IJISET - International Journal of Innovative Science, Engineering & Technology, Vol. 1 Issue 4, June 2014.
www.ijiset.com
ISSN 2348 – 7968

Implementation Image Retrieval and Classification with SURF Technique

Chandrika L

M.Tech Department of CS&E, VTU RC, Mysore
Chandrika.raj13@gmail.com

Abstract

Using a work in progress for a proposed method for Content Based Image Retrieval (CBIR) and Classification. Here interest points detector and descriptor called Speeded-Up Robust Features (SURF) combined with Bag-of-Visual-Words (BoVW) are used in this proposed method. The combination yields a good retrieval and classification result when compared to other methods. however, a new dictionary building method in which each group has its own dictionary is also proposed. Hence method is tested on the highly diverse CORELIOOO database and has shown a more discriminative classification and retrieval result.

Keywords : *Speeded-Up Robust Features (SURF), Bag-of-Visual-Words (BoVW), COREL1000, content Based Image Retrieval (CBIR)*

1. Introduction

Digital imaging is on the rise in the last few decades. With the internet and World Wide Web (WWW) now widely accessible anywhere and anytime, people are embracing the possibility of accessing images stored thousands of miles away and use it for their own purposes. But retrieving a desired image within a large scale collection with thousands of images is a stressful task. Most image retrieval systems rely heavily on the text based descriptions or annotation [1]. But the text based image retrieval has a heavy limitation in which it relies heavily on manually annotating images one by one. It also depends on the annotator interpretation of the image which can vary from one person to another. Problems with the traditional method of image annotation have led to the rise of interest in techniques for retrieving images based on the content.

Early CB[R system made use of low level visual features such as color and texture. Some early works include the work by MJ. Swain and D. H. Ballard [2] in which they proposed the concept of color histogram as well as introduced the concept of histogram intersection distance metric to measure the distance between the histogram of images. Another early work is the work by S.K Chang and S.H Liu [3] in which abstraction operations are formulated to perform clustering and classification of picture object. Low level visual features are sensitive to factors such as rotation and illumination. Eventhough there are a lot of works trying to fix that, there still exist a 'semantic gap' [4] between low level visual features and

The richness of human semantics [5] because of the difference between computer machine and human brains.

2. Related works

Other works that have been done in CBIR have been focusing on narrowing the 'semantic gap' between human and computers. Such works made use of many methods: using object ontology to define high level concept, using machine learning methods to associate low level features with query concepts, using relevance feedback to learn user's intention, generating semantic template to support high level image retrieval, and fusing the evidences from HTML text and the visual content of images for WWW image retrieval [6].

In this research, we utilize a sophisticated way of image feature extraction and indexing using SURF and BoVW. SURF algorithm [7], or Speeded-Up Robust Features, is a robust image local features detector which detects interest points and produces their descriptors. The interest points are not only distinctive, but also robust to noise, detection errors, as well as geometric and photometric changes. Interest points are key points that have well-defmed locations in image scale space. They roughly represent the object of the image. Meanwhile, BoVW is the computer vision application of Bag-of-Words (BoW) model for text retrieval that assumes text documents as an unordered collection of words. The BoW model will be further explained in Chapter [I.

Our proposed method which we call Grouped BoVW (GBoVW) is different with the

normal BoVW. The normal BoVW only has 1 global dictionary and our GBoVW has a dictionary for each group or class in our test database, which make our method more discriminative and results in higher accuracy.

This paper is organized as follows: Chapter II describes the algorithm used. Chapter III describes how the system is built and also the experiment setup. Chapter [V discusses the result of the experiment. Chapter V draws conclusion from all of the experiment.

3. ALGORITHMS

A. Speeded-Up Robust Features (SURF)

Herbert Bay et. al. [7] first introduced the SURF algorithm as a novel scale- and rotation-invariant interest point detector and descriptor. SURF produces a set of interest points for each image and a set of 64-dimensional descriptors for each interest points.

To detect interest points, SURF algorithm is based on the Hessian Matrix, but uses a very basic accurate approximation of Hessian determinant using the Difference-of-Gaussian (DoG). DoG is a very basic Laplacian -based detector. The descriptor uses a distribution of Haar-wavelet responses around the interest point's neighborhood.

SURF algorithm is very similar to SIFT algorithm [8], introduced by David G. Lowe, in term that they are both an interest points detector and descriptors as image features. In SIFT, these features are identified by using a staged filtering approach. The first stage identifies key locations in scale space by looking for locations that are maxima or minima of a Difference-of-Gaussian (DoG) function. Each point is used to generate a feature vector that describes the local image region sampled relative to its scale-spacecoordinate frame.

The major difference between SIFT and SURF is that, in the implementation of scale-space, SIFT typically implemented image pyramid where the input image is iteratively convolved with Gaussian kernel and repeatedly sub-sampled (reduced in size) [9]; while SURF created scale-space by applying kernels of increasing size to the original image. Another difference is that SURF descriptor has 128 dimensions while SURF descriptor only has 64 dimensions. Some comparison papers such as [10], [11], and [12] have stated that SURF outperforms SIFT in terms of result and computational time, thus we chose SURF instead of SIFT as our feature extractor.

SURF has 4 major steps as explained in [9] and [13]:

1.Integral Image

•Creates the integral image representation of supplied input image.

•Calculates pixel sums over upright rectangular areas.

2.Fast Hessian

•Builds the determinant of Hessian response map.

•Performs a non-maximal suppression to localize interest points in a scale-space resulting in vector of localized interest point.

•Uses the determinant of Hessian Matrix

$$H(f(x,y)) = \begin{bmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x^\partial y} \\ \frac{\partial^2 f}{\partial x^\partial y} & \frac{\partial^2 f}{\partial y^2} \end{bmatrix} \quad (1)$$

$$\det(H) = \frac{\partial^2 f}{\partial x^2}\frac{\partial^2 f}{\partial y^2} - \left(\frac{\partial^2 f}{\partial x \partial y}\right)^2 \quad (2)$$

(

•Interpolates detected points to sub-pixel accuracy.

3.SURF Descriptor

•Calculates dominant orientation of the interest points.

•Constructs a 4x4 window around the interest point.

•Calculates Haar Wavelet responses from each sub-region at 5x5 regularly-spaced sample points.

•Extracts 64-dimensional descriptor vector based on sums of wavelet responses.

4.Salient Features

•Stores data associated with each individual interest point.

Figure 1 shows an example of SURF Interest points in image number 414 (Dinosaur) from COREL1000 database:

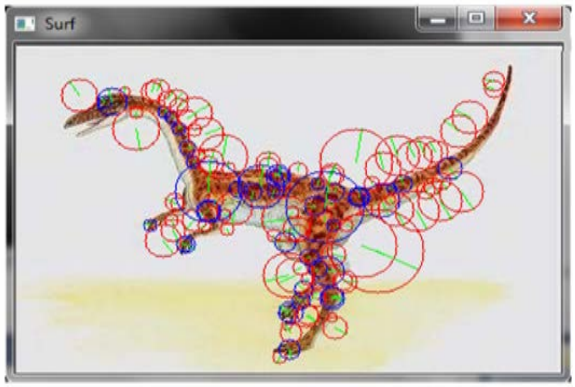

Figure 1. Example of SURF Interest Points

B. Bag-of-Visual-Words (BoVW)

Bag-of-Words (BoW) was originally devised as a text document retrieval algorithm. It describes a document based on the words it contained and the frequency of the word appearance. So the BoW considers "John loves Jane" the same as "loves Jane John" because both contains the same 3 words and the same frequencies of occurrences.

In [8], the BoVW approach was first tried out by clustering SIFT features introduced for object recognition. Ke Gao et. al. [14] proposed a filtration for SIFT interest points using attention model and then used BoVW model to efficiently index the filtered interest points. Pedro Quelhas et. al [15] use BoVW (in the paper they call it Bag-of-Visterms) to index and classify scenery images with DoG interest points. Tom Botterill, Steven Mills, and Richard Green [16] used the combination of SURF and Normal BoVW (with global dictionary) for robot localization through scene recognition. Anne Bosch [17] et al concluded in their review paper that BoW method achieved the best classification result for scenery classification.

Generally, the BoW consists of3 main steps:

1.Automatically extract the interest points and descriptor from the images.

2.Quantize the keypoints and descriptors to form the visual dictionary.

3.Find the occurences of each visual words in the image in order to build the BoW histogram.

The prototype consists of 2 main phases: Training Phase and Query Phase.

A. Training Part

Training images from the first group are first fed into the SURF function. It will extract the interest points from each image with its respective 64 dimensions descriptors. The interest points will then clustered into k clusters using k-means algorithm, using Euclidean distance, with respect to their descriptors. For this experiment, we choose k = 100. We chose k = 100 because from our experiment, k = 100 have the best accuracy, precision, and computational time ratio. We could see the comparison of different k in term of accuracy and precision in Figure 2 below:

Using our method, all values of k yields 0.96 and above in term of average accuracy, while in term of precision, all values of k yields above 0.75. But, the computational time increases significantly every time the value of k increases. For example, for k = 50 and k = 100, the training phase took approximately 6 hours of computational time while for k = 150, the training phase took 8 hours 30 minutes. For k = 300 which yields the highest precision, it took more than 12 hours for the training phase. Thus, we decided that it is not feasible to utilize k = 300 in our training and chose to use k = 100 instead. We could see from the result explained in Chapter IV that using k = 100, our method still outperforms the other methods.

Hence took the center of each cluster, chose it as the 'representative' of the cluster, and called it a visual word. Thus, we have a visual dictionary for the first group which consisted of 100 visual words.

This process was then repeated to the training images from the other groups. So in the end, we will have 10 visual dictionary, consisting of 100 visual words each, for the 10 groups from COREL1000 database.

We took the extracted features (interest points and descriptors) from the images in the training phase and calculated the Euclidean distance of each interest point with each visual words in its

respective group visual dictionary and then clustered them according to the smallest distance (nearest neighbor). In other words, for each image; we mapped the features back to the group visual dictionary. For each cluster, we count the number of interest points clustered in it and produce a histogram that showed how many interest points are clustered for each visual word. This histogram is what we call 'Bag-of-Visual-Word Histogram' and it represents each image according to its group visual dictionary.

Our method is different from Hierarchical K-means [19]. Hierarchical K -means (HKM) is one of the variant of K-means clustering algorithm, which aims to classify variables into similar groups without prior knowledge of assigned groups; while our method proposed a novel dictionary building algorithm for BoVW to achieve better classification and retrieval with prior known classes, not for clustering variables. HKM could be employed within our proposed algorithm to replace K-means as the clustering algorithm. However, our experiment showed that K-means performs better than HKM in our case, which can be seen . Thus, we decided to use K-means instead of HKM.

B. Query part

When a user submitted a query image, interest points and descriptors will be extracted using the same SURF algorithm. It will then calculate the distance from each interest point in the query image to each visual word in the visual dictionary for the first group using Euclidean Distance. From each interest points the shortest distance is chosen and then summed up from all the interest points in the query image. This way, we have the minimum distance of the query image to the first group.

The process is then repeated to all other visual dictionary for the other groups. Once the process ends, the query image should have 10 minimum distances, representing the distance of the query image to each group. We will then choose the smallest minimum distance and classify the query image to the group with the smallest distance to the query image.

When a query has found its matching group, its features will also be mapped back to the codebook. The extracted interest points and descriptors will be then clustered to the visual words by calculating the distance using Euclidean distance and choosing the smallest distance of each interest point to each visual words. Then we produce a histogram again which showed how many interest points clustered for each visual word. This way the query image will have its own BoVW Histogram.

4 Conclusion and Future work

In this paper we presented a new approach of building visual dictionary for the Bag-of-Visual-Words (BoVW) method. Our method created visual dictionary for each group in the COREL1000 database, as opposed to the global dictionary which normal BoVW employ. Compared to the the normal BoVW and a few other methods related to BoVW, our method

outperforms them in terms of accuracy and precision. Our GBoVW method is more discriminatory due to the individual group visual dictionary.

Our major challenge to the work is that our method is highly supervised. Highly supervised method means we need to determine the number of group before we perform classification.

For our future work, we would like to combine SURF features with some other methods, such as color histogram or color correlogram, which might produce even higher accuracy and precision.

References

[1]John Eakins and Margaret Graham,"Content-Based Image Retrieval," JISC Technology Applications,University of Northumbria at Newcastle, October 1999.

[2]Michael 1. Swain and Dana H. Ballard," Color Indexing," International Journal of Computer Vision,7:I, 11-32,Kluwer Academic Publishers, 1991.

[3]Shi-Kuo Chang and Sho-Hung Liu," Picture Indexing and Abstraction Techniques for Pictorial Databases," IEEE Transaction of Pattern Analysis and Machine Intelligence,1984.

[4]A. W. M. Smeulders,M. Worring,A. Gupta,R. Jain,"Content-Based Image Retrieval at the End of the Early Years," IEEE Transaction of Pattern Analysis and Machine Intelligence,2000.

ISSN 2348 – 7968

[5]X. S. Zhou and T. S. Huang,"CBIR: From Low-Level Features to High Level Semantics," Proceedings of the SPIE Image and Video Communicaitons and Processing,Vol. 3974,January 2000.

[6]Ying Liu,Dengsheng Zhang,Guojun Lu,Wei-Ying Ma,"A Survey of Content-Based Image Retrieval with High Level Semantics," The Journal of the Pattern Recognition Society,ELSEVIER,2006.

[7]Herbert Bay, Tinne Tuytelaars,Luc Van Gool, "Speeded-Up Robust Features," Computer Vision and Image Understanding (CVIU), Vol. 110,No. 3,pp. 346-359,EECV,2008.

[8]David G. Lowe "Object Recognition from Local Scale-Invariant Features," The Proceedings of the Seventh IEEE International Conference on Computer Vision,Vol. 2,pp. 1150-1157,1999.

[9]Christopher Evans "Notes on the OpenSURF Library," University of Bristol,2009.

[10]Shihua He,Chao Zhang,Pengwei Hao,"Comparative Study of Features for Fingerprint Indexing," 161h IEEE International Conference on Image Processing (ICIP),pp. 2749;2752,November 2009.

[11]Maya Dawood,Cindy Cappelle,Maan E. EI Najjar,Mohamad Khalil, Denis Pmorski, "Harris, SIFT, and SURF Features Comparison for Vehicle Localization based on Virtual 3D Model and Camera," 3'd International Conference on Image Processing Theory, Tools, and Applications (IPTA),pp. 307;312,October 2012.

[12]Luo Juan, Oubong Gwun, "SURF Applied in Panorama Image Stitching," 2nd International Conference on Image Processing Theory, Tools,and Applications (IPTA),pp. 495;499,July 2010.

[13]Anderson Rocha,Siome Goldenstein,Tiago Carvalho,Jacques Wainer, "Points of Interest and Visual Dictionary for Retina Pathology Detection," Instituto De Computacao, Universidade Estadual De Campinas,March 2011.

[14]Ke Gao, Shouxun Lin, Yongdong Zhang, Sheng Tang, Huamin Ren, "Attention Model Based SIFT Keypoints Filtration for Image Retrieval," Seventh IEEE/ACIS International Conference on Computer and Information Science,May 2008.