FEATURE EXTRACTION FOR LOCALIZED CBIR
What You Click is What you Get

Steven Verstockt, Peter Lambert, and Rik Van de Walle
Department of Electronics and Information Systems – Multimedia Lab, Ghent University,
Gaston Crommenlaan 8, bus 201, B-9050 Ledeberg-Ghent, Belgium

Keywords: Object recognition, Feature extraction, Localized CBIR, Query by selection, SIFT.

Abstract: This paper addresses the problem of localized content based image retrieval. Contrary to classic CBIR systems which rely upon a global view of the image, localized CBIR only focuses on the portion of the image where the user is interested in, i.e. the relevant content. Using the proposed algorithm, it is possible to recognize an object by clicking on it. The algorithm starts with an automatic gamma correction and bilateral filtering. These pre-processing steps simplify the image segmentation. The segmentation itself uses dynamic region growing, starting from the click position. Contrary to the majority of segmentation techniques, region growing only focuses on that part of the image that contains the object. The remainder of the image is not investigated. This simplifies the recognition process, speeds up the segmentation, and increases the quality of the outcome. Following the region growing, the algorithm starts the recognition process, i.e., feature extraction and matching. Based on our requirements and the reported robustness in many state-of-the-art papers, the Scale Invariant Feature Transform (SIFT) approach is used. Extensive experimentation of our algorithm on three different datasets achieved a retrieval efficiency of approximately 80%.

1 INTRODUCTION

Content-based image retrieval has been a lively discipline and fast advancing research area during the past decade (Smeulders, 2000). CBIR systems use visual content such as color, texture, and simple shape properties to search images from large scale image databases (Del Bimbo, 1999). Although they improve text-based image retrieval systems, these systems are not yet a commercial success. One of the major reasons for this limited success is that CBIR rely upon a global view of the image, sometimes leading to a lot of irrelevant image content that is used in the search process. A solution for the global view problem can be found in localized CBIR. These systems only focus on the portion of the image the user is interested in.

The localized CBIR system described in this paper uses an interactive click and search technology to retrieve the relevant object. The proposed system can be used in a wide spectrum of application areas such as youtube-like video retrieval, search engines, and interactive television.

The test case used in our work is interactive advertising by logo recognition, i.e., a technique to extract logos or logo-like objects in digital images, so it can be used in a 'what you click is what you get' functionality on iDTV.

The remainder of this paper is organized as follows. Section 2 to 5 describes the major steps in our logo recognition algorithm, i.e., logo extraction, feature extraction, and matching. A scheme of this algorithm is showed in Figure 1. Section 5 reports on the performance results obtained from a set of experiments on 3 different datasets. Finally, section 6 concludes the paper and points out directions for future work.

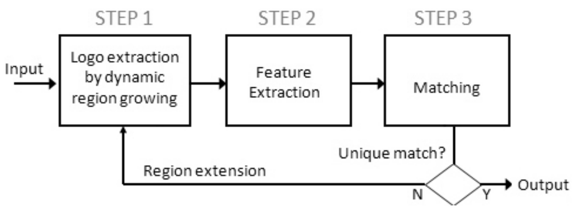


Figure 1: Logo recognition algorithm.

Verstockt S., Lambert P. and Van De Walle R. (2009).
FEATURE EXTRACTION FOR LOCALIZED CBIR - What You Click is What you Get.
In *Proceedings of the Fourth International Conference on Computer Vision Theory and Applications*, pages 371-374
DOI: 10.5220/0001755703710374

2 LOGO EXTRACTION

The first step in our logo recognition algorithm is logo extraction. The logo extraction itself can be further divided into five steps, as illustrated in Figure 2. The first three steps, i.e., contrast stretch, automatic gamma correction, and bilateral filtering, pre-process the input image to improve the quality of the subsequent steps in the logo extraction. Next, a dynamic region growing process searches the logo in the neighborhood of the click position. As soon as the region growing finishes, the algorithm makes an image crop, i.e., the result of our logo extraction.

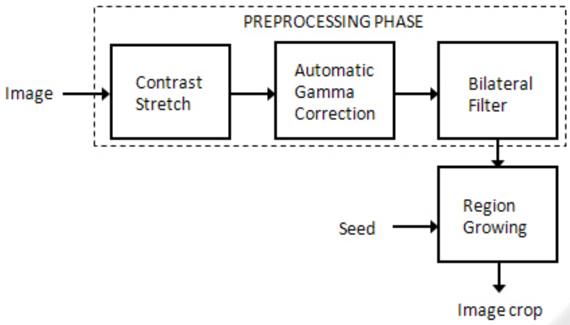


Figure 2: Logo extraction by region growing.

Automatic gamma correction

Since logo extraction works on all kinds of images, gamma correction needs to calculate a different gamma value for each individual image. To automatically generate an appropriate value we created an automatic gamma correction. This correction starts with an RGB to HSL color space conversion. Using the histogram of the HSL lightness component, our algorithm computes the mean and standard deviation of the lightness component. Based on these values the algorithm computes the gamma value using (Eq. 1).

$$
\begin{aligned}
&\texttt{if } \bar{l} > 0.5 \\
&\quad \gamma = 1 + \frac{|0.5-\bar{l}|}{\sigma_l} \\
&\texttt{else} \\
&\quad \gamma = \frac{1}{1+\frac{|0.5-\bar{l}|}{\sigma_l}}
\end{aligned}
\qquad (1)
$$

Bilateral filtering

Noise makes logo extraction more difficult. Many solutions, i.e., filters, exist in literature to remove image noise. However the majority of these filters have the undesirable side-effect of blurring the edges. For region growing these edges are very important and must be easily distinguishable.

Bilateral filtering prevents averaging across edges, while still averaging within smooth regions. It is a non-linear filtering technique introduced by (Tomasi and Manduchi, 1998). It extends the concept of Gaussian smoothing by weighting the Gaussian filter coefficients with their corresponding relative pixel intensities, i.e., combining gray levels or colors based on their geometric closeness and their photometric similarity.

Dynamic region growing

Since the proposed pre-processing operates on the global image, and, standard region growing grows unbounded, the computational cost can become very high and the retrieved image crop is sometimes too big to find a unique match. Dynamic region growing solves these problems since it investigates only a small part of the image, i.e., a 100 by 100 pixel area centered at the click position. If this small image crop contains enough information to retrieve a unique match, the dynamic region growing finishes. Otherwise the pixel area is extended and the logo extraction restarts.

3 FEATURE EXTRACTION

Many algorithms exist to recognize logo-like objects, e.g., global features, shape description- & matching techniques, and local features (invariant to transformations and variations). Based on our results and on studies comparing recognition techniques (Veltkamp, 2001), local features seem to perform best for solving real-life image matching problems.

Local feature-based image matching is usually done in two steps. The first step is the feature detection, i.e., keypoint or interest point detection. The second step involves computing descriptors for each detected interest point. These descriptors are then used for matching keypoints of the input image with keypoints in the logo database.

During the last decade, a lot of different detectors and descriptors have been proposed in literature (Mikolajczyk et al., 2004). For logo recognition the descriptor should be distinctive and at the same time robust to changes in viewing conditions. The feature should also be resilient to changes in illumination, image noise, uniform scaling, rotation, and minor changes in viewing direction. The descriptor should also minimize the probability of mismatch and finally, it should also be relatively easy and fast to extract the features and compare them against a large database of local features. Based on these requirements and the reported robustness in other research projects, Scale Invariant Feature Transform (SIFT) is used. A detailed description of SIFT can be found in the work of (Lowe, 2004).

4 FEATURE MATCHING

After generation of keypoint descriptors each of the resulting descriptors is compared with the keypoint descriptors in the logo database. The keypoint with minimum Euclidean distance, i.e., the closest neighbor, is selected in the database. However, this candidate is not necessarily matched because many features from an image may not have any correct match in the training database. For this reason the ratio of the distance with the closest neighbor and the distance with the second closest neighbor is computed. The match is rejected if this ratio exceeds the experimentally determined threshold of 0.6. Once comparison of descriptors is performed the database image with maximum number of matches is returned as recognized image.

When logo colors in the input image differ too much with colors of the corresponding logo in the database, feature matching fails. Proposed solutions in literature for making SIFT more invariant for background and color changes are color invariant CSIFT and Shape-SIFT (Abdel-Hakim and Farag, 2006). Although both methods seem to produce acceptable results, a much simpler method giving similar results is used. If no match can be found the image crop is inverted, the SIFT descriptors for the negative image crop are calculated, and, the matching restarts.

5 EVALUATION

The datasets used in the evaluation are obtained from three different sources: Caltech standard gadgets dataset, flickr.com dataset of top 5 relevant images for each of the 20 most famous advertisers in Flanders in 2008, and, real-life city pics taken with a mobile phone. Each image is annotated with 3 click positions per visible logo and for each of these click positions the region and the name of the logo that should be retrieved is specified. This forms the ground truth.

The evaluation uses two criterions: region coverage (Eq. 2) and retrieval efficiency (Eq. 3). Region coverage is used to measure the efficiency of dynamic region growing. Retrieval efficiency, i.e., a common measure for evaluating CBIR systems (Müller et al., 2001), is used to evaluate the outcome of the WYCIWYG system, i.e., the strongness of the proposed localized CBIR.

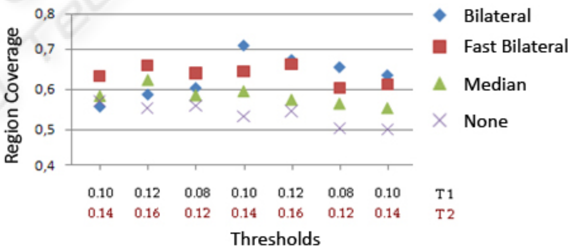
$$\text{Region coverage} = \frac{R1o}{R2n} \times \frac{R1o}{R1n} \quad (2)$$

R1o = overlapping pixels of the retrieved region
R2n = total number of pixels of the ground truth region
R1n = total number of pixels of the retrieved region

$$\text{Retrieval efficiency} = \frac{\text{No. correctly retrieved logos}}{\text{Total No. logo queries}} \quad (3)$$

The first experiment consists of determining the region growing thresholds and measuring the influence of four different pre-processing techniques: none, median filtering, bilateral filtering, and fast bilateral filtering (Paris and Durand, 2006). The efficiency is measured by the region coverage between the retrieved region and the region described in the ground truth.

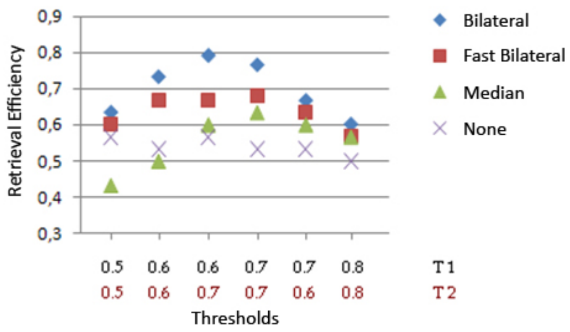
As can be seen in Graph 1, bilateral and fast bilateral filtering perform best. For optimal thresholds the region coverage of the bilateral filter exceeds 0.7, while without pre-processing the region growing only achieves a maximum coverage of 0.58.



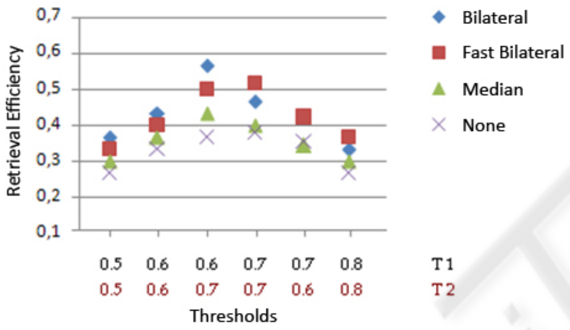
Graph 1: Region growing.

In the second test we measure the effect of the pre-processing on the retrieval efficiency of the whole system. During this test feature matching thresholds are varied to retrieve optimal values.

This test is not only covered for SIFT, but also the efficiency of SURF, i.e., speeded up robust features, is subject of this evaluation. SURF is a SIFT-like local feature descriptor introduced by (Bay et al., 2006) which has a similar performance as SIFT for classic CBIR. Graph 2 shows the retrieval efficiency of SIFT. Using optimal thresholds and the bilateral filter SIFT reaches a retrieval efficiency of approximately 80%. As can be seen in Graph 3 SURF's retrieval efficiency is much lower than the retrieval efficiency of SIFT. A reason for this can be found in the fact that the number of retrieved SURF feature vectors for the logo crops is sometimes too limited to find a match. Although for classic CBIR SIFT and SURF have a similar performance, SIFT improves SURF in our approach.



Graph 2: Retrieval efficiency of SIFT.



Graph 3: Retrieval efficiency of SURF.

6 CONCLUSIONS

An efficient localized content based image retrieval system has been developed to recognize logos and logo-like objects in digital images. The building blocks of the underlying algorithm are pre-processing by bilateral filtering, dynamic region growing, and, SIFT feature extraction and matching. WYCIWYG achieves a retrieval efficiency of approximately 80% over all the datasets. Using the proposed algorithm, it is possible to recognize a logo by clicking on it, as is illustrated in Figure 3. Even clicking in the neighborhood of the logo is sufficient to do successful logo retrieval.

WYCIWYG is implemented using MATLAB. Depending on the image size and the logo characteristics the average execution time is approximately 10 seconds on a Pentium M 740 with a default clock speed of 1.73 GHz. The memory usage is rather high, but 1GB should be sufficient. Further investigation will be carried out to accelerate the recognition process and decrease the memory usage.

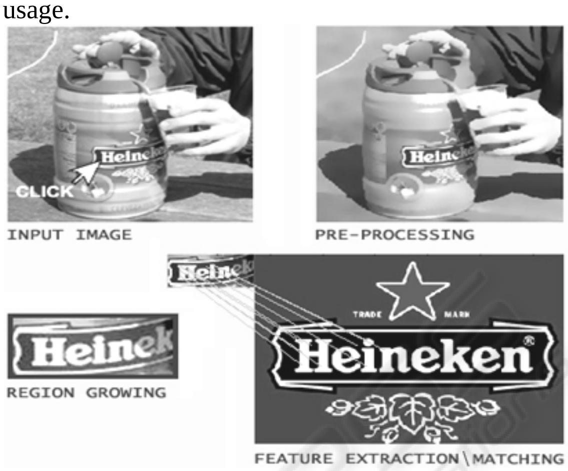


Figure 3: WYCIWYG demo.

REFERENCES

Abdel-Hakim, A.E., Farag, A.A., 2006. CSIFT: A SIFT Descriptor with Color Invariant Characteristics. In *CVPR, volume 2*, pages 1978-1983.

Bay, H., Tuytelaars, T., Van Gool, L., 2006. SURF: Speeded Up Robust Feature. In *ECCV*, pages 404-417.

Del Bimbo, A., 1999. *Visual Information Retrieval*, Morgan Kaufman. San Francisco, 1st edition.

Lowe, D.G., 2004. Distinctive image features from scale-invariant keypoints. *IJCV*, 60(2):91-110.

Mikolajczyk, K., Schmid, C., 2004. Comparison of affine-invariant local detectors and descriptors. In *EUSIPCO*, pages 69-81.

Müller, H., Müller, W., David McG. Squire, Marchand-Maillet, S., Pun, T., 2001. Performance evaluation in Content-Based Image retrieval: Overview and Proposals. *Pattern recognition letters*, 22(5):593-601.

Paris, S., Durand, F., 2006. A fast approximation of the bilateral filter using a signal processing approach. In *ECCV*, pages 568-580.

Smeulders, A., Worring, M., Santini, S., Gupta, A., Jain, R., 2000. Content-Based Image Retrieval at the End of the Early Years. *TPAMI*, 22(12):1349-1380.

Tomasi, C., Manduchi, R., 1998. Bilateral filtering for gray and color images. In *ICCV*, pages 839-846.

Veltkamp, R.C., Burkhardt, H., Kriegel, H.P., 2001. *State-of-the-Art in Content-Based Image and Video Retrieval*, Kluwer AP. Dordrecht, 1st edition.