

Enhanced Logo Matching and Recognition using SURF Descriptor

C. Aswini¹

¹Department of Computer Science and Engineering,
P. A. College of Engineering and Technology,
Pollachi, Tamil Nadu, India

D. Chitra²

²Department of Computer Science and Engineering,
P. A. College of Engineering and Technology,
Pollachi, Tamil Nadu, India

Abstract - Logos are graphically designed emblem used by business enterprises and organizations to aid and promote immediate public recognition. Individuality and uniqueness of a well-defined logo are necessary to avoid confusion among clients, suppliers and users. The proposed system implements Speeded-Up Robust Features (SURF) to extract local features from logos and to match the features. Interesting points on objects are extracted to provide a feature description of the object. The SURF method employs fast hessian detector to build integral images forming a stack. The use of integral images improves speed of logo matching. Extracted features from the logo and an image are matched using Euclidean distance measure. Experiments are done using MATLAB on a challenging dataset called Media Integration and Communication Center Logos that contains 13 classes of logos resulting in the collection of 720 images. Simulation results show increased accuracy in the recognition and computation time for feature extraction and matching is reduced.

Keywords - Fast Hessian Detector, Integral images, Local features, SURF.

I. INTRODUCTION

Computer vision deals with modeling and replication of human vision with the help of computer hardware and software. Functions in computer vision are pre-processing, image acquisition, detection, segmentation, high-level processing, feature extraction and decision making. Classical problem in computer vision is determining the image data containing some specific object, feature and activity. Logos are purely graphic symbols composed of name of the organization. Logo detection [16] and recognition has become significant in number of scenarios. Some logos are corrupted by light effects, noise and partial occlusions. Object recognition is the process of determining identity of an object being observed in image from a set of known labels. Most object recognition systems use either global or local features exclusively. A feature is defined as an interesting part of an image and features are used as a starting point for many computer vision algorithms.

Extracted features may be surface patches, corners and linear edges [4], [26]. Feature based methods are used to find feasible matches between logo features and image features. Feature extracted from logos are detectable even under changes in image scale, noise, illumination and light effects to perform efficient recognition. Such images and logos have low resolution and quality. It is more challenging to extract and match the features under such conditions.

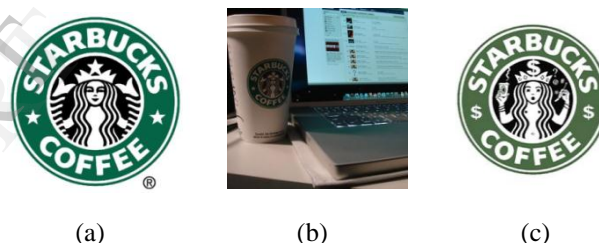


Fig. 1 (a) Starbucks logo (b) Logo in bad light conditions (c) Logo with small changes

The desirable property for a feature detector is its repeatability to find same feature in two or more different images of same scene [8]. Features are extracted from both logo and an image using SURF descriptors [25]. SURF descriptors are efficient in repeatability, distinctiveness and robustness. Features are matched using distance measures that can be computed and compared faster. Matching is based on number of features matched. If number of features matched is above threshold value, the given image containing logo is authorized. If number of feature matches is below the threshold, logo is not recognized [7], [19].

II. RELATED WORK

A large number of object matching and recognition techniques have been proposed. Transformation of images into its local features [6] is efficient for matching and recognition process since these features provide essential information such

as orientation and gradient. Visual feature extraction is a research topic that has received much interest in recent years. Among extraction techniques Scale Invariant Feature Transform (SIFT) is the most widely adopted approach that provides stable visual interesting points from objects for reliable and robust object detection. SIFT [5] detects and describes local feature points in images for object detection in different scenes. These feature points describe strength and direction of object. Feature points are matched using distance measures.

Lowe et al [4] and Sivic et al [14] presented a method for extracting distinctive invariant features from images that can be used to perform reliable matching between different views of a scene. Features are invariant to image scale and rotation and provide robust matching across a substantial range of affine distortion, change in 3D viewpoint, addition of noise and change in illumination. Features are highly distinctive for object recognition. Recognition proceeds by matching individual features to a database of features from known objects using a fast Nearest Neighbor (NN) algorithm.

A method for trademark image database retrieval based on object shape information is proposed by Jain et al [5], Belongie et al [21] and Ballan et al [17]. Two stages achieve both desired efficiency and accuracy. In first stage, shape features are used to browse through the database to generate a moderate number of plausible retrievals if a query is presented. In second stage, candidates from first stage are screened using a template matching process to discard incorrect matches.

A semi-automatic segmentation is proposed by Shih et al [11] to extract shapes of representative objects called masks. Features like invariant moments and histogram of edge directions are selected to describe the mask. Moments are invariant to scaling, rotation and translation. Based on rank of feature distance a grade evaluation method is provided for similar trademark retrieval. A feedback algorithm is to automatically determine weight of each feature.

Various types of object recognition and detection algorithms proposed by Merler et al [18] are color histogram matching, SIFT matching and boosted Haar-like features. Histograms of 16 bins are calculated separately for a total of 32 bins. Histograms belonging to same product are subsequently averaged to obtain a final template histogram representative of object. SIFT key-points are generated for every in-vitro image in data set to represent images using scale and rotation invariant descriptors.

Quack et al [23] and Lazebnik et al [22] presented a novel approach to automatically find spatial configurations of local features occurring frequently on instances of a given object class and rarely on background. The lowest layer is built on a set of local features extracted in each image. Difference of Gaussian (DoG) detector is used to extract regions and SIFT descriptor to describe appearance. SIFT feature vectors are clustered into an appearance codebook with a hierarchical agglomerative clustering method.

A semi-automatic system for detecting and retrieving trade-mark appearances in sports videos is proposed by Bagdanov et al [1] and Watve et al [3]. The original Motion Pictures Expert Group (MPEG) videos are sub-sampled and SIFT points are detected at 5fps. Visual quality is estimated

measuring blurriness, number of edges and number of SIFT points. Local neighborhood descriptors of salient points are used to obtain a matching technique that is robust to partial occlusions.

Wu et al [26] proposed a novel scheme so that image features are bundled into local groups. Each group of bundled features becomes much more discriminative than a single feature. SIFT descriptor assembles a 4x4 array of 8 gradient orientation histograms around keypoint making it robust to image variations induced by both photometric and geometric changes. Maximally Stable Extremal Region [13] (MSER) detects affine-covariant stable regions.

Sahbi et al [9] and Kleban et al [10] proposed a novel variational framework able to match and recognize multiple instances of multiple reference logos in image archives. Reference logos and test images are seen as constellations of local features and matched by minimizing an energy function mixing. Context is used to find interest point correspondences between two images to tackle logo detection. Adjacency matrices [2] are defined in order to model spatial and geometric relationships between interest points belonging to two images.

III. SURF FEATURE DESCRIPTOR FOR DIMENSIONALITY REDUCTION

The proposed algorithm uses SURF feature descriptor for logo recognition. Feature vectors that are invariant to image scaling and rotation are extracted by SURF. Features are matched by sign of Laplacian. SURF local descriptors have better computational efficiency than standard SIFT local descriptors because of integral images computed in SURF.

Interest points are selected at discrete locations in the image, such as corners, blobs, and T-junctions. Neighborhood of every key point is represented by a feature vector. The feature descriptor has to be distinctive, robust to noise, detection errors, geometric and photometric deformations. Finally the SURF descriptor vectors are matched between different images. The matching is based on sign of Laplacian. Different stages of SURF algorithm are performed to build feature space. Stages of SURF algorithm are interest point detection, constructing SURF descriptor for each key point and descriptor matching.

A. Constructing Integral image

The use of calculating integral images in SURF is its speed. Integral image is an intermediate representation and consists of sum of pixel values of the image. Integral images are also called as Summed Area Tables. Integral image is given by (1).

$$U(x, y) = \sum_{i \leq x, j \leq y} u(i, j) \quad (1)$$

B. Interest Point Detection

SURF uses Fast Hessian feature detector. The fast Hessian is based on the determinant of Hessian matrix $H(f)$. Hessian matrix consists of partial derivatives of two dimensional function $f(i, j)$ given by (2).

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

The Hessian matrix determinant is calculated by

$$\det(H) = \frac{\partial^2 f}{\partial x^2} \frac{\partial^2 f}{\partial y^2} - \left(\frac{\partial^2 f}{\partial x \partial y} \right)^2 \quad (3)$$

C. SURF Descriptor

A circular region around the detected interest points is constructed to assign a unique orientation and to gain invariance to image rotations. The Haar wavelet responses are used to compute orientation in both x and y directions. The Haar wavelets [25] can be quickly computed through integral images. In the next step, SURF feature descriptors are constructed by extracting square regions around the interest points. The windows are split up into 4 x 4 sub-regions to retain spatial information. In each sub-region, Haar wavelets are extracted at regularly spaced sample keypoints. The wavelet responses in both horizontal and vertical directions are summed up over each sub-region. The responses results in the descriptor vector for all 4 x 4 sub-regions of length 64 resulting in the standard SURF descriptor.

D. Feature Matching

The feature matching speed is achieved by a single step of indexing based on sign of the Laplacian of interest point. The sign of the Laplacian distinguishes bright blobs from a dark background. Bright interest points are matched against other bright interest points and also for the dark interest points. The process minimizes the matching speed and it has no computational costs.

IV. PERFORMANCE EVALUATION

A. Logo Detection Performance

Logo detection is achieved by finding interesting points in the reference logo and in a test image and to match these vectors using Euclidean distance measure. The performance of logo detection is evaluated by measuring False Acceptance and False Rejection Rates.

$$FAR = \frac{\text{No of incorrect logo detection}}{\text{No of logo detections}} \quad (4)$$

$$FRR = \frac{\text{No of unrecognized logo appearance}}{\text{No of logo appearances}} \quad (5)$$

B. Simulation Results

A number of object matching and recognition experiments are performed on novel challenging dataset called Media Integration and Communication Center Logos Logos dataset.

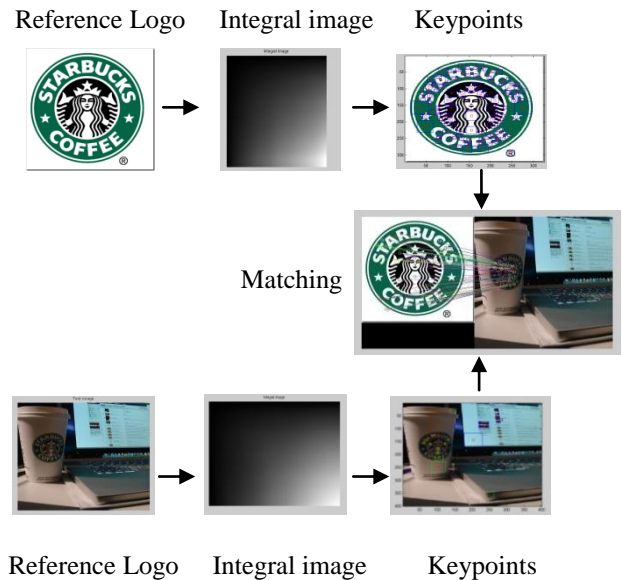


Fig. 2 Working of SURF Algorithm

Input logo and an image are given as input. Features extracted using SURF are matched. If matched points are above threshold and logo is recognized as shown in Fig. 2.



Fig. 3 Logo matching of Reference logo and an Input image

The twelve classes of logos in a dataset and an input image are chosen randomly except apple class. These features are extracted using SURF descriptor. These features are matched and logos are recognized.

TABLE1
Simulation Time

Logo	PCA-SIFT Computation time(s)	SURF Computation time(s)
Agip	9.92	7.97
Barilla	7.94	5.89
Birra moretti	4.79	3.40
Coca-cola	8.27	7.17
Cinzano	4.92	3.51
Esso	9.30	6.15
Ferrari	5.68	2.94
Heineken	6.50	3.98
Marlboro	7.92	4.62
McDonald's	4.51	2.99
Pepsi	10.04	7.43
Starbucks	7.04	4.51

Computation time for logos shown in Fig. 3 using SURF and PCA-SIFT is listed in the Table 1. SURF reduces the computation time comparing to PCA-SIFT algorithm. SURF descriptor outperforms previous existing methods when images are rotated to different angles and also under light effects. Apple class of logos is matched with less accuracy due to its large size and illumination changes.

Fig. 4 shows the comparison of accuracy of PCA-SIFT and proposed SURF. Accuracy rate has performed well compared to existing PCA-SIFT method.

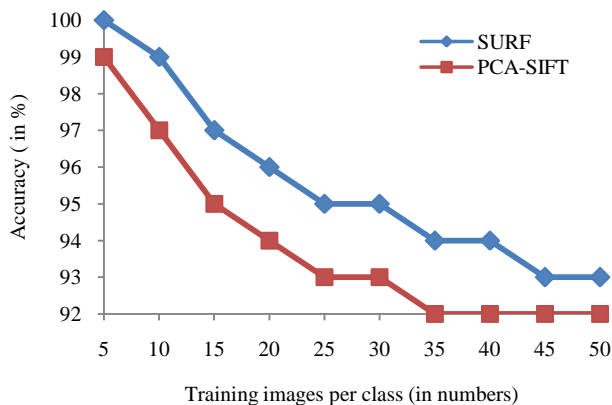
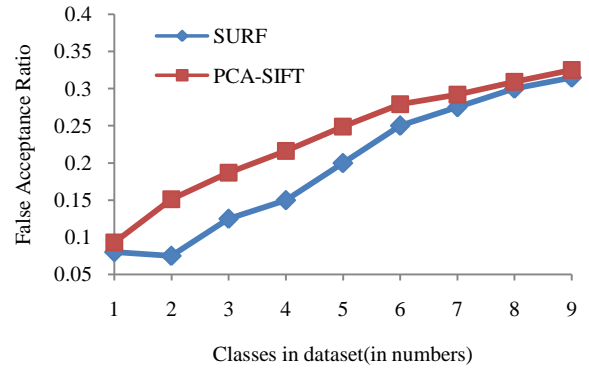


Fig. 4 Number of training images per classes Vs Accuracy

False acceptance ratios of both PCA-SIFT and SURF algorithms are shown in Fig. 5. Number of incorrect logo detections per total number of logo detections is minimized in proposed method.



per total number of appearances compared to earlier PCA-SIFT method. The measure is shown in Fig. 6.

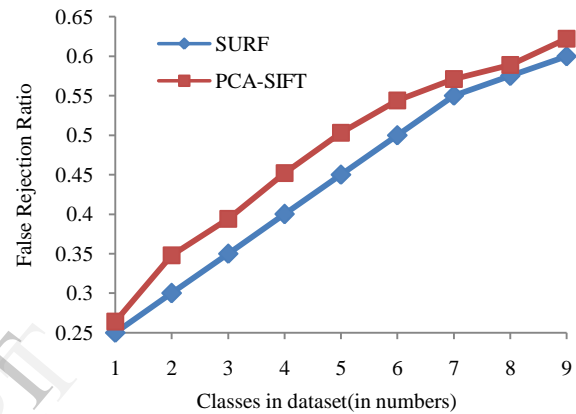


Fig. 6 Number of training images per classes Vs False Rejection Ratio

Table 2 shows the comparison of various existing methods with SURF descriptors. SURF method provides logo detection of accuracy 95%.

TABLE 2
Comparison of accuracy

Title	Year	Accuracy
Spatial pyramid mining for logo detection in natural scenes	2008	80%
Efficient visual search of videos cast as text retrieval	2009	81%
Context-Dependent Logo Matching and Recognition	2013	89%
Enhanced Logo Matching and Recognition using local feature descriptors	2013	92%
Enhanced Logo matching and Recognition using SURF Descriptor	2014	95%

V. CONCLUSION

The approach to speed up feature extraction in logo recognition is considered. The features extracted using SURF algorithm are more distinctive, more robust to image deformations and more compact than standard SIFT representation. The proposed SURF algorithm is quite simple and compact. The result shows that using these descriptors in logo matching results in increased accuracy and faster matching. SURF feature vector is significantly smaller than standard SIFT feature vector and can be used with the matching algorithms. Euclidean distance between two feature vectors is used to determine the two vectors correspond to same keypoint in different images. The future work is to apply Principal Component Analysis to Speeded Up Robust Feature descriptor. SURF can be extended to other object recognition problems such as object tracking and object matching with large datasets.

REFERENCES

- [1] A. D. Bagdanov, L. Ballan, M. Bertini, and A. Del Bimbo, "Trademark matching and retrieval in sports video databases," in *Proc. ACM Int. Workshop Multimedia Inf. Retr.*, Augsburg, Germany, 2007, pp. 79–86.
- [2] A. Joly and O. Buisson, "Logo retrieval with a contrario visual query expansion," in *Proc. ACM Multimedia*, Beijing, China, 2009, pp. 581–584.
- [3] A. Watve and S. Sural, "Soccer video processing for the detection of advertisement billboards," *Pattern Recognit. Lett.*, vol. 29, no. 7, pp. 994–1006, 2008.
- [4] D. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.
- [5] E. Mortensen, H. Deng, and L. Shapiro, "A SIFT descriptor with global context," in *Proc. Conf. Comput. Vis. Pattern Recognit.*, San Diego, CA, 2005, pp. 184–190.
- [6] G. Carneiro and A. Jepson, "Flexible spatial models for grouping local image features," in *Proc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2. Washington, DC, 2004, pp. 747–754.
- [7] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (SURF)," *Comput. Vis. Image Understand.*, vol. 110, no. 3, pp. 346–359, 2008.
- [8] H. Sahbi, J.-Y. Audibert, and R. Kerivan, "Context-dependent kernels for object classification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 4, pp. 699–708, Apr. 2011.
- [9] H. Sahbi, L. Ballan, G. Serra, and A. Del Bimbo, "Context-Dependent Logo Matching and Recognition," *IEEE Trans. Image Process.*, vol. 22, no. 3, March 2013.
- [10] J. Kleban, X. Xie, and W.-Y. Ma, "Spatial pyramid mining for logo detection in natural scenes," in *Proc. IEEE Int. Conf. Multimedia Expo*, Hannover, Germany, 2008, pp. 1077–1080.
- [11] J. L. Shih and L.-H. Chen, "A new system for trademark segmentation and retrieval," *Image Vis. Comput.*, vol. 19, no. 13, pp. 1011–1018, 2001.
- [12] J. Luo and D. Crandall, "Color object detection using spatial-color joint probability functions," *IEEE Trans. Image Process.*, vol. 15, no. 6, pp. 1443–1453, Jun. 2006.
- [13] J. Matas, O. Chum, M. Urban, and T. Pajdla, "Robust wide-baseline stereo from maximally stable extremal regions," *Image Vis. Comput.*, vol. 22, no. 10, pp. 761–767, 2004.
- [14] J. Sivic and A. Zisserman, "Efficient visual search of videos cast as text retrieval," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 4, pp. 591–606, Apr. 2009.
- [15] Jain and A. Vailaya, "Shape-based retrieval: A case study with trademark image databases," *Pattern Recognit.*, vol. 31, no. 9, pp. 1369–1390, 1998.
- [16] K. Gao, S. Lin, Y. Zhang, S. Tang, and D. Zhang, "Logo detection based on spatial-spectral saliency and partial spatial context," in *Proc. IEEE Int. Conf. Multimedia Expo*, 2009, pp. 322–329.
- [17] L. Ballan, M. Bertini, and A. Jain, "A system for automatic detection and recognition of advertising trademarks in sports videos," in *Proc. ACM Multimedia*, Vancouver, BC, Canada, 2008, pp. 991–992.
- [18] M. Merler, C. Galleguillos, and S. Belongie, "Recognizing groceries in situ using in vitro training data," in *Proc. IEEE Comput. Vis. Pattern Recognit. SLAM Workshop*, Minneapolis, MN, May 2007, pp. 1–8.
- [19] O. Chum and J. Matas, "Unsupervised discovery of co-occurrence in sparse high dimensional data," in *Proc. Conf. Comput. Vis. Pattern Recognit.*, San Francisco, CA, 2010, pp. 3416–3423.
- [20] R. Fergus, P. Perona, and A. Zisserman, "Object class recognition by unsupervised scale-invariant learning," in *Proc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2. Madison, WI, 2003, pp. 264–271.
- [21] S. Belongie, J. Malik, and J. Puzicha, "Shape matching and object recognition using shape contexts," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 4, pp. 509–522, Apr. 2002.
- [22] S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories," in *Proc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2. 2006, pp. 2169–2178.
- [23] T. Quack, V. Ferrari, B. Leibe, and L. Van Gool, "Efficient mining of frequent and distinctive feature configurations," in *Proc. Int. Conf. Comput. Vis.*, Rio de Janeiro, Brazil, 2007, pp. 1–8.
- [24] Y. Kalantidis, L. G. Pueyo, M. Trevisiol, R. van Zwol, and Y. Avrithis, "Scalable triangulation-based logo recognition," in *Proc. ACM Int. Conf. Multimedia Retr.*, Trento, Italy, 2011, pp. 1–7.
- [25] Y. S. Kim and W. Y. Kim, "Content-based trademark retrieval system using visually salient feature," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, San Juan, Puerto Rico, Mar. 1997, pp. 307–312.
- [26] Z. Wu, Q. Ke, M. Isard, and J. Sun, "Bundling features for large scale partial-duplicate web image search," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Miami, FL, Mar. 2009, pp. 25–32.