Image Retrieval Using Speeded Up Robust Feature: An Effort to Improvement

SADGI SHARMA Computer Science and Engineering

ABDUL MOIZ SIDDIQUI Computer Science and Engineering

ABSTRACT

Due to tremendous technological advancements, need of image information systems has became an important issue, since visual media requires large amounts of memory and computing power for processing and storage, there is a need to efficiently index and retrieve visual information from image database. In recent years, the digital document image has become an important means of enhancing information management. Content-Based Image Retrieval (CBIR) is a challenging task which retrieves the similar images from the large database. Most of the CBIR system uses the low-level features such as color, texture and shape to extract the features from the images. In so many works available, interest points are used to extract the similar images with different view and accuracy. In this paper, the same is tried to retrieve with the use of SURF and fed into Support Vector Machine (SVM) for further classification. SURF is fast and robust interest points detector which is used in many computer vision applications and other methodologies available for the same have been discussed, followed by an experimental simulation and evaluation over a set of test images in MATLAB simulation environment.

Keywords

Content based Image Retrieval (CBIR), Speed up Robust feature (SURF), image databases

1. INTRODUCTION

In recent years, very large collections of images and videos have grown rapidly. In parallel with this growth, content based retrieval and querying the indexed collections are required to access visual information. Two of the main components of the visual information are texture and color. The history of the content-based image retrieval can be divided into three phases:

- The retrieval based on artificial notes.
- The retrieval based on vision character of image contents.
- The retrieval based on image semantic features.

The image retrieval that is based on artificial notes labels images by using text firstly, in fact it has already changed image retrieval into traditional keywords retrieval. Problem with the approach is that, it brings heavy workload and on the other hand, it still remains subjectivity and uncertainty. Because the image retrieval that is based on artificial notes still remains insufficiency, the farther study that adapts vision image features has been come up and become the main study. The character of this method is image feature extraction impersonally, whether the retrieval is good or not depends on the accuracy of the features extraction. So the research based on vision features is becoming the focus in the academic community. The feature of vision can be classified by semantic hierarchy into middle level feature and low- level feature. Low-level feature includes color, texture and inflexion. Middle level involves shape description and object feature[1].

Content based Image Retrieval systems try to retrieve images similar to a user-defined specification or pattern (e.g., shape sketch, image example). Their goal is to support image retrieval based on content properties (e.g., shape, color, texture), usually encoded into feature vectors. One of the main advantages of the CBIR approach is the possibility of an automatic retrieval process, instead of the traditional keyword-based approach, which usually requires very laborious and time-consuming previous annotation of database images.[2]

2. SPPED UP ROBUST FEATURE (SURF):

SURF (Speeded Up Robust Features) is a robust local feature detector, first presented by Herbert Bay et al. in 2006, that can be used in computer vision tasks like object recognition or 3D reconstruction. It is partly inspired by the SIFT descriptor. The standard version of SURF is several times faster than SIFT and claimed by its authors to be more robust against different image transformations than SIFT. SURF is based on sums of 2D Haar wavelet responses and makes an efficient use of integral images.

It uses an integer approximation to the determinant of Hessian blob detector, which can be computed extremely quickly with an integral image (3 integer operations). For features, it uses the sum of the Haar wavelet response around the point of interest. Again, these can be computed with the aid of the integral image.

SURF used in this approach to extract relevant features and descriptors from images. This approach is preferred over its predecessor due to its succinct descriptor length i.e.

Manuscript received November 5, 2014
Manuscript revised November 20, 2014

64 floating point values. In SURF, a descriptor vector of length 64 is constructed using a histogram of gradient orientations in the local neighborhood around each key point.[15]

3. IMAGE RETRIEVAL: LITERATURE REVIEW

In this section we try to subjectively identify some of the current trends in the research for image retrieval systems. A common ground in most of current IR systems is to exploit low-level features such as color, texture and shape, which can be extracted by a machine automatically. While semantic level retrieval would be more desirable for users [6], given the current state of technology in image understanding, this is still very difficult to achieve. This is especially true when one has to deal with a heterogeneous and unpredictable image collection.

Thus methods inspired by artificial intelligence [7], textual retrieval [8, 9], and psychology & human-computer interaction [10, 11], are starting to influence the research. Synthetically, image retrieval starts off by the design of a robust, meaningful and flexible feature set to characterize all plausible images in the collection. Then clever manipulation of the features tries to uncover some higher level similarity between the query and the database candidates. An interactive, iterative, and user-oriented query process then improves on the raw results.

Early IR systems [12, 13, 14] mainly relied on a global feature set extracted from images. For instance, color features are commonly represented by a global histogram. This provides a very simple and efficient representation of images for the retrieval purpose. However, the main drawback with this type of systems is that they have neglected spatial information. More recent on systems have addressed this problem. Spatial information is either expressed explicitly by the segmented image regions or implicitly via dominant wavelet coefficients.

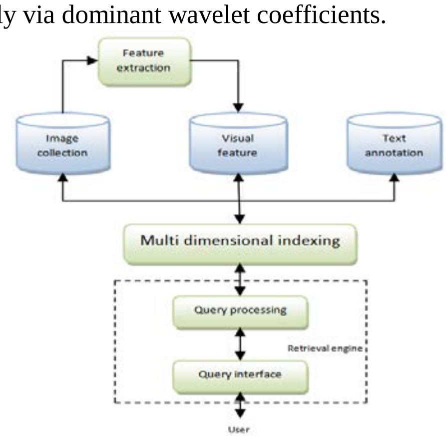

Figure 1: Architecture of basic Image retrieval system

Most systems use the query by example approach, where the user selects one or several images, and the system returns the similar ones. Other uses the way of querying the image database based on content. Other few special systems allow the user to specify spatial constraints on the dominant objects.

All of these methods suffer somewhat from the drawback that the system relies on the users abilities and does not adapt to his/her needs.

In a research work carried out in [16] the researcher makes a study on content-based image retrieval algorithm for document image database. Given a query image the system returns overall similar images in database. For document images, we propose the algorithm based on hierarchical matching tree. First segment an image into several regions with paragraph marking based on paragraph height estimation, and then segment the region into line blocks, the algorithm for document image retrieval by regions and line blocks with hierarchical matching tree is presented.

Content in an image can be described by using either its semantic or its visual information in terms of

features. Labelling an image using semantic information will depend on the viewer and is time consuming. One solution is to retrieve images based on features suppose to represent the visual content in the image. Visual content is described, in the simplest form, by low level features like color, shape, texture, etc extracted from the image. Color features include color histogram, color moments, color sets (ex: HSV). Shape features are boundary based and region based and could be local or global. Texture features are derived from texture patterns present in image; could be structural or probabilistic or spectral [17].

Effectiveness of content-based image retrieval (CBIR) system depends on the choice of the set of visual features and on the choice of the similarity metric that models the user perception of similarity.

Eakins mentioned three levels of queries in CBIR [18]

• Level 1: Retrieval by primitive features such as color, texture, shape or the spatial location of image elements. Typical query is query by example, 'find pictures like this'.

• Level 2: Retrieval of objects of given type identified by derived features, with some degree of logical inference. For example, 'find a picture of a flower'.

• Level 3: Retrieval by abstract attributes, involving a significant amount of high-level reasoning about the purpose of the objects or scenes depicted. This includes retrieval of named events, of pictures with emotional or religious significance, etc. Query example, 'find pictures of a joyful crowd'.

Levels 2 and 3 together are referred to as semantic image retrieval, and the gap between Levels 1 and 2/3 as the semantic gap. Some challenges in learning semantics in CBIR are:

• Semantic gap characterization

• Huge amount of objects to search among.
• Incomplete query specification.
• Incomplete image description.

Some state-of-the-art techniques in reducing the 'semantic gap' [17]:

1) Using object ontology to define high-level concepts and generating Semantic Template (ST) to support high level image retrieval.
2) Using machine learning tools to associate low level features with query concepts.
3) Introducing Relevance Feedback (RF) into retrieval loop for continuous learning of users' intention.
4) Making use of both the visual content of images and the textual information obtained from the Web for WWW (the Web) image retrieval.

In another piece of research in [19] a novel approach for content based color image classification using Support Vector Machine (SVM).In this approach, color image classification is done on features extracted from histograms of color components. The benefit of using color image histograms are better efficiency, and insensitivity to small changes in camera view-point i.e. translation and rotation.

4. FEATURE EXTRACTION

Feature Extraction is the basis of content based Image retrieval. Broadly speaking, feature may include both text based feature and visual based features. However since there exist a vast literature on text based feature extractions, we will confine ourselves to visual feature extraction techniques. In context of visual feature scope, it is further classified as general features and domain specific features. The former includes colour, texture, and shape feature while the latter is application dependant and may include for example human faces, finger prints etc.

4.1 Colour

Colour is the one mostly used visual feature in Image retrieval. It is relatively robust to background complication and independent of image size and orientations. Some representative studies of colour perception and colour spaces can be found in [2, 3]. In image retrieval, colour histogram is the most commonly used colour feature representations. Statistically it denotes the joint probabilities of three colour channels.

Color Histogram is commonly based on the intensity of three channels. It represent represents the number of pixels that have colors in each of a fixed list of color ranges. Color Moment is based used to overcome quantization effect in color histogram. It represents to calculate the color similarity by weighted Euclidean distance. Color set is used for fast search over large collection of image. It is based on the selection of color from quantized color space.

A histogram is the distribution of the number of pixels for an image. The color histogram represents the color content of an image. It is robust to translation and rotation. Color histogram is a global property of an image. The number of elements in a histogram depends on the number of bits in each pixel in an image. For example, if we suppose a pixel depth of n bit, the pixel values will be between 0 and 2n-1, and the histogram will have 2n elements. The HSV space color histogram is calculated and the joint histogram is calculated by using Hue and Saturation Histogram by calculating the total number of pixel in both the Hue and Saturation Histogram.

4.2 Texture

Texture [4] refers to visual patterns with properties of homogeneity that do not result from the presence of only a single color such as clouds and water. Texture features typically consist of contrast, uniformity, coarseness, and density. There are two basic classes of texture descriptors, namely, statistical model-based and transform-based. The former one explores the grey-level spatial dependence of textures and then extracts some statistical features as texture representation. One example of this group is co-occurrence matrix representation. The latter approach is based on some transform such as DWT. 2D Discrete Wavelet Transform is the wavelet representation of a discrete signal X consisting of N samples can be computed by convolving X with the low pass and high pass filters and down sampling the output signal by 2, so that the two frequency bands each contains N=2 samples. With the correct choice of filters, this operation is reversible. This process decomposes the original image into two sub bands: the lower and the higher band. This transform can be extended to multiple dimensions by using separable filters.

4.3 Shape

In image retrieval, depending on the applications, some require the shape representation to be invariant to translation, rotation, and scaling, while others do not. shape descriptor is some set of numbers that are produced to describe a given shape feature. A descriptor attempts to quantify shape in ways that agree with human intuition (or task-specific requirements). Good retrieval accuracy requires a shape descriptor to be able to effectively find perceptually similar shapes from a database. Usually, the descriptors are in the form of a vector.

Shape descriptors should meet the following requirements:

1. Descriptors should be as complete as possible to represent the content of the information items.

2. Descriptors should be represented and stored compactly. The size of descriptor vector must not be too large.
3. Computation of distance between descriptors should be simple; otherwise the execution time would be too long.

Obviously, if a representation satisfies the former requirement, it will satisfy the latter as well. Therefore, in the following we will focus on shape representations that are transformation invariant. In general, the shape representations can be divided into two categories, boundary-based and region-based. The former uses only the outer boundary of the shape while the latter uses the entire shape region. The most successful representatives for these two categories are Fourier descriptor and moment invariants [5].

5. CONTENT BASED IMAGE RETIEVAL (CBIR)

The role of CBIR starts when a query image and a large data base of images are available, then CBIR extracts visual contents (features) of the query image and compares these with the visual contents of each image in the data bank. Those images in the data bank, whose visual contents closely match those of the query image, are then retrieved. These retrieved images are supposed to be looking "similar" to the query image. However, in practice, only a few retrieved images will look similar, because the extracted visual features from any image will not fully characterize represent that image. Images that are close in feature space are, in general, not close semantically.

Content-based image retrieval (CBIR), also known as query by image content (QBIC) and content-based visual information retrieval (CBVIR) is the application of computer vision techniques to the image retrieval problem, that is, the problem of searching for digital images in large. Content-based image retrieval is opposed to concept-based approaches (see concept-based image indexing).

"Content-based" means that the search analyzes the contents of the image rather than the metadata such as keywords, tags, or descriptions associated with the image. The term "content" in this context might refer to colors, shapes, textures, or any other information that can be derived from the image itself. CBIR is desirable because most web-based image search engines rely purely on metadata and this produces a lot of garbage in the results. Also having humans manually enter keywords for images in a large database can be inefficient, expensive and may not capture every keyword that describes the image. Thus a system that can filter images based on their content would provide better indexing and return more accurate results.

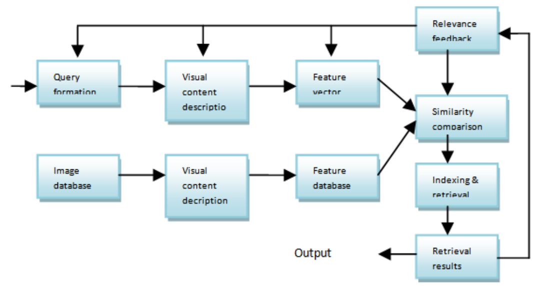

Figure 2: CBIR System

6. PROPOSED METHODOLOGY

In the approach proposed, we used rotation invariant Interest point descriptor and detector known as Speeded Up Robust Feature (SURF) which is faster and more accurate than its counterparts. The SURF descriptor is based on similar properties as its counterparts possess, with a complexity stripped down even further.
The proposed methodology is also been depicted by following flowchart of activities which consists of modules like orientation followed by exctraction of SURF descriptors which is clustered into centroids then after which a process of preparing histogram is followed and lastly which is fed in Support vector machine. All these process are elaborated in following section after the image drawn below in figure 3 which represents the flowchart of proposed methodology. The poposed approach is being tested over a limited range of images.

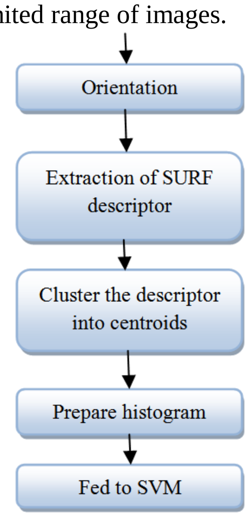

Figure 3: Flowchart depicting Proposed Methodology

The first step consists of fixing a reproducible orientation based on information from a circular region around the interest point.

Then, we construct a square region aligned to the selected orientation, and extract the SURF descriptor from it. Our approach extract the key points and descriptor from the set of training images by converting two dimensional information of images into one dimensional array of information by using Bag-of-words(BoW) approach and then clusters the descriptors into N (let's say) centroids.
Then a histogram is prepared according to centroids of respective images. Lastly, the information is fed into SVM to classify the result based on BoW available.
This in turn results in classification of different images fed into the model and results hence will be evaluated in comparison to existing SURF Histogram approach in terms of correct detection of test images.

7. SIMULATION & RESULTS

The approach proposed is tried to simulate and evaluate over MATLAB which is a high-level language and interactive environment for numerical computation, visualization, and programming. Using MATLAB, one can analyze data, develop algorithms, and create models and applications. The language, tools, and built-in math functions enable you to explore multiple approaches and reach a solution faster than with spreadsheets or traditional programming languages, such as C/C++ or Java.
In the scenario discussed above, the database includes 8 topics (classes) of different objects that in each class there are 10 types of that object. There are 41 pictures of each species. In fact, there are 80 objects and 41 images is available from each of object. All objects have been selected so those include different categories of objects in the real world such as fruits, vegetables, living creatures, vehicles and etc. and the results are compared in terms rate of correct detection of the provided test images depending on various topics and compared with SVM and without SVM. Out of which, two images of car and dog are displayed below in figure 2 and fig 3 respectively, with SURF descriptors.

Table: 1 Rate of correct detection of test images

Images	SURF Histogram	Proposed SURF Hstogram
Apple	72.68	73.4
Tomato	77.80	78.9
Pear	71.46	80.4
Cow	75.12	76.1
Dog	78.78	76.38
Horse	73.90	74.20
Cup	73.90	77.58
Car	80.24	82.35
TOTAL	**75.49**	**77.41**

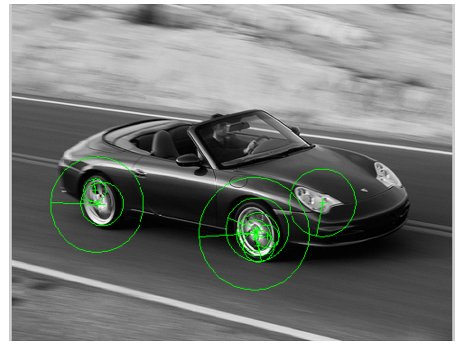

Figure 2: Test image of car

Figure 3: Test image of Dog

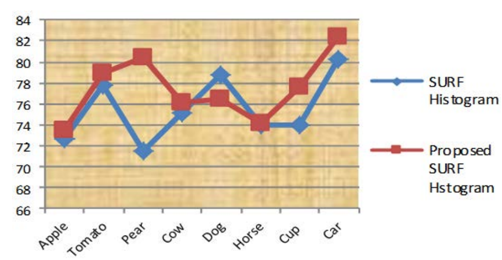

Figure:4 Graphical evaluations of results

In above shown results it is observed that proposed SURF Histogram approach proves better than conventional and existing approach, only in few cases of complex images it is showing a slight downfall, which one can ameliorate by using more efficient classification techniques.

8. CONCLUSION

Content-Based Image Retrieval (CBIR) is a challenging task which retrieves the similar images from the large database. Most of the CBIR system uses the low-level features such as colour, texture and shape to extract the features from the images. In Recent years the Interest points are used to extract the most similar images with different view point and different transformations. In this paper the SURF is combined with the colour feature to improve the retrieval accuracy. SURF is fast and robust interest points detector/descriptor which is used in many computer vision applications.

Content of an image as perceived by a human is subjective and cannot be obtained from these visual features. Humans interpret images at an abstract/high level (concept) and sometimes the features used in CBIR systems are, therefore, insufficient to capture such characteristics. This is commonly known as Semantic Gap in CBIR systems. In general, there is no direct link between the high level concepts and the low-level features.

In so many works available, interest points are used to extract the similar images with different view and accuracy. The work can be improvised and modified by the use of SURF. SURF is fast and robust interest points detector which is used in many computer vision applications and other methodologies available for the same have been discussed.

REFERENCES

[1] J. P. Eakins, "Towards Intelligent Image Retrieval," Pattern Recognition, Institute for Image Data Research, University of Northumbria at Newcastle, PERGAMON, vol. 35, pp. 3 - 14, 2002. www.elsevier.com/locate/patcog

[2] Makoto Miyahara, “Mathematical transform of (R, G, B) color data to Munsell (H,S,V) color data”, Int. Proc. SPIE Visual Communications and Image Processing Volume 1001, 650-6 57.

[3] Jia Wang, Wen Jann Yang, raj Acharya, “Color clustering techniques for color content based image retrieval from image databases”, Int. Proc. IEEE Conf. on Multimedia Comput. & Sys, 1997.

[4] C. Gotlieb, H. Kreyszig. Texture description based on co–occurrence matrices. CVGIP, 51:70–86, 1990.

[5] Y. Rui, A. C. She, and T. S. Huang, Modified fourier descriptors for shape representation—a practical approach, in Proc. of First International Workshop on Image Databases and Multi Media Search, 1996.

[6] J.P. Eakins. Automatic image content retrieval - are we getting anywhere? In Proc. of Third International Conference on Electronic Library and Visual Information Research, pages 123–135, May 1996.

[7] T. Minka. An image database browser that learns from user interaction. Master’s thesis, MIT Media Laboratory, 1995.

[8] S. I. Gallant and M. F. Johnston. Image retrieval using image context vectors: first results. In Storage and Retrieval for Image and Video Databases III, volume 2420, pages 82–94, 1995.

[9] Z. Pecenovic. Intelligent image retrieval using Latent Semantic Indexing. Master’s thesis, Swiss Federal Institute of Technology, Lausanne, Vaud, April 1997.

[10] D. McG. Squire and T. Pun. Assessing agreement between human and machine clusterings of image databases. Pattern Recognition,1998.

[11] M. Richeldi and P. L. Lanzi. ADHOC: A tool for performing effective feature selection. In Proceedings of the International Conference on Tools with Artificial Intelligence, pages 102–105, 1996.

[12] M. Flickner et al. Query by image and video content: The QBIC system. Computer, pages 23–32, September 1995.

[13] R.W. Piccard A. Pentland and S. Sclaroff. Photobook: Content-based manipulation of image databases. International Journal of Computer Vision, 18(3):233–254, 1996.

[14] J.R. Bach et al. The Virage image search engine: An open framework for image management. In Storage and Retrieval for Image and Video Databases III, volume 2420 of SPIE, pages 76–87, 1995.

[15] Herbert Bay et. Al., SURF: Spped Up Robust Features, ETH Zurich.

[16] Hou Dewen Wan Xichang, Liu Zhiang “A Content-Based Retrieval Algorithm for Document Image Database”, 2010

[17] B.L. Deekshatulu, “Learning Semantics in Content Based Image Retrieval (CBIR) – A brief review” International Conference on Information Technology for Real World Problems,2010.

[18] Lokesh Setia, Machine Learning Strategies for Content Based Image Retrieval, M.Sc Dissertation, 2008. Albert Ludwigs-Universität, Freiburg im Breisgau.

[19] Saurabh Agrawal, Nischal K Verma, Pradip Sircar, “Content based colour Image classification using SVM”, Eighth International Conference on Information Technology: New Generations, 2011.

[20] Velimurugan K., Baboo S., “Content based Image retrieval using SURF and Colour moments”, Global Journal of Computer Science and Technology, Vol 11, Issue 10, May 2011.

[21] Remi Vieux, Jenni Benios, Jean Phillipe, “Content based image retrievalusing Bag Of Regions” , Int. Conf. on Multimedia modeling, Klagenfurt, Austria, 2012.

[22] Arndt Bergner, “Content based image retrieval using Intrest point algorithm in context of scientific cultural image collection of Hebraic tombstone” Otto Friedrich University of Bramberg, Germany, August 2013.

[23] Soumia Benkarma, Lynda Zaoui, Christophe Charrier, “Accurate Image Search using Local descriptors into a compact image representation” IJCSI, Vol 10, Issue 1, No. 10, January 2013.