Volume 2, No. 3, May-June 2011
ISSN No. 0976-5697
International Journal of Advanced Research in Computer Science
RESEARCH PAPER
Available Online at www.ijarcs.info

Image retrieval using SURF features and Annotated Data

Ashish Kumar*
Computer Science and Engineering Department
Thapar University
Patiala, India
ashish.k.217@gmail.com

Shalini Batra
Asst. Prof.
Computer Science and Engineering Department
Thapar University
Patiala, India
sbatra@thapar.edu

Abstract: The ubiquity of image capturing devices has led to creation of large image collections for both personal purposes and commercial hosting of images on the Web. With this enormous increase in image storage retrieve of efficient image from a large dataset in real time has become an important task. Traditional methods of database management do not suffice here as they are not able to capture the information within the contents of an image. This paper proposes a method of image retrieval based on Speeded Up Robust Features (SURF) to detect and retrieve the most relevant image from an image dataset. Experimental results indicate that Content Based Image Recognition is a better method than any method that relies solely on annotation.

Keywords: image retrieval, computer vision, kd-tree, OpenCV, SURF

I. INTRODUCTION

Content Based Image Retrieval (CBIR) is the process of retrieving digital images from large databases based on an image's content. It consists of four steps in general:

(1) Data acquisition and processing
(2) Feature representation
(3) Image indexing
(4) Query and feedback processing

The image data may be acquired from image files already existing on hard disk drive or from an image capture source such as a camera. The subsequent processing may involve resizing of the image, change in color space properties of the image, etc. to make the image more amenable to processing. Further processing involves identifying unique features of the image that can be used for matching the image. The selection of type of features is dependent on the needs and the various methodologies that the feature supports. Some of the feature types are based on color gradients, color blobs, texture, corner detection, objects, etc. Once feature detection is done, the user can then use a query image to search for similar images within the dataset.

Finding image using simple information like image dimensions or annotation provided in the images do not provide any desired solution. Using features such as Speeded Up Robust Features (SURF)[1], based on image content as search criteria is a better option than. Further to support efficient and fast search operations for storing and matching features data structures like kd-trees [3] is used.

II. RELATED WORK

The work to extract the digital information began as soon as the idea of digitizing content that was present in physical mediums such as books, vinyl records gained foothold. The contributing ideas that would lay the foundations came from a variety of fields such as artificial intelligence, computational vision to psychology. At the forefront however was field of computer vision which provided some of the first algorithms for searching features in video, audio, and images. With the growth of the internet Web engines caught on, and started to provide image searches. Efforts were also made for integration of such systems directly into commercial database systems. It was realized by scientists during the course of developing media information systems that there was a widening semantic gap between the low level features that were used in computations by scientists and high level features that general users used their daily language when searching for images of interest. One of the earliest image based search engine to address the semantic gap was the Imagescape and it could provide searches on over 10 million images on topics like sky, trees, river etc [2].

This paper covers only images though feature techniques applied on image can be applied on videos and vice versa. Even though a video is inherently composed of images, an image collection presents unique challenge because unlike a video where subject matter is unlikely to change suddenly from frame to frame and the object is a bit easier to find due to dominant presence in consecutive frames of the video, images in a collection may differ widely in the scope of matter they cover [2].

Early attempts were mainly focused towards human face recognition but steadily the search diversified towards detecting objects in general. The difficulty of detecting objects led to research and advancements in proposal of novel and better similarity features like color features based on NF, RGB and m color space. Other techniques have included texture matching on the basis of histogram and use of computational geometry to match shapes [2].

Learning algorithms became an important part of image retrieval systems as it was realized that pattern recognition between underlying relationships of features would yield better results than was possible by simple matching of features. They allow the computer system to build a semantic understanding of the image collection and reduce the effects of noise introduced due to real world clutter in the image contents as well as ordering of images in vast collections.

The earliest learning systems were based on neural networks, also used are components based searches, statistics based methods have also shown great promise. Some image retrieval systems also integrate human feedback in training the systems so there is a more human centered relevance to search results [2].

It is not only essential to create good image similarity measures but it is also important that the result of searches be available to the user in a reasonable period of time. One of the first approaches was to create image retrieval systems based on SQL databases. But they were found to have poor performances as some of the basic assumptions on data integral to their design don't hold in case of image data. Researchers then turned to similarity based databases using tree like structures to perform similarity matches [2]. Hybrid approaches that involve traditional Relational Database Management Systems (RDBMS) have also been proposed [4].

Some recent work include using Multi-Channel based CBIR systems that work using multiple color channel representation of an image to find the relevant images[5] , using texture for image similarity and retrieval [6], using SIFT with user feedback to determine the closest match[7]. Some new technologies being introduced use not only the image content information but also the associated metadata like GPS data to segment images based on location for better image data segmentation [8]. The popular image search services provided by Google and Microsoft ,through Google Search and Bing , respectively are prominent examples of large scale Web based proprietary CBIR implementations that use not only image content but also text and other metadata like hyperlinks to provide accurate searches.

The search for better features continues to capture more relevant information as well as in terms of computation scalability in regards to application use in a server application or running on the latest Smart phones.

III. APPROACH AND ALGORITHMS

A feature is a metric on the basis of which image contents can be expressed. These features can then be further stored, retrieved and matched more efficiently in terms of memory and computation costs as compared to direct image data. There are two kind of features :(1) Global features (2) Local features. Global features represent the whole image information as a single feature vector not considering the many regions and objects that are part of the image content. Local features on the other hand are calculated from the individual regions and objects giving a better overview of an image's content [9]. Scale Invariant Feature Transform (SIFT) [10] and Speeded Up Robust Features (SURF) are type of local features. SURF was chosen because it has been shown to be computationally less time to calculate and descriptors are almost as good as generated by SIFT[11][12].

A. *Speeded Up Robust Features(SURF)*

SURF [1] is a scale and rotation invariant detector and descriptor. Scale and rotation invariance mean that an object can be identified even though if its representation gets scaled in size or it is rotated about an axis in its representation in an image. Variance occurs due to the way information exists in reality and the incompleteness with which it can be captured from a recording. Invariance is important as applied to chosen features, as measurement of similarity is possible only with respect to those features which do not change image to image [13]. The algorithm first detects key points within the image, these are points which have been deemed to be acceptable for creating unique descriptor representation of image contents. The next step calculates the descriptors for all detected key points which may or may not necessarily result in descriptors. The descriptors are vectors of floating point values with descriptor of increasing lengths computationally costly but also accurate though the accuracy does not scale with increase in length.

B. *Randomized kd-trees*

Kd-tree data structure was proposed in 1975 as a multidimensional binary search tree (or kd-tree, where k is the dimensionality of the search space) as a data structure for storage of information to be retrieved by associative searches. In addition to its storage efficiency, a significant advantage of this structure is that a single data structure can handle many types of queries very efficiently [3]. It is a binary tree that recursively splits the whole input space into partitions, in a manner similar to a decision tree acting on real-valued inputs. Each node in the kd-tree represents a certain hyper-rectangular partition of the input space; the children of this node denote subsets of the partition. Hence, the root of the kd-tree is the whole input space, while the leaves are the smallest possible partitions this kd-tree offers and each leaf explicitly records the data points that reside in the leaf. The tree is built in a manner that adapts to the local density of input points and so the sizes of partitions at the same level are not necessarily equal to each other [13].

A randomized kd-tree is built by choosing the split dimension randomly from the first D dimensions on which data has the greatest variance. Speed of search increases by building multiple randomized kd-trees and searching across them in parallel. The creation of randomized kd-trees is such that they are mostly unique, thus avoiding any wasteful results caused due to similarity of tree structures. As linear search is too costly, an approximate nearest neighbor search is done, in which non-optimal neighbors are sometimes returned. Such approximate algorithms can be orders of magnitude faster than exact search, while still providing near optimal accuracy [14].

The original implementation of SURF algorithm is proprietary, the implementation used for the experiment is an open source implementation and part of Open Source Computer Vision Library (OpenCV). OpenCV is an open source library that consists of implementations for various algorithms and routines that support image processing and computer vision techniques. The algorithm for storing features and retrieving nearest approximate matches is based on Fast Library for Approximate Nearest Neighbor (FLANN) that has been integrated into OpenCV. SURF will detect key points and create descriptors that will enable finding similar content in other image and a data structure that supports fast retrieval and matching of stored content.

IV. METHODOLOGY

The method used can be broken into three steps
(1) Detect and extract features from images
(2) Store the features in randomized kd-trees
(3) Use a query image to find similar images

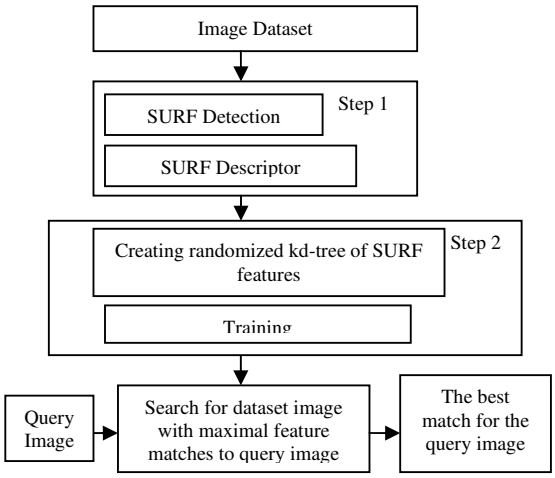

Figure 1. Block diagram of Image retrieval system

MIRFlickr [15] dataset has been used for testing the proposed methodology. The first five thousand pictures are broken into five sets of thousand images each. Images are read for each set sequentially and SURF feature detection and descriptor extraction are done sequentially. Once all the images have been processed, the descriptors are passed to a FLANN based descriptor matcher which stores the features in randomized kd- trees. After training randomized kd tree, the query image SURF detection and descriptor extraction is done. The query images are from within the image dataset but are not part of the five sets of images on which the querying was done. The descriptors from query image are then used to find the best image from the image collection using nearest neighbor search. The search for nearest neighbors is approximate and is limited to 2 nearest neighbors. Since the MIRFlickr dataset is annotated, the top 100 image matches are matched against the annotated data to predict accuracy of the search.

V. EXPERIMENTAL SETUP

As discussed earlier the image dataset was from MIR Flickr. The computer system was a 1.5 GHz and 2GB RAM machine, the program was developed on Visual Studio 2010 Express Edition and OpenCV version 2.2.

Table 1. Match results

Percentage of accurate feature matches with respect to annotated data for images		
Image Set	***SURF-64***	***SURF-128***
Set-1	10.39%	10.92%
Set-2	10.36%	10.52%
Set-3	11.07%	11.01%
Set-4	10.64%	10.99%
Set-5	10.39%	10.65%

VI. DISCUSSION

From the table we see that for the dataset there is no substantial difference between SURF-64 and SURF-128.
The higher percentage of match in SURF-64 than SURF-128 can be attributed to the use approximate nature of nearest neighbor search. In general, the low percentage of recognition can be explained in part by the incomprehensiveness of annotated data, the usage of default parameters of various functions which were not adjusted for the dataset and the approximate neighbor search. Also, the query images may not have had many features of same category as the images in the data set used. The successful correlations may be increased using user feedback and a better learning algorithm.

VII. CONCLUSION

A SURF feature based image retrieval system was built and found to be workable, though with a low success rate using annotations as a truth based feature. The SURF feature based CBIR may be used to annotate images as comprehensive manual annotation is costly in terms of money, time and will still lack comprehensiveness. The capabilities of the system can be enhanced by incorporating human feedback in learning algorithms.

VIII. REFERENCES

[1] H. Bay, T. Tuytelaars, and L. V. Gool, "SURF: Speeded up robust features," in Proc. of the 9th European Conference on Computer Vision (ECCV'06), ser. Lecture Notes in Computer Science, A. Leonardis, H. Bischof, and A. Pinz, Eds., vol. 3951. Graz, Austria: Springer-Verlag, May 2006, pp. 404–417.

[2] Michael S. Lew, Nicu Sebe, Chabane Djeraba, Ramesh Jain, "Content-Based Multimedia Information Retrieval:State of the Art and Challenges, ACM Transactions on Multimedia Computing, Communications, and Applications," vol. 2, issue 1, 2006, pp. 1-19.

[3] J. L. Bentley, "Multidimensional binary search trees used for associative searching." Communications of the ACM (CACM), Vol. 18(9):509–517, 1975.

[4] Seung-Hoon Lee, Gi-Hwa Jang, Su-Hyun Lee, Sung-Hwan Jung; Yong-Tae Woo , "A content-based image retrieval system using extended SQL in RDBMS ," Information, Communications and Signal Processing, 1997. ICICS., Proceedings of 1997 International Conference on , vol., no., pp.1069-1072 vol.2, 9-12 Sep 1997.

[5] James C. French, James V. S. Watson and Xiangyu Jin and W. N.Martin, "Integrating Multiple Multi-Channel CBIR Systems (Extended Abstract)," Proc. Inter. Workshop on Multimedia Information Systems,pp.85—95, 2003.

[6] Selim Aksoy, Professor Robert, M. Haralick, "Textural features for content-based image database retrieval," In Proceedings of IEEE Workshop on Content-Based Access of Image and Video Libraries, in conjunction with CVPR'98, 1998.

[7] X. Wangming, W. Jin, L. Xinhai, Z. Lei, and S. Gang. "Application of image sift features to the context of cbir." In CSSE '08: Proceedings of the 2008 International Conference on Computer Science and Software Engineering, pages 552{555, Washington, DC, USA, 2008. IEEE Computer Society.

[8] Yan-Tao Zheng, Ming Zhao, Yang Song, Hartwig Adam, Ulrich Buddemeier, Alessandro Bissacco, Fernando Brucher, Tat-Seng Chua and Hartmut Neven, "Tour the World: Building A Web-Scale Landmark

Recognition Engine." NUS Grad. Sch. for Integrative Sci. & Eng., Nat. Univ. of Singapore, Singapore, Singapore.

[9] Mohamed Aly, Peter Welinder, Mario Munich, and Pietro Perona, "Automatic Discovery of Image Families: Global Vs. Local Features." IEEE International Conference on Image Processing (ICIP), Cairo, Egypt, November 2009.

[10] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," International Journal of Computer Vision, vol. 60, no. 2, pp. 91–110, Nov. 2004.

[11] Bauer, J., S¨underhauf, N., & Protzel, P. (2007). "Comparing Several Implementations of Two Recently Published Feature Detectors." In Proc. of the International Conference on Intelligent and Autonomous Systems, IAV, Toulouse, France. Kan Deng.

[12] Christoffer Valgren and Achim Lilienthal, "SIFT, SURF and Seasons: Long-term Outdoor Localization Using Local Features." Proc. of 3rd European Conference on Mobile Robots,2007.

[13] Kan Deng, OMEGA: On-Line Memory-Based General Purpose System Classifier. PhD thesis, Robotics Institute, Carnegie Mellon University,1998.Technical Report CMU-RI-TR-98-33.

[14] Marius Muja, David G. Lowe , "Fast Approximate Nearest Neighbors With Automatic Algorithm Configuration". In VISAPP International Conference on Computer Vision Theory and Applications 2009, pp 331-340.

[15] M. J. Huiskes, M. S. Lew (2008). The MIR Flickr Retrieval Evaluation. ACM International Conference on Multimedia Information Retrieval (MIR'08), Vancouver, Canada.

