Volume 4, Issue 1, January 2014 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

A Survey of Content-Based Image Retrieval Systems using Scale-Invariant Feature Transform (SIFT)

Dr.K.VELMURUGAN
Associate Professor, Department of MCA
SNR Sons College, Coimbatore, Tamil Nadu, India

Abstract— ***Content-based image retrieval (CBIR) is a method for finding similar images from large image databases. As the network and development of multimedia technologies are becoming more popular, users are not satisfied with the traditional information retrieval techniques. In recent years, local descriptors are used as image features to improve the performance of CBIR. The SIFT is one of the most local feature detector and descriptors used in many computer vision applications. This paper provides the surveys of CBIR systems that using SIFT algorithm to extract the local features of images.***

Keywords— Content-Based Image Retrieval Systems, Scale-Invariant Feature Transform

I. INTRODUCTION

Content-based image retrieval (CBIR) is a system, in which retrieves visual-similar images from large image database based on automatically-derived image features, which has been a very active research area recently. In most of the existing CBIR systems[1], the image content is represented by their low-level features such as color, texture and shape[2][3]. In general, image features can be either local or global [4]. The global features describe the visual content of the entire image. The retrieval systems based on global features cannot represent all the characteristics of the image. Therefore, the global features are not suitable for tasks like partial image matching or searching for images that contain the same object or same scene with different viewpoints. In order to avoid using global features, the interest points detectors were introduced to represent the local features of images in image retrieval systems [5,6]. The interest points are the salient image patches that contain rich local information about an image. Many algorithms have been developed for the purpose of detecting and extracting the interest points like Harris[7], Hessian[8] , Scale invariant [9] , affine-invariant [10] and Difference-of- Gaussians (DOG) [11],SIFT[12] and SURF[13]. The SIFT is one of the most interest point detector and descriptors used in many computer vision applications. SIFT descriptors, which are invariant to image scaling and transformation and rotation, and partially invariant to illumination changes and affine, present the local features of an image. Recently SIFT is used in many CBIR system to describe the content of images. In the SIFT based CBIR system, a few thousand keypoints are extracted from each image. For matching the key descriptors of the images a nearest neighbour search (NNS), an algorithm is used to detect similarities between keypoints between the two images.

II. OVERVIEW OF SIFT ALGORITHM

Lowe [12] has presented a powerful framework Scale Invariant Features Transform (SIFT) to recognize/retrieve objects. The SIFT algorithm identifies features of an image that are distinct, and these features can in turn be used to identify similar or identical objects in other images. The SIFT algorithm produces keypoint descriptors. A keypoint is an image feature which is so distinct that image scaling, noise, or rotation does not, or rather should not, distort the keypoint. A keypoint descriptor is a 128-dimensional vector that describes a keypoint and contains a lot of information about the point it describes.

The SIFT algorithm can be viewed as a keypoint descriptor composed by four major stages. The output of the SIFT algorithm is a set of keypoint descriptors of images. Once such descriptors have been generated for more than one image, one can begin image matching.

1) Scale-space extrema detection
2) Keypoint localization
3) Orientation assignment
4) Keypoint description

In the following, we describe each one of these steps

1) ***Scale-space Extrema Detection :-*** In the first stage, the method identifies locations and scales same object. For detecting locations that are invariant to scale change of the image can be accomplished by searching for stable features across all possible scales, using a continuous function of scale known as scale space. The scale space of

an image *I (x, y)* is defined as a function *L(x, y, σ)*, that is produced from the convolution of *I (x, y)* with a variable-scale Gaussian *G(x, y, σ):*

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y) \qquad (1)$$

where ∗ is the convolution operation in *x* and *y*, and *G(x, y, σ)* is a variable-scale Gaussian and *I(x, y)* is the input image.

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} \; e^{-(x2+y2)/2\sigma} \qquad (2)$$

To efficiently detect stable keypoint locations in scale space,it is used a scale space extrema based on the difference-of-Gaussian function, D(x, y, σ), which can be computed from the difference of two nearby scales separated by a constant multiplicative factor *k[12] :*

$$D(x, y, \sigma) = L(x, y, k\sigma) - L(x, y, \sigma) \qquad (3)$$

Figure 2 shows an efficient approach to construction of *D(x, y, σ).* To detect the local maxima and minima of *D(x, y, σ)* each point is compared with its 8 neighbours at the same scale, and its 9 neighbours up and down one scale. If this value is the minimum or maximum of all these points then this point is an extrema.

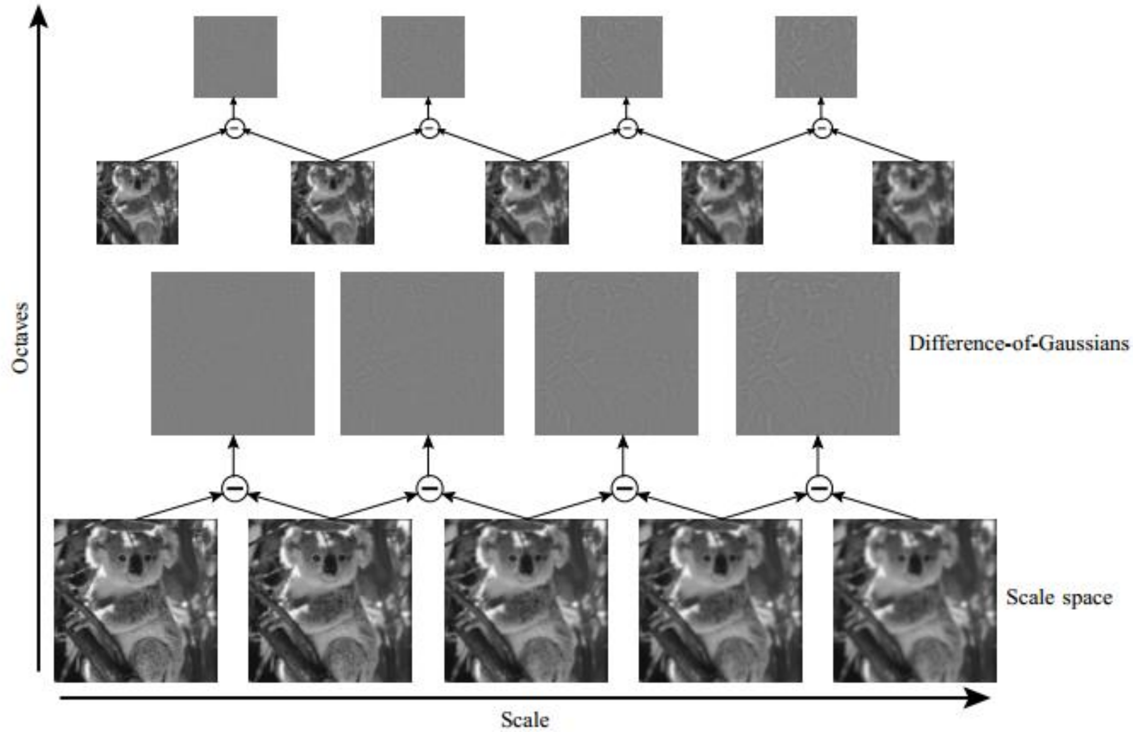

Figure 2. The construction of scale space extrema based on difference-of-Gaussians.

2) ***Keypoint localization:-*** Once a keypoint candidate has been found, the next step is to adjust its accuracy. For all interest points keypoint a detailed model is created to determine location and scale. The Keypoints are selected based on their stability. A stable keypoint is thus a keypoint resistant to image distortion. It is performed by a Taylor expansion of the scale-space function, D(x, y, σ), shifted so that the origin is at the sample point.

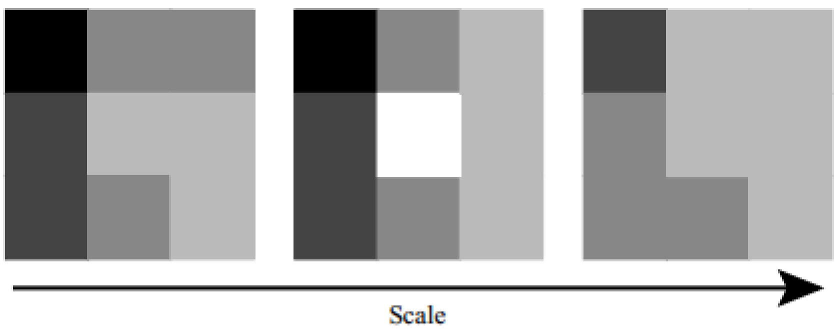

Figure 3. Keypoint localization at different scales.

3) ***Orientation assignment:-*** For each of the keypoints identified the SIFT computes the direction of gradients around. One or more orientations are assigned to each keypoint based on local image gradient directions. By assigning a consistent orientation to each keypoint based on local image properties, its feature vector can be represented relative to this orientation and therefore achieve invariance to image rotation. This keypoint orientation is calculated from an orientation histogram of local gradients from the closest smoothed image L(x, y, σ). For each image sample L(x, y) at this scale, the gradient magnitude m(x, y) and orientation θ(x, y) is computed using pixel differences:

$$m(x, y) = ((L(x + 1, y) - L(x - 1, y))^2 + (L(x, y + 1) - L(x, y - 1))^2)^{½} \quad (4)$$

$$\theta(x, y) = \tan^{-1}((L(x, y + 1) - L(x, y - 1)) / (L(x + 1, y) - L(x - 1, y))) \quad (5)$$

The orientation histogram has 36 bins covering the 360 degree range of orientations. Each point is added to the histogram weighted by the gradient magnitude, *m(x, y)*, and by a circular Gaussian with σ variance that is 1.5 times the scale of the keypoint. Additional keypoints are generated for keypoint locations with multiple dominant peaks whose magnitude is within 80% of each other [12]. The dominant peaks in the histogram are interpolated with their neighbors for a more accurate orientation assignment.

4) ***Keypoint Descriptor:-*** The local image gradients are measured at the selected scale in the region around each keypoint. These are transformed into a representation that allows for significant levels of local shape distortion and change in illumination. Figure 4 illustrates the computation of the feature vector of each keypoint.

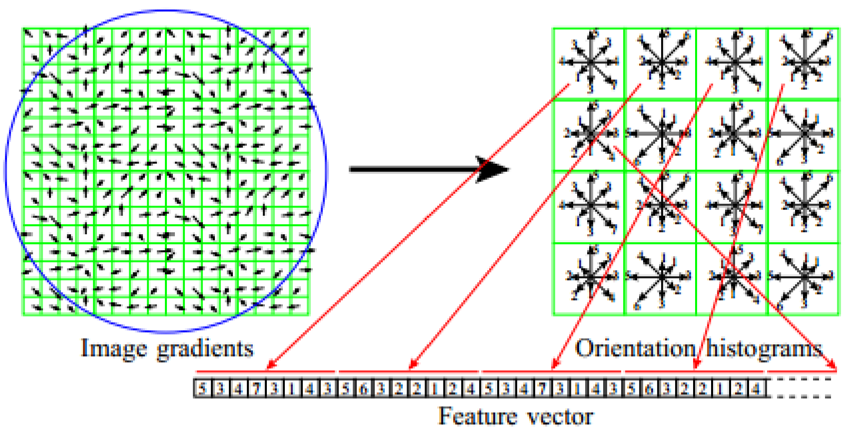

Figure. 4. The computation of the feature vector of a keypoint

III. A SIFT APPROACH FOR CBIR

In this approach SIFT features are first extracted from a set of reference images and stored in a database. A query image is matched by individually comparing each feature from the query image to this previous database and finding candidate matching features based on Euclidean distance of their feature vectors and then ranked as per number of matching keypoints. The block diagram of SIFT based CBIR system is shown in figure 5.

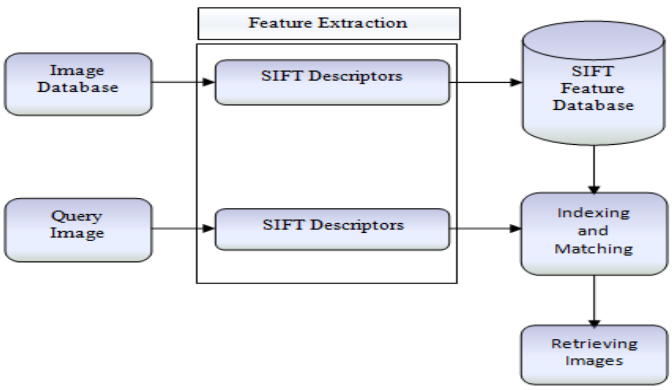

Figure 5. A CBIR System with SIFT Algorithm

IV. THE CBIR SYSTEMS USING SIFT ALGORITHM

A CBIR system used SIFT feature sets for indoor image retrieval and robot localization [14]. This system reduced the complexity of each SIFT feature and the number of SIFT features required to describe a scene. The *kd*-tree with the Best Bin First(BBF), an Approximate Nearest Neighbors(ANN) search algorithm, was used to index and match the SIFT features[15]. And a modified voting scheme called nearest neighbor distance ratio scoring (NNDRS) was used to calculate the aggregate scores of the corresponding candidate images in the database. A content-based Web image search engine [16] saved SIFT feature as XML files. For matching, a dynamic probability function that replaced the original fixed value to determine the similarity distance of ROI (Region of interest) and database from Web training images. The SIFT based neural network and Graph-based segmentation technique is combined in system [17]. A novel image retrieval system based on bag-of-features (BoF) model by integrating scale invariant feature transform (SIFT) and local binary pattern (LBP) is proposed in [18]. The SIFT and LBP features yield complementary and substantial improvement on image retrieval even in the case of noisy background and ambiguous objects.

The Image retrieval system is also used full for the agricultural field [19] to determine growth and insect attack and used for the detecting weeds in the field, whether or not a plant is damage by a specified illness and distinguish weeds form soil regions. This system using SIFT features to capture local characteristics of the plant, as well as global shape descriptors based on the outer contour of the plant. A few thousand keypoints are extracted from each image and image matching involves distance computations across all pairs of SIFT feature vectors from both images, which is quite costly. The SIFT features performed surprisingly well even after quantizing each component to binary, when the medians were used as the quantization thresholds [20]. Quantized features preserved both distinctiveness and matching properties. The SIFT descriptor vectors for each image was indexed by making the use of vocabulary tree and relevance feedback technique [21] also used to bridge the gap between low level features and high level concepts. A local image descriptor used VQ-SIFT for more effective and efficient image retrieval [22]. Instead of SIFT's weighted orientation histograms, the vector quantization (VQ) histogram was applied as an alternate representation for SIFT features. The integration of SIFT features with VQ-based local descriptors achieved better image retrieval accuracy than the conventional algorithm.

The SIFT is also applied for trademark image retrieval [23] by combining the image global features with local features. The Zernike moments of retrieved images are extracted and sorted them according to similarity and candidate images are formed. Then, the SIFT features are used for matching the query image accurately with candidate images. A 2D medical image retrieval system [11] which employed interest points derived from superpixels in a bags of visual words (BVW) framework which relied on stable interest points so that the local descriptors can be clustered into representative, discriminative prototypes (the visual words). The usage of the centers of mass of superpixels as interest points yields higher retrieval accuracy when compared to using Difference of Gaussians (DoG) or a dense grid of interest points.

The SIFT based Radial Basis Function (RBF) method is proposed for efficient image retrieval [25]. The changes of illumination does not affect the system performance due to these SIFT features. The distances between feature vectors of database and the query image were computed using the classifier RBF. The CBIR used video as input using SIFT algorithms and Edge detection algorithm [26]. The system discussed the efficient result using SIFT and Edge detection algorithm. A feature-based method for matching facial sketch images to face photographs also presented [27]. In this system the descriptors are calculated at selected discrete points using SIFT. The attempt to evaluate the application of the SIFT to refine CBIR is proposed [28].

A system for face recognition and retrieval was proposed in [29]. The retrieval rate drastically increased especially for LFW images by extracting LBP and SIFT features of training images and arranging them in sparse representation; shape context and inner distance shape contexts methods were applied on test image for deriving relevant images with better performance. A bag-of-words (BoW) or bag-of-features model in image retrieval system was proposed [30]. First quantizing local descriptors into visual words, and then applying scalable textual indexing and retrieval schemes. The SIFT algorithm used in CBIR for color based and shape based image retrieval [31]. This approach relied on the choice of several parameters which directly impact its effectiveness when applied to retrieve images. To match the key-points of images k-means clusters were used to compare similarity.

V. CONCLUSION

The purpose of this survey is to provide an overview of the CBIR systems that using SIF algorithm to describe the content of images. Many sophisticated algorithms have been designed to describe the low level features. But these algorithms cannot adequately model image semantics. They have many limitations when dealing with broad content image databases. Extensive experiments on CBIR systems show that low-level contents often fail to describe the high level semantic concepts in user’s mind. Therefore, nowadays the CIBR systems using the SIFT algorithm to extract the features of the images to improve the performance of the system. The SIFT owes the excellent performance in image matching and to retrieve desired images from the database.

REFERENCES

[1] R. C. Veltkamp and M. Tanase, “Content-based image retrieval systems:A survey,” Department of Computing Science, Utrecht University, Tech.Rep. UU-CS-2000-34, 2000.

[2] W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain,“Content-based image retrieval at the end of the early years,” IEEE Trans. Pattern Analysis and Machine Intelligence , vol. 22, no. 12, pp.1349–1380, 2000.

[3] R. Datta, D. Joshi, J. Li, and J. Z. Wang, “Image retrieval: Ideas, influences, and trends of the new age,” ACM Comput. Surv. , vol. 40, no. 2, pp. 1–60, 2008.

[4] A. Halawani, A. Teynor, L. Setia, G. Brunner, and H. Burkhardt, "Fundamentals and Applications of Image Retrieval: An Overview", presented at Datenbank- Spektrum, 2006, pp.14-23.

[5] C. Schmid and R. Mohr. Local grayvalue invariants for image retrieval. IEEE Transactions on Pattern Analysis and Machine Intelligence, 19(5):530.534, 1997. [6] T.Tuytelaars and L.VanGool, “Content-based Image Retrieval based on Local Affinely Invariant Regions”.3rd Int. Conf. on Visual Information Systems, Visual99, Amsterdam, The Netharlands, 2-4 June 1999,pp.493-500.

[7] Harris and M.J. Stephens. "A combined corner and edge detector". In Proc. Alvey Vision Conference, pp. 147-152, 1988.

[8] K. Mikolajczyk. Interest point detection invariant to affine transformations. PhD thesis. Institut National Polytechnique de Grenoble. 2002.

[9] K. Mikolajczyk and C. Schmid. "An affine invariant interest point detector" in Proc. of the 7th European Conf. on Computer Vision. Copenhagen, Denmark, vol. I, pp. 128–142. 2002.

[10] D. Lowe. "Object Recognition from Local Scale- Invariant Features", in Proc. of the 7th Int. Conf. on Computer Vision. vol. 2, p. 1150. September 20–25, 1999. IEEE, 1999.

[11] K. Mikolajczyk and C. Schmid (2001), Indexing based on scale invariant interest points., International Conference on Computer Vision, pp. 525-531.

[12] Lowe, D. G., "Distinctive Image Features from Scale-Invariant Keypoints", International Journal of Computer Vision, 60, 2, pp. 91-110, 2004.

[13] Herbert Bay, Andreas Ess, TinneTuytelaars, Luc Van Gool, "SURF: Speeded Up Robust Features", Computer Vision and Image Understanding (CVIU), Vol. 110, No. 3, pp. 346- -359, 2008.

[14] Ledwich, Luke, and Stefan Williams. "Reduced SIFT features for image retrieval and indoor localisation." *Australian conference on robotics and automation*. Vol. 322. 2004.

[15] 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 Interna- tional Conference on Computer Science and Software Engineering, pages 552–555, Washington, DC, USA, 2008. IEEE Computer Society.

[16] Wang Z, Jia K, Liu P (2009) An effective web content-based image retrieval algorithm by using SIFT feature. Software Eng 1:291–295.

[17] N.D. Anh, P.T. Bao, B.N. Nam, and N.H. Hoang, "A New CBIR System Using SIFT Combined with Neural Network and Graph-Based Segmentation", in Proc. ACIIDS (1), 2010, pp. 294-301.

[18] X. Yuan, J. Yu, Z. Qin and T.Wan, A SIFT-LBP retrieval model based on bag-of features. In 18th IEEE International Conference on Image Processing(ICIP), pp. 1061-1064, 2011.

[19] Mandlik, MrVinay S., and Sanjay B. Dhaygude. "Agricultural Plant Image Retrieval System Using CBIR."International Journal of Emerging Technology and Advanced Engineering, Volume 1, Issue 2, December 2011,(ISSN 2250-2459).

[20] K. Peker, "Binary sift: Fast image retrieval using binary quantized sift features," in CBMI, pp.217–222, 2011.

[21] Anil BalajiGonde, R.P. Maheshwari, R. Balasubramanian ,SIFT Feature with Relevance Feedback for Image Retrieval,TECHNIA – International Journal of Computing Science and Communication Technologies, VOL. 3, NO. 2, Jan. 2011. (ISSN 0974-3375).

[22] Qiu Chen, Feifei Lee, Koji Kotani, TadahiroOhmi ,Local Image Descriptor using VQ-SIFT for Image Retrieval, World Academy of Science, Engineering and Technology 59 2011.

[23] WANG, Zhenhai, and Kicheon HONG. "A Novel Approach for Trademark Image Retrieval by Combining Global Features and Local Features." *Journal of Computational Information Systems* 8.4 (2012): 1633-1640.

[24] Haas, Sebastian, et al. "Superpixel-Based interest points for effective bags of visual words medical image retrieval." *Medical Content-Based Retrieval for Clinical Decision Support*. Springer Berlin Heidelberg, 2012.

[25] Shraddhakumar, Shwetasinghal, Ankitasharma,Radial Basis Function used in CBIR for SIFT Features, International Journal of Advanced Research in Computer Science and Software Engineering, Volume 2, Issue 4, April 2012,ISSN: 2277 128X.

[26] HANWATE, PS, and UL KULKARNI. "APPROACH TOWARDS CBIR USING SIFT" International Journal of Engineering Science and Technology (IJEST,Vol. 4 No.07 July 2012 , ISSN : 0975-5462

[27] RakeshS ,KailashAtal, Ashish Arora, PulakPurkait,BhabatoshChanda'Face Image Retrieval Based on Probe Sketch Using SIFT Feature Descriptors'-1st Indo-Japan Conference on Perception and Machine Intelligence,2012.

[28] Kamath, M., Punjabi, D., Sabnis, T., Upadhyay, D., &Shrawne, S. Improving Content Based Image Retrieval using Scale Invariant Feature Transform. International Journal of Engineering and Advanced Technology (IJEAT), Jun 2012.

[29] Tayade, Yogesh R., and S. M. Bansode. "An Efficient Face Recognition and Retrieval Using LBP and SIFT."International Journal of Advanced Research in Computer and Communication Engineering Vol. 2, Issue 4, April 2013

[30] Liu, Jialu. "Image Retrieval based on Bag-of-Words model." *arXiv preprint t arXiv:1304.5168* (2013).

[31] Dr. Sambhavjain ,Manojchouhan, CONTENT BASED IMAGE RETRIEVAL BY CLUSTERING USING (SIFT) Algorithm,International Journal of Computer Science and Information Security (IJCSIS) Volume (1) : Issue (1),June 2013.

