


Figure 3.7: Image split into 25 Cells

Then, for each block B_i in the image, the colour histogram H_i is calculated making up a total of 25 colour histograms. Given two images A and B, A and B are matched by comparing H_{Ai} with H_{Bi} using the standard colour histogram comparison method explained in section 2.4.2, where H_{Ai} and H_{Bi} are the i^{th} colour histogram that have been extracted from A and B respectively.

EIMJ
المجلة الإلكترونية الشاملة متعددة التخصصات
العدد الرابع والثلاثون شهر (٣) ٢٠٢١

This yields a total of 25 comparison results obtained from each block pair. Suppose that D_i (H_{Ai}, H_{Bi}) is the distance calculated between the i^{th} histograms in A and B, Thus the total distance between two images is calculated as follows:

$$D_{total} = \sum_{i=1}^{25} D_i(H_{Ai}, H_{Bi})$$

To perform classification, the total distance between two images are compared, the shorter the distance, the greater the similarity.

3.2.2 Approaches to Model Construction

Image classification is the task of associating images with classes by comparing query images to a training dataset. In most image classification applications, the training dataset is randomly selected from standard image databases such as ImageNet, the Corel database and many more. In our project, the training dataset is automatically retrieved from Google Images using specific keyword searches. However, what is novel in our approach is that the training dataset is not chosen at random but rather images are automatically selected and models are constructed according to some image properties.

When images are searched in Google Images using a particular keyword, the top 40 images are retrieved and different models are constructed from that set. In our project, we have proposed three approaches of constructing models from a dataset: I) finding the most typical image. II) detecting image clusters III) forming a composite image. As three image classification techniques are adopted in our system, these ways of constructing models are applicable for each technique. To illustrate these approaches, we are going to show how models are constructed on “bank” dataset. The dataset is shown in figure 3.8.

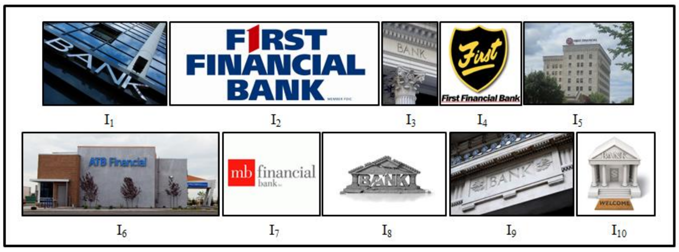


Figure 3.8: Bank Dataset Used to Construct Models

3.2.2.1 Finding Most Typical Image

The first approach that has been adopted in this project is finding the most typical image in the image dataset retrieved from Google Images for a given keyword. A typical image can be defined as the image that strongly represents the group selected from. This approach is included as a matter of interest since the training dataset should adequately represent the class that belongs to. So, the problem here is that when a keyword is looked up in Google Images, not all retrieved images are sensible.

"Meaningless" images and/or less representative images might be retrieved and this may cause confusion to image classification, especially if the query is, to a certain degree, "ambiguous". Thus, eliminating those kinds of images is a key step in our algorithm. The measure that has been used to identify such images is finding the image(s) that is more like what any other.

To find the most typical image in a set of 10 images, there is a need to compare each of them with each other and then score them. Comparing images will be done using SURF, colour histograms and spatial colour histogram.

Suppose that I_1, I_2, I_3, … I_{10} are the first ten images retrieved from Google Images. Each image I_i is compared to every image in the set yielding 55 comparisons. Thus, a similarity matrix M (i, j) is generated from the distance between two images as shown in table 3.1:

Table 3.1: Distance Matrix

Image	$\mathbf{I_1}$	$\mathbf{I_2}$	**…**	$\mathbf{I_9}$	$\mathbf{I_{10}}$	**Score**
$\mathbf{I_1}$	$D(I_1, I_1)$	$D(I_1, I_2)$	…	$D(I_1, I_9)$	$D(I_1, I_{10})$	S_1
$\mathbf{I_2}$	$D(I_2, I_1)$	$D(I_2, I_2)$	…	$D(I_2, I_9)$	$D(I_1, I_1)$	S_2
…	…	…	…	…	…	…
…	…	…	…	…	…	…
$\mathbf{I_9}$	$D(I_9, I_1)$	$D(I_9, I_2)$	…	$D(I_9, I_9)$	$D(I_9, I_{10})$	S_9
$\mathbf{I_{10}}$	$D(I_{10}, I_1)$	$D(I_{10}, I_2)$	…	$D(I_{10}, I_9)$	$D(I_{10}, I_{10})$	S_{10}

The distance here refers to the similarity measure that is used for SURF descriptors, colour histograms or spatial colour histograms.

By applying SURF algorithm the distance here refers to the number of matched keypoints where the larger the value the more the similarity and vice versa. For a given image I_i, the four most similar images are identified based on their scores with I_i and then the sum of their scores with I_i is calculated as follows:

$$\mathbf{S_i = D(I_i, I_x) + D\left(I_i, I_y\right) + D(I_i, I_z) + D(I_i, I_w)}$$

where I_x, I_y, I_z and I_w are the most four similar images to I_i.

Thus, each image achieves a final score that represents the degree of similarity to the four most similar images. This leads the image with the high score to be the most typical image as it has the least overall distance from the others. To perform image classification, query images are compared to the most typical image that has been selected.

To show how the most typical image is selected from the “bank” dataset, a 10 ×10 matrix is constructed as shown in table 3.2.

Table 3.2: Distance Matrix for Bank Dataset

Image	I_1	I_2	I_3	I_4	I_5	I_6	I_7	I_8	I_9	I_{10}	Score
I_1	1120	11	18	6	26	16	0	14	5	20	**80**
I_2	11	2669	447	15	128	362	5	3	0	5	**942**
I_3	18	447	2790	5	18	27	2	5	14	15	**510**
I_4	6	15	5	291	2	10	10	8	8	6	43
I_5	26	128	18	2	23155	112	1	1	4	1	**284**
I_6	16	352	27	10	112	9296	1	6	2	5	**507**
I_7	0	5	2	10	1	1	152	1	0	11	28
I_8	14	3	5	8	1	6	1	576	15	2	43
I_9	5	0	14	8	4	2	0	15	553	5	42
I_{10}	20	5	15	6	1	5	11	2	5	642	52

From the table, it is apparent that I_2 is selected to be the most representative image as it scores highly.

Regarding colour histograms and spatial colour histograms, the distance measure that is used is different from the one that is used for SURF descriptors, where the smaller the value the more the similarity and vice versa. Consequently, the image with low score is the image that is selected to be the most typical image.

3.2.2.2 Detecting Image Clusters

In the previous section, we have found the image that is most representative of the class. Now, we are going to find a group of images that form a good cluster of the class to which they belong. The idea here is to find four images that are thought to be similar to each other and strongly represent the class.

To find such a cluster, the similarity matrix M (i, j) is generated from the distance between two images, as discussed in the previous section. Again, the final score that represents the degree of similarity to the four most similar images is calculated for each image. Using SURF algorithm, the images that form a good cluster are the ones that score most highly while images that form a good cluster using colour histograms and spatial colour histograms are the ones that score low. Then, the top four images are selected, excluding the most typical image. These images are found to be the images that appear in lots of clusters, which are identified in section 3.3.2.1 when the final score is calculated.

By referring back to table 3.2, $I_{1,}$ $I_{3.}$ I_5 and I_6 are selected as image clusters. These images are shown in figure 3.9.

Figure 3.9: Image Clusters Selected from Bank Dataset

To perform image classification, query images are compared to the cluster that has been found. This will be accomplished by comparing the query image with every image that is a member of the cluster and the average of the four scores is calculated to be the overall score.

3.2.2.3 Forming Composite Model

This section describes another approach of constructing a model using the notion of composite model. What we mean by a composite model is the image that is obtained by means of averaging models belongs to certain class. The composite model is made out of the models that are most representative of the class to which they belong to or in other words the ones that form a good cluster. Thus, the composite model is thought to be a generalized model of the entire class. In our approach, SURF descriptors, colour histograms and spatial colour histograms are used to make the composite model.

The task of making a composite SURF descriptor is divided into two major steps: finding matched points and averaging descriptors. The procedure for forming a composite descriptor is explained further in the following. Suppose that I_1 and I_2 are two model that the composite model will be made out of. By applying SURF algorithm, a pairwise comparison of descriptors is made between I_1 and I_2 yielding a training-query pair of descriptors (D_{1i}, D_{2i}), where i is the number of detected keypoints of either I_1 or I_2. After that, the best matches are selected based on some conditions. Basically, the composite model is made out of descriptors that are involved in the best match while excluding others. Having found the descriptors that correspond to each other, now we can make a composite model by averaging these pairs of descriptors yielding a new feature vector that is believed to be a composite version of the two feature vectors.

To form a composite model out of a set of models using colour histograms, the average of the number of pixels (N) for each bin is calculated and the colour histogram is computed as follows:

$$\left[\frac{\mathbf{N_{11} + N_{12} + \cdots + N_{1m}}}{\mathbf{m}}, \frac{\mathbf{N_{21} + N_{22} + \cdots + N_{2m}}}{\mathbf{m}}, \frac{\mathbf{N_{n1} + N_{n2} + \cdots + N_{nm}}}{\mathbf{m}}\right]$$

where m and n indicate the number of images used and the number of bins respectively. The same process is done using spatial colour histograms and the average histogram is calculated for every corresponding cells. In our approach, the composite vector is made out of the four images that form a cluster (see section 3.3.2.3). This composite vector is then compared to the query images in order to perform image classification.

To have some confidence that the composite descriptor captures some features about images it is made of, a simple experiment has been carried out to see how the composite image scores on each of the images that made it up compared to other images.

The composite image is made out of I_2, I_4 and I_6. Then, it is compared to all 10 images, including the images that are used to form the composite. Table 3.3 shows the scores that have been achieved by comparing each image to the composite image.

Table 3.3: Scores Achieved by Comparing Each Image with the Composite Image

Image	$\mathbf{I_1}$	$\mathbf{I_2}$	$\mathbf{I_3}$	$\mathbf{I_4}$	$\mathbf{I_5}$	$\mathbf{I_6}$	$\mathbf{I_7}$	$\mathbf{I_8}$	$\mathbf{I_9}$	$\mathbf{I_{10}}$
Score	68	1645	86	1903	76	4543	48	41	43	59

Interestingly, we found that the images to which the composite is most similar are pretty much the images that it is made of.

3.2.3 Combining Classification Techniques

So far, we have developed three different techniques of comparing images: SURF, colour histograms and spatial colour histograms. In our project, we propose a novel approach which combines these techniques to perform image classification. What is of interest is that this approach balances between colour information and geometrical features. Three different methods for combining these techniques are proposed: average, degree of confidence and majority voting. These methods will be explained further herein.

- **Average**

Basically, after constructing a model for each class, a query image is compared to N models obtaining N scores, where N indicates the number of classes. Then, all scores are normalized to take values between 0 and 1, where 1 indicates the highest similarity. The above process is repeated using all three classification techniques described earlier: SURF, colour histograms and spatial colour histograms.

Suppose that $C_{1,}$ $C_{2,}$ …, C_n are the given classes and S_{1i}, S_{2i} and S_{3i} are the three scores that are achieved by comparing an image with a model of each class C_i using SURF, colour histograms and spatial colour histograms respectively. To obtain an overall score for the query image out of the three scores, the overall average A is calculated as follows:

$$\mathbf{A_i = \frac{S_{1i} + S_{2i} + S_{3i}}{3}}$$

The query image is said to belong to the class Ci if the overall average scores highly with that class.

- **Degree of Confidence**

Instead of comparing the average scores obtained by each technique, a confidence degree is calculated. The confidence degree refers to the estimated likelihood of the selected choice to be correct. In this method, each classification technique is applied independently and the query image is classified to either class. Having applied three techniques, different classification results are obtained. Thus, in order to choose between these results, confidence degree is obtained for each result.

(Alabbas, 2013)have invented this method for doing tagging in the study of “Textual Entailment for Modern Standard Arabic”, where there are three classifiers. They have found that actually taking the classifier that is most confidant produces better accuracy than taking anything else.

In our project, confidence degree is calculated by finding the ratio between the highest score and the average of the others. Suppose that a query image QI scores S_1, S_2, …, S_n (after normalization) when compared to models of classes C_1, C_2, …, C_n respectively where high score indicates high similarity. QI is said to belong to class C_i if Si is the highest score among other scores. Once the winning class has been selected, the degree of confidence is calculated as follows:

$$\mathbf{C = \frac{S_{max}}{\frac{(\sum_{i=1}^{n} S_i) - S_{max}}{n}}}$$

where S_{max} indicates the highest score. By calculating the confidence degree for each technique, the classification result that is selected is the one that has high degree of confidence.

- **Majority Voting**

The third method that is adopted to combine classification techniques is based on using majority voting, one of the simplest combining methods. If the two of the techniques agree on the classification result, then this result will be the choice that is made. However, if three disagree, then a back off strategy is applied by using either average method or degree of confidence method.

4. Experimental Results and Discussion

The aim of this section is to assess the image classification system that has been developed using a set of images from different classes. It contains a description of the experiments that have been conducted throughout the project and the results obtained. It starts with identifying the image dataset that has been used in evaluating the system. Then, it goes on to investigate various experiments along with the experimental results and an analysis of these results. The experiments are split into three categories: model construction, image comparison and classification and selecting optimal parameters and combining results. Experiments that have been conducted on image pre-processing are also demonstrated.

4.1 Image Dataset

Throughout the project, we did large number of experiments with different image datasets from different classes. However, the image database that is used in this experiment consists of 120 different colour images divided into three equal size classes: landscapes, faces and buildings. These images are retrieved automatically from Google Images using keyword searches with no restrictions specified such as camera type, resolution, brightness, etc. The images are of different sizes varying from 300 × 300 to 4000 × 4000 and are represented by 24-bits per pixel. No pre-processing was done on the data prior to experiments. This dataset has been used as a standard to evaluate the accuracy of our classification system. The dataset is also used to construct different kind of models.

4.2 Image Pre-processing

As mentioned in section 3, the K-Nearest Neighbour is the classifier that has been used in our project. However, we were experimenting with other classifiers and configurations of classifiers including Nearest Neighbour and FLANN based classifiers. One of the interesting experiments that have been conducted using Nearest Neighbour classifier is to try to reduce the number of SURF keypoints that will be used in the matching phase without affecting the accuracy.

In recent years, invariant features based image matching algorithms, such as SURF, have proven high performance. However, the execution time of such algorithms is relatively long. This is due to the fact that a large number of high dimensional descriptor vectors are compared in order to find similarity between features. Thus, to speed up image matching process, we investigate reducing the number of generated SURF features while the matching is still accurately performed, since the matching runtime is influenced by the number of key points.

a stabilization test has been done as follows. First, keypoints were extracted from the images and the matching was done using the original number of keypoints. Then, for all images, the number of extracted keypoints was reduced by a fixed number, say 10, and the process was repeated until finding a point at which the clusters that were formed by the matching stabilised.

To measure the stability of the clustering and find the minimum number of keypoints the process tends to be stable, an experiment has been conducted as follows. Given 10 images, suppose that N_1, N_2, …, N_{10} indicates the number of keypoints of the 10 images and N_i is the highest number of keypoints of any image in the set. Using the original number of keypoints without any reduction, scores that are achieved by comparing a given image to all other images were ranked, where the higher the score, the higher the rank. The number of keypoints that are above $N_i - 10$ was reduced to $N_i - 10$ for all images. Using the reduced number of keypoints, the comparison scores were ranked again. Then, a counter of how the rankings change from one stage to the next is kept. In other words, the number of images that were in the top five and are no longer in the top five after the reduction is counted. What we would like to find from this experiment was the number of keypoints at which ranking measure does not change. So, the process of choosing different number of keypoints continued until the minimum number of keypoints, which could be used with no change in the ranking, was reached.

The experiment has been conducted on 6 different set of images each set consists of 10 images and the threshold was set to 300. The most striking result to emerge from the data is that reducing the number of keypoints for all images in the dataset to be equal to the half of the largest number of keypoints leaves the ranking unchanged. Figure 4.1 below shows the change in ranking for all datasets when the number of key points is reduced.

Even though the pattern of results was interesting, using the K-Nearest Neighbour classifier rather than other classifiers gives even better results in addition to the fact that the K-Nearest Neighbour classifier is more robust than the Nearest Neighbour classifier.

We have also conducted some experiments on reducing the image size to reach a specific number of keypoints with the hope of speeding up the algorithm. Unfortunately, what we found is that for different images, similar amount of reduction leads to different outcomes. We have also found that the first N keypoints of the original image are different from the first N keypoint of the reduced image. These two findings have led us to preserve the original size of all images.

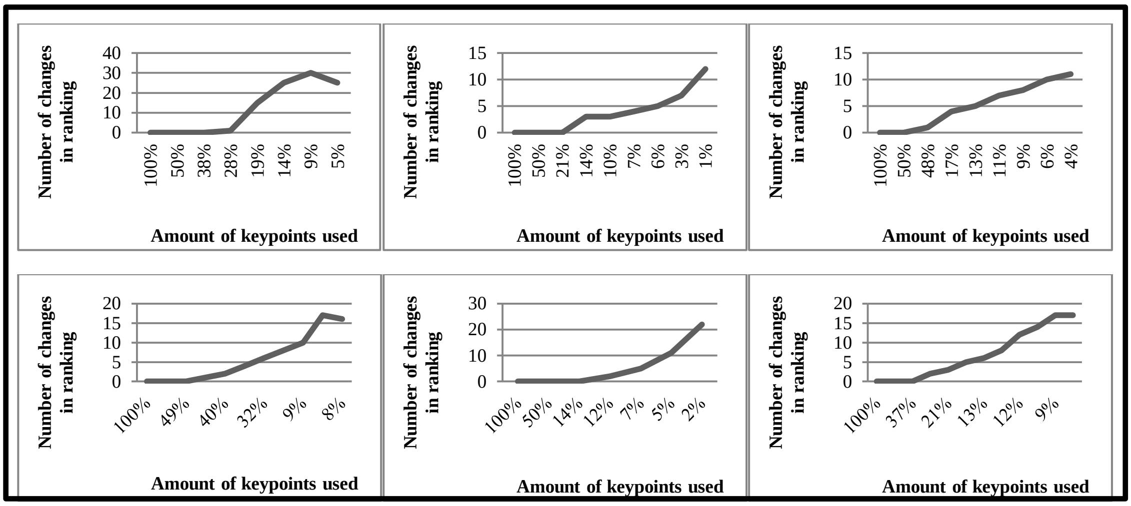


Figure 4.1: Effect of Reducing Key points on Image Ranking

4.3 Experiments and Results Analysis

As mentioned, we evaluate our approach of image classification using a set of images retrieved automatically from Google Images. Thus, classes that are assigned to these images are based on keyword searches. In order to further show how performance of image classification is improved, we compare our approach with the original SURF.

4.3.1 Model Construction

In our project, we propose three approaches of constructing models namely most typical image, clusters and composite image. These models are then used to classify query images. Models are constructed using three image classification techniques: SURF, colour histograms and spatial colour histograms. For each technique, what we are going to do is evaluate each kind of model on the image dataset in order to select the model that gives more accurate classification results when compared to query images. The goal from this experiment is to find models that are sort of representing a class they belong to.

For each class, models are constructed using SURF, colour histograms and spatial colour histograms. To illustrate the experiment, we are going to show how different kinds of models are constructed on landscape images which are similarly constructed for face and building images. From each class, 40 images were used for model construction. Now, let us have a look at how these models are constructed using each technique.

4.3.1.1 Constructing Models Using SURF

SURF is the first image classification technique that has been used in our approach. The threshold was set at 300 and the classifier that has been used is the K-Nearest Neighbour classifier with K = 2. To construct a model, 40 images were compared to each other (using SURF algorithm), yielding a 40 × 40 matrix of scores which indicate the number of matched points between two images.

- **Find Most Typical Image**

The first approach of constructing a model is finding the most typical image that represents the class it belongs to. To find the most typical image, the top four most similar images to every image in the set were identified. Then, the sum of the matched points of these images was calculated for every image. The measure that is used to identify the most typical image is finding the image that is most similar to the images which is similar to. Thus, the image that achieves the highest score was selected as the most typical image.

As shown in table 4.1, I_{25} was the image that has been selected to be the most typical image. This image is then used to be compared with query images.

- **Detect Image Clusters**

Image clusters is the second model that was constructed in our project. By referring back to the matrix constructed earlier using SURF algorithm in table 4.1, image clusters are detected by finding the 4 images that achieve highest scores apart from the typical image. Thus, images I_{13}, I_{28}, I_{32} and I_{36} form a cluster which will be then used to classify query images.

- **Composite Model**

The last model that was constructed in our project is the composite model. The composite model was made out of the models that formed a cluster. The process of making a composite model using SURF has gone through three stages. First, the average of descriptor vectors of matched points between two models was computed yielding average descriptor vectors (ADVs). These ADVs were then matched with the third model in order to identify only matched points between the model and ADVs. Finally, the average of the descriptor vectors of matched points between the third model and the ADVs was calculated and the same process was repeated again with the fourth model making up overall ADVs of the four models. The resulted overall ADVs were then used classify query images, which will be described later in this section. Forming average descriptor vectors is illustrated in figure 4.2.

Image 1
Descriptors of Matched Keypoints

$$\left\{\begin{array}{c} [X1_{1,1} \;\; X1_{1,2} \;\; X1_{1,3} \; \dots \; X1_{1,64}] \\ [X1_{2,1} \;\; X1_{2,2} \;\; X1_{2,3} \; \dots \; X1_{2,64}] \\ [X1_{3,1} \;\; X1_{3,2} \;\; X1_{3,3} \; \dots \; X1_{3,64}] \\ [X1_{n,1} \;\; X1_{n,2} \;\; X1_{n,3} \; \dots \; X1_{n,64}] \end{array}\right\}$$

Image 2
Descriptors of Matched Keypoints

$$\left\{\begin{array}{c} [X2_{1,1} \;\; X2_{1,2} \;\; X2_{1,3} \; \dots \; X2_{1,64}] \\ [X2_{2,1} \;\; X2_{2,2} \;\; X2_{2,3} \; \dots \; X2_{2,64}] \\ [X2_{3,1} \;\; X2_{3,2} \;\; X2_{3,3} \; \dots \; X2_{3,64}] \\ [X2_{n,1} \;\; X2_{n,2} \;\; X2_{n,3} \; \dots \; X2_{n,64}] \end{array}\right\}$$

Average Descriptors Vectors of
Image 1 and Image 2

$$\left\{\begin{array}{c} \left[\frac{X1_{1,1}+X2_{1,1}}{2} \;\; \frac{X1_{1,2}+X2_{1,2}}{2} \;\; \frac{X1_{1,3}+X2_{1,3}}{2} \; \dots \; \frac{X1_{1,64}+X2_{1,64}}{2}\right] \\ \left[\frac{X1_{2,1}+X2_{2,1}}{2} \;\; \frac{X1_{2,2}+X2_{2,2}}{2} \;\; \frac{X1_{2,3}+X2_{2,3}}{2} \; \dots \; \frac{X1_{2,64}+X2_{2,64}}{2}\right] \\ \left[\frac{X1_{3,1}+X2_{3,1}}{2} \;\; \frac{X1_{3,2}+X2_{3,2}}{2} \;\; \frac{X1_{3,3}+X2_{3,3}}{2} \; \dots \; \frac{X1_{3,64}+X2_{3,64}}{2}\right] \\ \left[\frac{X1_{n,1}+X2_{n,1}}{2} \;\; \frac{X1_{n,2}+X2_{n,2}}{2} \;\; \frac{X1_{n,3}+X2_{n,3}}{2} \; \dots \; \frac{X1_{n,64}+X2_{n,64}}{2}\right] \end{array}\right\}$$

Figure 4.2: Forming Average Descriptor

4.3.1.2 Constructing Models Using Colour Histograms

The second image classification technique that has been used in our project is colour histograms. Colour histograms of hue channels were calculated for images in the Hue-Saturation-Value (HSV) colour space. Bhattacharyya distance was used to compare two histograms. In the following, we demonstrate the construction of the three models using colour histograms. To construct a model, 40 images were compared to each other (using colour histograms), yielding a 40×40 matrix of scores which indicate the distance between two colour histograms.

- **Find Most Typical image**

To find the most typical image using colour histograms, the top four closest images to every image in the set were identified. Then, the sum of the distances of these images was calculated. The image that achieved the lowest score was selected as the most typical image.

As shown in table 4.2, I_{25} is the image that has been chosen to be the most typical image. This image is then used to be compared with query images.

- **Detect Image Clusters**

By referring back to the matrix constructed earlier using colour histograms, image clusters are detected by finding the 4 images that achieve lowest scores apart from the typical image. Thus, images I_{18} I_{28} I_{32} and I_{36} formed a cluster which will be then used to classify query images.

- **Composite Image**

The composite image was made out of the images that formed a cluster. Making a composite image using colour histograms is straightforward. Simply, for all 4 images, average counts of the number of pixels that fall in each bin was calculated resulting in an average histogram of the hue channels. This average histogram is then compared to query images histograms in order to perform image classification.

4.3.1.3 Constructing Models Using Spatial Colour Histograms

Spatial colour histogram is the last image classification technique that has been used in our project. What is different between colour histograms and spatial colour histograms is that the former are calculated for the whole image while the latter are calculated for image regions. In our approach, we split the image into 25 equal-size cells and we calculated the colour histogram for each cell. Colour histograms of hue channels were calculated in the Hue-Saturation-Value (HSV) colour space. Bhattacharyya distance was used to compare two spatial colour histograms. To compare two images using spatial colour histograms, every spatial colour histogram from one image was compared to the corresponding spatial colour histogram from the other image yielding 25 different distances which were then added to form an overall distance.

The construction of the three models using colour histograms is illustrated in the following. 40 images were compared to each other (using spatial colour histograms), yielding a 40×40 matrix of scores which indicate the overall distance between spatial colour histograms of two images.

- **Find Most Typical image**

Similar to what has been done previously using colour histograms, to find the most typical image using spatial colour histograms, the top four closest images to every image in the set were identified. Then, the sum of the distances of these images was calculated. The image that achieved the lowest score was selected as the most typical image.

Table 4.3 show that I_{25} is the image that has been chosen to be the most typical image which is then used to be compared with query images.

- **Detect Image Clusters**

Image clusters were detected on the same way as we did using colour histograms. The 4 images that achieved lowest scores apart from the typical image were identified. Thus, images I_{16}, I_{28}, I_{32} and I_{36} formed a cluster that will be then used to classify query images.

- **Composite Image**

The composite image was made out of the images that formed a cluster. To make a composite image using spatial colour histograms, averages of spatial colour histograms were calculated from the four images yielding 25 average histograms of the hue channels. To classify images, these average histograms are then compared to query image spatial histograms.

4.3.2 Image Comparison and Classification

Evaluation of classification performance is an important step in the task of image classification. The previous section has demonstrated the process of making different kinds of models of landscape images using three different image classification techniques. The process was repeated on face and building images. The main goal of that experiment was to examine every model of every technique on a set of images and evaluate the performance of image classification of such models to determine their speed and accuracy.

In this section, what we are going to do is, for each technique, compare every image in the dataset to a kind of model from each class. A given query image is said to belong to class C_i if its distance to a model of class C_i is smaller than its distance to models of either classes. The process is then repeated to all kinds of model. The following three subsections will provide an evaluation of all models from each technique.

4.3.2.1 Classification Using SURF Models

As has been demonstrated in the previous section, three different kinds of models have been constructed using SURF algorithm. Now, what we are going to do is evaluating these kinds of models on our image dataset by comparing each kind of model from each class to every image using SURF algorithm. An image is said to belong to a class if the number of matched points between the image and a model of that class is larger than number of matched points between the image and models of the other classes.

To illustrate the classification performance of model 1 (most typical image), every query image was compared to three images: the most typical image of landscape set, face set and building set, selected using SURF algorithm, and the image was classified to the closest class. Classification performance results in an accuracy of 36.67%. However, classification results revealed that 95% of landscape and face images were classified as buildings. This result may be explained by the fact that images with higher number of keypoints achieve higher scores than others. When we checked the number of keypoints of all three models (landscape, face and building), we found that “landscape model” has the highest number of keypoints. Disappointingly, the normalisation process that was done earlier does not pay off as much as would expected using the K-Nearest Neighbour classifier. Figure 4.3 compares the experimental data on model 1 using SURF algorithm.

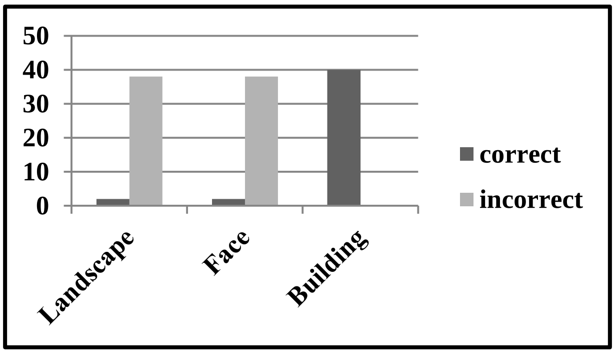


Figure 4.3: Experimental Data on Model 1 Using SURF

To show how our classification system performs when model 2 (image clusters) is used, every query image was compared to a cluster from each class i.e. compared to the four images that form a cluster using SURF algorithm. Once an image is compared to a cluster, average of the 4 scores that were achieved from the comparisons was calculated to be the overall score. Since query images are compared to a cluster of landscapes, a cluster of faces and a cluster of buildings, the query image was classified to the class that its cluster achieved the highest average score.

Classification performance results in an accuracy of 62.5%. Figure 4.5 compares the experimental data on model 2 using SURF algorithm. Data from this figure can be compared with the data in figure 4.4 which shows an increase in accuracy of 25.8%.

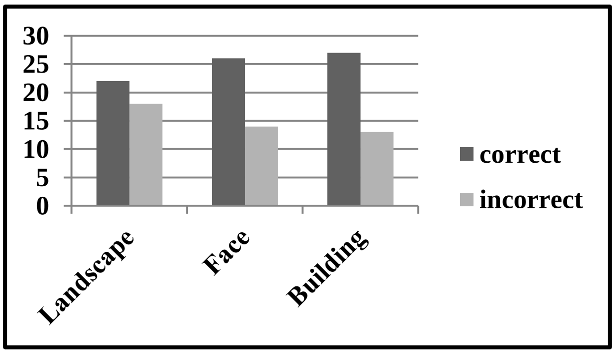


Figure 4.4: Experimental Data on Model 2 Using SURF

To classify query images using model 3 (composite images), every query image was compared to a composite image from each class using SURF algorithm. This has been done by matching query image descriptors to composite descriptors that were formed previously. Finally, the query image was classified to the class that its composite image achieved the highest score when compared to the query image.

Classification performance results in an accuracy of 60.8%. Figure 4.5 compares the experimental data on model 3 using SURF algorithm. It is apparent from the figure that model 3 performs roughly as well as model 2 with a decrease of about 1.67%.

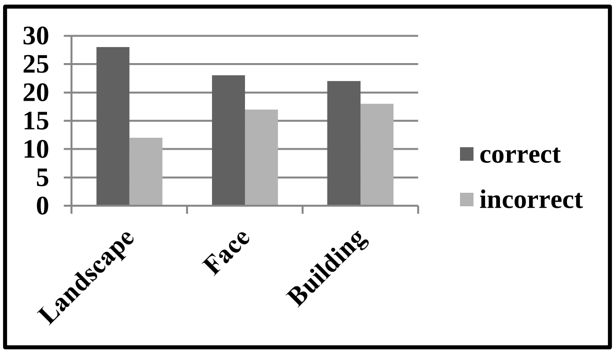


Figure 4.5: Experimental Data on Model 3 Using SURF

4.3.2.2 Classification Using Colour Histograms Models

In the previous section, we have constructed three kinds of models using colour histograms. Now, what we are going to do is to compare each kind of model from each class with every image using colour histograms. An image is said to belong to a class if the distance between the image and a model of that class is shorter than the distance between the image and models of the other classes.

To illustrate the classification performance of model 1 (most typical image), every query image was compared to three images: the most typical image of landscape set, face set and building set, selected using colour histograms, and the image was classified to the closest class.

Classification performance results in an accuracy of 63.3%. However, classification results revealed that 85% of landscape images were classified as buildings. It seems possible that these results are due to the similarity of colours in particular regions, as both are outdoor images. About 95% of face and building images were classified correctly. Figure 4.6 compares the experimental data on model 1 using colour histograms.

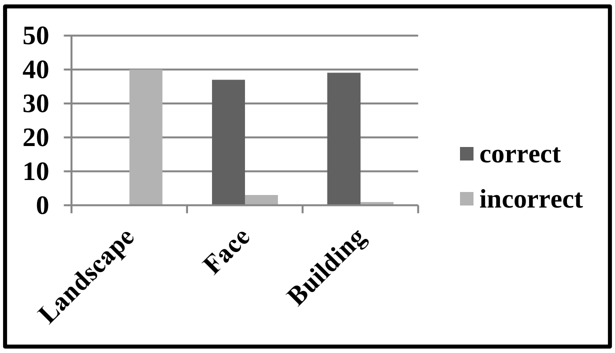


Figure 4.6: Experimental Data on Model 1 Using Colour Histograms

To show how our classification system performs when model 2 (image clusters) is used, every query image was compared to a cluster from each class i.e. compared to the four images that form a cluster using colour histograms. Once an image is compared to a cluster, average of the 4 scores that were achieved from the comparisons was calculated to be the overall score. As colour histograms were compared using Bhattacharyya distance, the query image was classified to the class that its cluster achieved the lowest average score.

Classification performance results in an accuracy of 80.8%. Half of landscape images were classified correctly while face and building images result in accuracy of 96.25%. Figure 4.7 compares the experimental data on model 2 using colour histograms. Data from this figure shows an increasing in accuracy of 18.33% compared to experimental results on model 1.

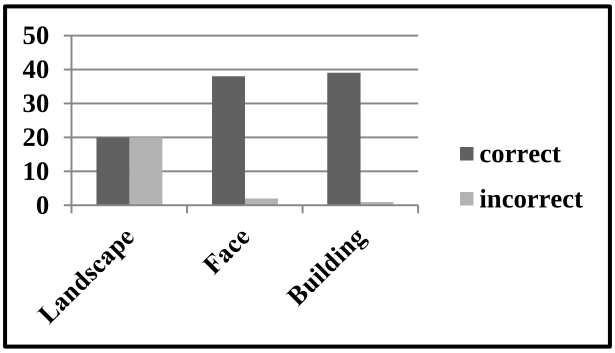


Figure 4.7: Experimental Data on Model 2 Using Colour Histograms

To classify query images using model 3 (composite images), every query image was compared to a composite image from each class using colour histograms. This has been done by comparing a query image histogram to the composite histogram that was calculated previously. Finally, the query image was classified to the class that its composite image achieved the lowest score when compared to the query image.

Classification performance results in an accuracy of 81.67%. About 40% of landscape images were incorrectly classified. The reason for misclassification of some landscape images might be related to the image colours and brightness as some of those images were dark while others contained lots of purple areas which are unusual in landscape images. Figure 4.8 compares the experimental data on model 3 using colour histograms. It is apparent from the figure that model 3 performs roughly as well as model 2 with an increase of about 0.87%.

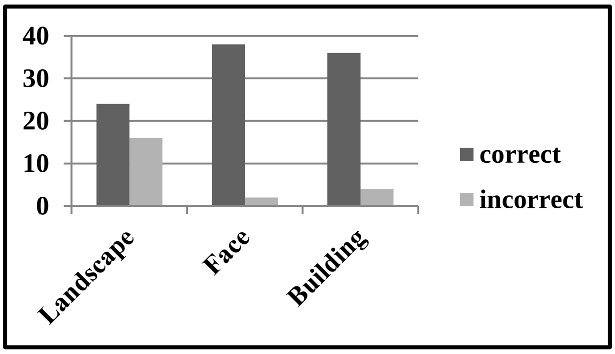


Figure 4.8: Experimental Data on Model 3 Using Colour Histograms

4.3.2.3 Classification Using Spatial Colour Histograms Models

To evaluate the three kinds of models using spatial colour histograms, each kind of model from each class was compared to every image using spatial colour histograms. An image is said to belong to a class if the overall distance between the image and a model of that class is shorter than the overall distance between the image and models of the other classes.

To illustrate the classification performance of model 1 (most typical image), every query image was compared to three images: the most typical image of landscape set, face set and building set, selected using spatial colour histograms, and the image was classified to the closest class.

Classification performance results in an accuracy of 64.17%. However, classification results revealed that 80% of landscape images were classified as buildings. Again, these results could be attributed to the similarity of colours in particular regions, as both are outdoor images. About 96.25% of face and building images were classified correctly. Figure 4.9 compares the experimental data on model 1 using spatial colour histograms.

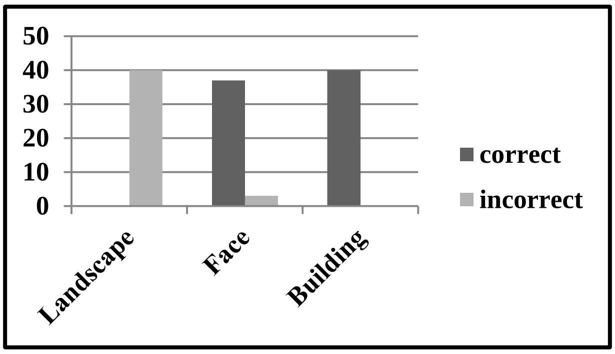


Figure 4.9: Experimental Data on Model 1 Using Spatial Colour Histograms

To show how our classification system performs when model 2 (image clusters) is used, every query image was compared to a cluster from each class i.e. compared to the four images that form a cluster using spatial colour histograms. Once an image is compared to a cluster, average of the 4 scores that were achieved from the comparisons was calculated to be the overall score. As spatial colour histograms of image cells were compared using Bhattacharyya distance, the query image was classified to the class that its cluster achieved the lowest average score.

Classification performance results in an accuracy of 75.83% with more than half of landscape images were classified as buildings. Face and building images result in accuracy of 95%. Figure 4.10 compares the experimental data on model 2 using spatial colour histograms. Data from this figure shows an increasing in accuracy of 11.67% compared to experimental results on model 1.

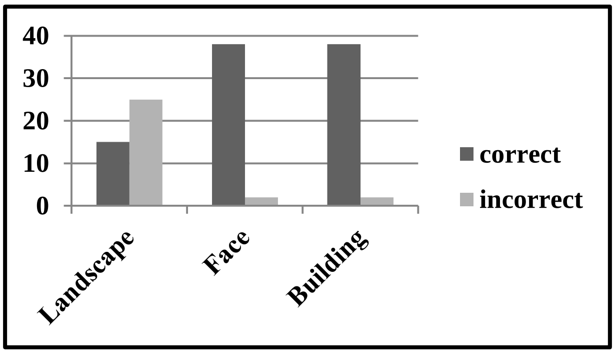


Figure 4.10: Experimental Data on Model 2 Using Spatial Colour Histograms

To classify query images using model 3 (composite images), every query image was compared to a composite image from each class using spatial colour histograms. This has been done by comparing query image spatial histograms to the composite spatial histograms that were calculated previously. Finally, the query image was classified to the class that its composite image achieved the lowest score when compared to the query image.

Classification performance results in an accuracy of 63.33%. About 86.25% of face and building images were classified correctly. Figure 4.11 compares the experimental data on model 3 using spatial colour histograms. It is apparent from the figure that model 3 performs roughly as well as model 1 with a decrease of about 0.83%. The most striking result to emerge from the data is that model 3 using spatial colour histograms approach does not perform as well as model 3 using SURF algorithm and colour histograms. It is difficult to explain this result, but it might be related to the fact that the spatial colour histogram is calculated by averaging the number of pixels for a region in each image. Thus, it might do not have sufficient discrimination power since it is entirely spatial and restricted to that region.

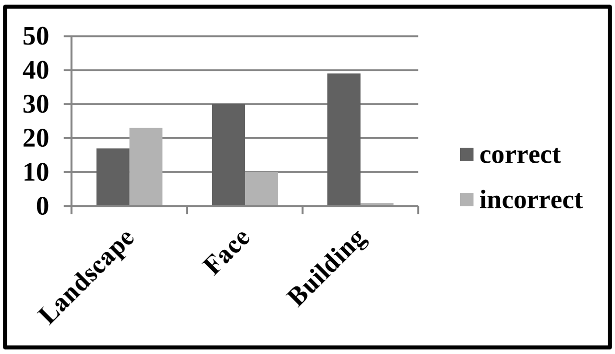


Figure 4.11: Experimental Data on Model 3 Using Spatial Colour Histograms

4.3.3 Selecting Optimal Parameters & Combining Results

The previous section has compared the performance of three different models using three different classification techniques. The correlation between each model and its classification performance was tested and the results obtained from the preliminary analysis of these models are summarised in figure 4.12.

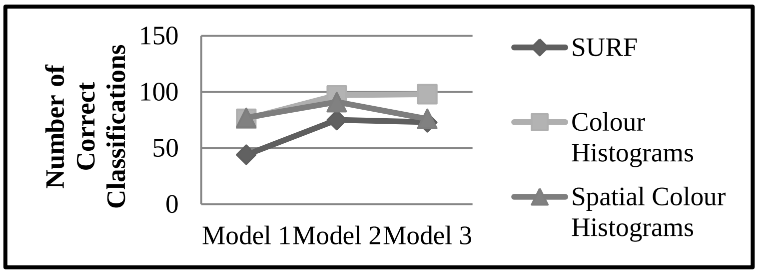


Figure 4.12: Classification Results of Each Technique

From this data, we can see that model 1 resulted in the lowest value of accurate classifications. Using SURF and colour histogram, there is a clear trend of increasing in classification accuracy when model 2 and model 3 were used, which both behave quite similarly. On average, model 2 and model 3 were shown to have the optimal performance for SURF and colour histograms techniques. Hence, it could conceivably be hypothesised that for these techniques, matching query images with either image clusters or composite images increases the accuracy of the classification by about 21.5%. Regarding spatial colour histograms, model 2 has shown optimal performance compared to model 1 and model 3.

In our project, we proposed a new classification approach that combines colour features along with geometry features using SURF algorithm, colour histograms and spatial colour histograms. Since model 2 and model 3 perform quite similarly using SURF and colour histograms, model 3 has been selected as it makes classification faster (comparing to a single image is faster than comparing to 4 images). Regarding spatial colour histograms, model 2 has been selected since it has the optimal performance.

After we have done a set of experiments and find the optimal parameters, we have tested our classification algorithm on the dataset which consists of 120 images from different classes. Then, we have applied the three methods of combining classification scores achieved from each technique after normalising them. Table 4.4 shows the classification results for the proposed classification algorithm.

Table 4.4: Classification Results for the Proposed Classification Algorithm

Method of Combination	Accuracy
Average	80.83%
Degree of Confidence	75.83%
Majority Voting	85.83%

From table 4.4, we find that using the majority voting procedure to combine scores results in the best accuracy (85.83%). In order to further show how our approach performs, we compare it with the original SURF. We have found that our approach shows an increase of classification accuracy by about 31.67%.

5. Conclusion

The main goal of this project was to develop and assess our new approach to image classification that balances between colour information and geometrical features using a combination of different image classification techniques. A number of objectives were identified in order to achieve this general goal. The literature review was carried out to satisfy the first objective, which was acquiring a deep understanding of the current image classification techniques and efforts that has been made recently to improve image classification. We gave an overview of the basics of image classification by demonstrating the importance of image classification and addressing its major challenges. Then, we identified the position of image classification within various computer vision applications. Classification techniques that are currently used were investigated including Speeded Up Robust Features (SURF) and colour histograms along with the working mechanism of both. Finally, a discussion of the related work that has been carried out in this area was provided.

In addition to the theoretical research, a large part of the practical work was to experiment with a number of image classification techniques and construct different models for image comparison. Three kinds of models, including the most typical images, image clusters and composite images, were constructed using three different classification techniques. To develop software that satisfies the proposed objectives, prototyping development methodology was used. Thus, in order to enhance the system and obtain optimal results, different models were designed, implemented and tested iteratively.

Our approach was evaluated using a dataset of 120 images of landscapes, faces and buildings retrieved automatically from Google Images. Classification results have shown that colour features have more discrimination power than geometrical information for the classification problem considered in this study. Results have also shown that image clusters and composite models perform quite similarly when SURF and colour histograms are used. To speed up the execution time, composite models have been chosen as the optimal model since comparing with composites is four times quicker than comparing with image clusters. These composite models that have been formed from image clusters using SURF and colour histograms yielded good results with an accuracy of about 60.8% and 81.67% respectively, while image clusters model using spatial colour histograms yielded the best results with an accuracy of 75.83%.

Different classification scores resulted from different techniques were combined using three procedures: average, degree of confidence and majority voting. We have found that the majority voting was the best and resulted in an accuracy of 85.83. Evaluation results of our classification scheme has shown that combining colour histograms with SURF descriptors improves classification accuracy by about 31.67% when compared with the original SURF descriptors. The proposed classification approach assumes that the dataset consists of only images that fall into these three categories and the classifier classifies query images into landscapes, faces or buildings and does not reject any image that does not belong to either categories.

The major limitation of this study was the long execution time required to perform image matching using SURF algorithm. This has restricted us to reduce the sample size that has been used. In the future, it is **recommended** that further research be undertaken in the following directions.

First, it would be interesting to assess our image classification approach with larger test samples and conduct similar experiments with more image classes. Further research might investigate SURF-based image matching in order to identify significant keypoints that contribute more to the matching since research in this field would be of great help in increasing the runtime efficiency of image matching. Considerably, more work will need to be done to focus on execution time of our approach to be applied in real time applications.

Finally, future efforts could be directed towards experimenting further image classification techniques.

References

Alabbas, M. (2013). *Textual Entailment for Modern Standard Arabic.* The University of Manchester.

Bay, H., Tuytelaars, T., & Gool, L. (2006). SURF: Speeded Up Robust Features. *Springer Berlin Heidelberg*, 404-417.

Chapelle, O., Haffner, P., & Vapnik , V. (1999). Support Vector Machines forHistogram-Based Image Classification. *10*(5), 1055-1064.

Fan, P., Men, A., Chen, M., & Yang, B. (2009). Color-SURF: A surf descriptor with local kernel color histograms. *IEEE*, 726-730.

Khan, N., McCane, B., & Wyvill, G. (2011). SIFT and SURF Performance Evaluation against Various Image Deformations on Benchmark Dataset. *IEEE*, 501-506.

Lowe, D. (1999). Object recognition from local scale-invariant features. *The proceedings of the seventh IEEE international conference , 2*, 1150-1157.

OpenCV. (2014). *About*. Retrieved July 17, 2014, from http://opencv.org/about.html

OpenCV. (2014). *Histograms*. Retrieved July 17, 2014, from http://docs.opencv.org/trunk/doc/py_tutorials/py_imgproc/py_histograms/py_histogram_begins/py_histogram_begins.html

OpenCV. (2014). *Introduction to SURF (Speeded-Up Robust Features)*. Retrieved July 17, 2014, from http://docs.opencv.org/trunk/doc/py_tutorials/py_feature2d/py_surf_intro/py_surf_intro.html

Sergyan, S. (2008). Color Histogram Features Based Image Classification in Content-Based Image Retrieval Systems. *IEEE*, 221-224.

Shukla, T., Mishra, N., & Sharma, S. (2013). Automatic Image Annotation using SURF Features. *International Journal of Computer Applications, 68*(4), 17-24.

Szummer, M., & Picard, R. (1998). Indoor and Outdoor Image Classication. *IEEE*, 42-51.

Vailaya, A., Jain, A., & Zhang, H. (1998). On Image Classification: City Images vs. Landscapes. *Pattern Recognition, 31*(12), 1921-1935.

Vailaya, A., & Zhang, H. (2001). Image classification for content-based indexing. *IEEE, 10*(1), 117-130.

Zhang, H., Gao, W., Chen, X., & Zhao, D. (2006). Object detection using spatial histogram features. *Image and Vision Computing , 24*(4), 327-341.