



International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Special Issue 5, December 2014

Logo Detection and Recognition by Image Extraction

Shani S J¹, Anu Kuriakose²

Student, M-Tech. Electronics (Communication Engineering), Federal Institute of Science and Technology, Angamaly,
Kerala, India

Assistant Professor, Department of Electronics and communication Engineering, Federal Institute of Science and
Technology, Angamaly, Kerala, India

ABSTRACT: This paper contribute the design of a framework able to extract match and recognize multiple instances of multiple test logos with reference logos in image archives. Reference logos and test images are seen as constellations of local features (interest points, regions, etc.) and matched by minimizing an energy function mixing: 1) a fidelity term that measures the quality of feature matching, 2) a neighborhood criterion that captures feature co-occurrence/geometry, and 3) a regularization term that controls the smoothness of the matching solution.

KEYWORDS: Image detection, SIFT, Image recognition,

I.INTRODUCTION

The use of symbols in communications has been documented since the time of hieroglyphics. Symbols are used to convey depth of information in an efficient manner, and they can be used to replace what would normally require a lot of explanation. Symbols are a fantastic way to convey a concept. Graphic logos are a special class of visual objects extremely important to assess the identity of something or someone.

The expanding and massive production of visual data from companies, institutions and individuals, and the increasing popularity of social systems like Flickr, YouTube and facebook for diffusion and sharing of images and video, have more and more urged research in effective solutions for object detection and recognition to support automatic annotation of images and video and content-based retrieval of visual data. In industry and commerce, logos have the essential role to recall in the customer the expectations associated with a particular product or service. This economical relevance has motivated the active involvement of companies in soliciting smart image analysis solutions to scan logo archives to find evidence of similar already existing logos, discover either improper or non authorized use of their logo, unveil the malicious use of logos that have small variations with respect to the originals so to deceive customers, analyse videos to get statistics about how long time their logo has been displayed. Logos are graphic productions that either recall some real world objects, or emphasize a name, or simply display some abstract signs that have strong perceptual appeal.

Colour may have some relevance to assess the logo identity. But the distinctiveness of logos is more often given by a few details carefully studied by graphic designers semiologists and experts of social communication. The graphic layout is equally important to attract the attention of the customer and convey the message appropriately and permanently. Different logos may have similar layout with slightly different spatial disposition of the graphic elements, localized differences in the orientation, size and shape, or in the case of malicious tampering – differ by the presence/absence of one or few traits. Logos however often appear in images/videos of real world indoor or outdoor scenes superimposed on objects of any geometry, shirts of persons or jerseys of players, boards of shops or billboards and posters in sports playfields. In most of the cases they are subjected to perspective transformations and deformations, often corrupted by noise or lighting effects, or partially occluded. Such images and logos thereafter have often relatively low resolution and quality. Regions that include logos might be small and contain few information. Logo detection and recognition in these scenarios has become important for a number of applications such as the automatic



International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Special Issue 5, December 2014

identification of products on the web to improve commercial search-engines, the verification of the visibility of advertising logos in sports events the detection of near-duplicate logos and unauthorized uses. Special applications of social utility have also been reported such as the recognition of groceries in stores for assisting the blind. A generic system for logo detection and recognition in images taken in real world environments must comply with contrasting requirements. On the one hand, invariance to a large range of geometric and photometric transformations is required to comply with all the possible conditions of image/video recording. Since in real world images logos are not captured in isolation, logo detection and recognition should also be robust to partial occlusions. At the same time, especially if we want to discover malicious tampering or retrieve logos with some local peculiarities, we must also require that the small differences in the local structures are captured in the local descriptor and are sufficiently distinguishing for recognition.

II.SIFT

Scale Invariant Feature Transform (SIFT) is an approach for detecting and extracting local feature descriptors that are reasonably invariant to changes in illumination, image noise, rotation, scaling, and small changes in viewpoint.

A. *Detection of scale-space extrema*

SIFT key points of objects are first extracted from a set of reference images and stored in a database. An object is recognized in a new image by individually comparing each feature from the new image to this database and finding candidate matching features based on Euclidean distance of their feature vectors. From the full set of matches, subsets of key points that agree on the object and its location, scale, and orientation in the new image are identified to filter out good matches. Finally the probability that a particular set of features indicates the presence of an object is computed, given the accuracy of fit and number of probable false matches. Object matches that pass all these tests can be identified as correct with high confidence.

This step detect key points using a cascade filtering approach that uses efficient algorithms to identify candidate locations that are then examined in further detail. The first stage of key point detection is to identify locations and scales that can be repeatedly assigned under differing views of the same object. 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 efficient scale-space kernel is the Gaussian function. Therefore, the scale space of an image is defined as a function, $L(x, y, \sigma)$, that is produced from the convolution of a variable-scale Gaussian, $G(x, y, \sigma)$, with an input image, $I(x, y)$:

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

where $*$ is the convolution operation in x and y

$$G(x, y, \sigma) = [1/(2\pi\sigma^2)] * e^{-(x^2+y^2)/2\sigma^2}$$

To efficiently detect stable key point locations in scale space, we use scale-space extrema in the difference-of-Gaussian function convolved with the image, $D(x, y, \sigma)$, which can be computed from the difference of two nearby scales separated by a constant multiplicative factor k :

$$\begin{aligned} D(x, y, \sigma) &= (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y) \\ &= L(x, y, k\sigma) - L(x, y, \sigma) \end{aligned} \quad (1)$$

There are a number of reasons for choosing this function. First, it is a particularly efficient function to compute, as the smoothed images, L , need to be computed in any case for scale space feature description, and D can therefore be computed by simple image subtraction. In addition, the difference-of-Gaussian function provides a close approximation to the scale normalized Laplacian of Gaussian, $\sigma^2 \nabla^2 G$, the normalization of the Laplacian with the factor σ^2 is required for true scale invariance. The maxima and minima of $\sigma^2 \nabla^2 G$ produce the most stable image features compared to a range of other possible image functions, such as the gradient, Hessian, or Harris corner function. Once this DoG are found, images are searched for local extrema over scale and space. For eg, one pixel in an image is compared with its 8 neighbours as well as 9 pixels in next scale and 9 pixels in previous scales. If it is a local extrema, it is a potential key point. It basically means that key point is best represented in that scale.



International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Special Issue 5, December 2014

B. Orientation assignment

In this step, each keypoint is assigned one or more orientations based on local image gradient directions. This is the key step in achieving descriptor can be represented relative to this orientation and therefore achieve invariance to image rotation. First, the Gaussian smoothed image $L(x, y, \sigma)$ at the key point's scale σ is taken so that all computations are performed in a scale invariant manner. For an image sample $L(x, y)$ at scale σ , the gradient magnitude $m(x,y)$, and orientation $\theta(x,y)$ are precomputed using pixel differences:

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

$$\theta(x,y) = \text{atan2}((L(x,y-1) - L(x,y+1)), (L(x+1,y) - L(x-1,y)))$$

The magnitude and direction calculations for the gradient are done for every pixel in a neighboring region around the keypoint in the Gaussian – blurred image L . An orientation histogram with 36 bins is formed, with each bin covering 10 degrees. Each sample in the neighboring window added to a histogram bin is weighted by its gradient magnitude and by a Gaussian-weighted circular window with σ that is 1.5 times that of the scale of the key point. The peaks in this histogram correspond to dominant orientations. Once the histogram is filled, the orientations corresponding to the highest peak and local peaks that are within 80% of the highest peaks are assigned to the keypoint. In the case of multiple orientations being assigned, an additional keypoint is keypoint for each additional orientation.

C. Keypoint descriptor

Previous steps found key point locations at particular scales and assigned orientations to them. This ensured invariance to image location, scale and rotation. Now we want to compute a descriptor vector for each key point such that the descriptor is highly distinctive and partially invariant to the remaining variations such as illumination, 3D viewpoint, etc. This step is performed on the image closest in scale to the key point's scale.

First a set of orientation histograms is created on 4x4 pixel neighborhoods with 8 bins each. These histograms are computed from magnitude and orientation values of samples in a 16 x 16 region around the keypoint such that each histogram contains samples from a 4 x 4 subregion of the original neighborhood region. The magnitudes are further weighted by a Gaussian function with equal to one half the width of the descriptor window. The descriptor then becomes a vector of all the values of these histograms. Since there are 4 x 4 = 16 histograms each with 8 bins the vector has 128 elements. This vector is then normalized to unit length in order to enhance invariance to affine changes in dimension of the descriptor, i.e. 128, seems high, descriptors with lower dimension than this don't perform as well across the range of matching tasks and the computational low for finding the nearest-neighbor. Longer descriptors continue to do better but not by much and there is an additional danger of increased sensitivity to distortion and occlusion. It is also shown that feature matching accuracy is above 50% degrees. Therefore SIFT descriptors are invariant to minor affine changes. To test the distinctiveness of the SIFT descriptors, matching accuracy is also measured against varying number of keypoints in the testing database only very slightly for very large database sizes, thus indicating that SIFT features are highly distinctive.

D. SIFT feature representation

Once a keypoint orientation has been selected, the feature descriptor is selected as orientation histograms on 4 by 4 pixel neighborhoods. The orientation histograms are relative to the keypoint orientation, the orientation data comes from the Gaussian image closest in scale to the keypoint's scale. Just like before contribution of each pixel is weighted by the gradient magnitude, and by a Gaussian with σ 1:5 times the scale of the keypoint. Histograms contain 8 bins each, and each descriptor contains an array of 4 histograms around the keypoint. This leads to a SIFT feature vector with 4 * 4 * 8 = 128 elements.

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Special Issue 5, December 2014

III. EXISTING METHOD

Algorithm 1: CDS Logo Detection and Recognition

Input: Reference logo image: IX , Test image: IY , CDS parameters: $\epsilon, Na, Nr, \alpha, \beta, \tau$.

Output: A boolean value determining whether the reference logo in IX is detected in IY .

Extract SIFT from IX, IY and let $SX := \{x1, \dots, xn\}, SY := \{y1, \dots, ym\}$ be respectively the list of interest points taken from both images;

for $i \leftarrow 1$ **to** n **do**

 Compute the context of x_i , given ϵ, Na, Nr ;

for $j \leftarrow 1$ **to** m **do**

 Compute the context of y_j , given ϵ, Na, Nr ;

 Set $t \leftarrow 1, maxt \leftarrow 30$;

repeat

for $i \leftarrow 1$ **to** n **do**

for $j \leftarrow 1$ **to** m **do**

 Compute CDS matrix entry $K_{xi,yj}^{(t)}$ given α, β ;

 Set $t \leftarrow t + 1$;

until convergence (i.e. $\|K^{(t)} - K^{(t-1)}\| \approx 0$) OR

$t > maxt$;

$K \leftarrow K^{(t)}$

for $i \leftarrow 1$ **to** n **do**

for $j \leftarrow 1$ **to** m **do**

 Compute $K_{yj|xi} \leftarrow K_{xi,yj} \frac{K_{xi,yj}}{\sum_{s=1}^m K_{xi,ys}}$

 A match between x_i and y_j is declared iff

$K_{yj|xi} \geq \sum_{s \neq j}^m K_{ys/xi}$

if number of matches in $SY > \tau |SX|$ **then**

return true i.e. logo detection

else

return false;

An example of existing method is shown in figure 4.1. By comparing the feature points of both images and finding the value of τ it is possible to find the matching. Here in this example value of τ is greater than 0.5 and the logo is matched with that in test image.



Figure 4.1. Where (a) is the reference logo (b) is the test image (c) Feature points in reference image and (d) is the feature points of test image

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Special Issue 5, December 2014

IV. PROPOSED METHOD

If one image contains two logos the sift algorithm finds feature points of total image then during comparing this test image with reference image the algorithm gets confused. One solution to this problem is image segmentation that is segmenting the required logos from the test image and use for comparison. And there are different type of segmentations are there. In case of active segmentation it depends on some fixed parameters or measurements. But in case of logos in real world scenes segmentation depends on seed point selected by the user. Since it may contain other dominant objects than logos?

Here propose a segmentation algorithm which takes a fixation point as its input and outputs the region containing the given fixation point in the scene. For example the fixation point indicated by the red cross lines in figure this method segments the corresponding regions enclosing those fixations points



Figure 5.1: Fixing a point



Figure 5.2: Extracted image

The primary difference between the proposed segmentation framework and the standard approach is the fact that we always segment one region/object at a time. In order to segment multiple objects, the segmentation process will be repeated for the fixation points from inside each of the objects of interest. The diagram below shows that fixation has a critical role in proposed segmentation process and that is to identify the object of interest.

Standard Framework:

Image \longrightarrow Segmentation

New Framework:

Fixation
 \downarrow
Video/Image \longrightarrow  \longrightarrow Segmentation

Figure 5.1: Existing and proposed methods of segmentation



International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Special Issue 5, December 2014

For a given fixation point, segmenting the region/object containing that point is a twostep process.

1. **Cue Processing:** Visual cues such as color, texture, motion and stereo generate a probabilistic boundary edge map wherein the probability of a pixel to be at the boundary of any object in the scene is stored as its intensity.
2. **Segmentation:** For a given fixation point, the optimal closed contour around that point in the probabilistic edge map. This process is carried out in the polar space to make the segmentation process scale invariant

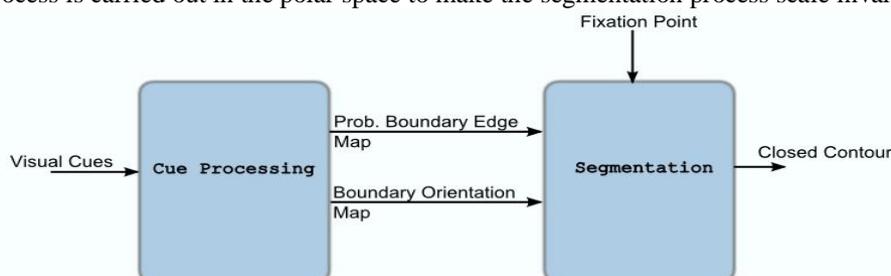


Fig 5.2: Proposed method

After segmenting the same CDS algorithm is sufficient for further process.

ACKNOWLEDGMENT

The heading of the Acknowledgment section and the References section must not be numbered. Causal Productions wishes to acknowledge Michael Shell and other contributors for developing and maintaining the IEEE LaTeX style files which have been used in the preparation of this template. To see the list of contributors, please refer to the top of file IEEETran.cls in the IEEE LaTeX distribution.

REFERENCES

- [1] S. M. Metev and V. P. Veiko, *Laser Assisted Microtechnology*, 2nd ed., R. M. Osgood, Jr., Ed. Berlin, Germany: Springer-Verlag, 1998.
- [2] J. Breckling, Ed., *The Analysis of Directional Time Series: Applications to Wind Speed and Direction*, ser. Lecture Notes in Statistics. Berlin, Germany: Springer, 1989, vol. 61.
- [3] S. Zhang, C. Zhu, J. K. O. Sin, and P. K. T. Mok, "A novel ultrathin elevated channel low-temperature poly-Si TFT," *IEEE Electron Device Lett.*, vol. 20, pp. 569–571, Nov. 1999.
- [4] M. Wegmuller, J. P. von der Weid, P. Oberson, and N. Gisin, "High resolution fiber distributed measurements with coherent OFDR," in *Proc. ECOC'00*, 2000, paper 11.3.4, p. 109.
- [5] R. E. Sorace, V. S. Reinhardt, and S. A. Vaughn, "High-speed digital-to-RF converter," U.S. Patent 5 668 842, Sept. 16, 1997.
- [6] (2002) The IEEE website. [Online]. Available: <http://www.ieee.org/>
- [7] M. Shell. (2002) IEEEtran homepage on CTAN. [Online]. Available: <http://www.ctan.org/tex-archive/macros/latex/contrib/supported/IEEEtran/>
- [8] *FLEXChip Signal Processor (MC68175/D)*, Motorola, 1996.
- [9] "PDCA12-70 data sheet," Opto Speed SA, Mezzovico, Switzerland.
- [10] A. Karnik, "Performance of TCP congestion control with rate feedback: TCP/ABR and rate adaptive TCP/IP," M. Eng. thesis, Indian Institute of Science, Bangalore, India, Jan. 1999.
- [11] J. Padhye, V. Firoiu, and D. Towsley, "A stochastic model of TCP Reno congestion avoidance and control," Univ. of Massachusetts, Amherst, MA, CMPSCI Tech. Rep. 99-02, 1999.
- [12] *Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specification*, IEEE Std. 802.11, 1997.