Logo Matching and Detection for Document Image Recognition

Dr.B.B.M.KrishnaKanth
Principal, Hindu College of Engineering & Technology,Amaravathi Road, Guntur,A.P,India-522002.
Mail:bbkkanth@yahoo.com

Abstract: **Graphics detection and recognition are fundamental research problems in document image analysis and retrieval. logo detection and recognition continues to be of great interest to the document retrieval community as it enables effective identification of the source of a document. We contribute, through this paper, to the design of a novel variation 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 (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. We also introduce a detection/recognition procedure and study its theoretical consistency.**

***Index Terms*—Context-dependent kernel, logo detection, logo recognition.**

INTRODUCTION

Logos are often used pervasively as declaration of document source and ownership in business and government documents. The problem of logo detection and recognition is of great interest to the document image analysis and retrieval communities because it enables immediate identification of the source of documents based on the originating organization.

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 [see Fig. 1(a)]. Color 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 [see Fig. 1(b)]. 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 [see Fig. 1(c)]. Logo detection and recognition in these scenarios has become important for a number of applications. Among them, several examples have been reported in the literature, such as the automatic identification of products on the web to improve commercial search-engines [4], the verification of the visibility of advertising logos in sports events [5]–[7],

Figure 1:Some Basic Logo Structures.

Fig. 1. (a) Examples of popular logos depicting real world objects, text, graphic signs, and complex layouts with graphic details. (b) Pairs of logos with malicious small changes in details or spatial arrangements. (c) Examples of logos displayed in real world images in bad light conditions, with partial occlusions and deformations.

the detection of near-duplicate logos and unauthorized uses [8], [9]. Special applications of social utility have also been reported such as the recognition of groceries in stores for assisting the blind [10].

II STATE OF REVIEW OF ART

EXISTING SYSTEM

The previous work on "Shape matching and object recognition using shape contexts," and "ANSIG—An analytic signature for permutation-invariant two-dimensional shape representation," have used different global descriptors of the full logo image either accounting for logo contours or exploiting shape descriptors such as shape context. A two-stage algorithm proposed in "Logo detection based on spatial-spectral saliency and partial spatial context," that accounts for local contexts of key points. They considered spatial-spectral saliency to avoid the impact of cluttered background and speed up the logo detection and localization. These methods assume that a logo picture is fully visible in the image, is not corrupted by noise and is not subjected to transformations. According to this, they cannot be applied to real world images. The major limitation of this approach is image resolution and their solution has revealed to be very sensitive to occlusions.

PROPOSED SYSTEM

In this paper, we present a novel solution for logo detection and recognition which is based on the definition of a "Context-Dependent Similarity" (CDS) kernel that directly incorporates the spatial context of local features. The proposed method is model-free, i.e. it is not restricted to any a priori alignment model. Context is considered with respect to each single SIFT key point Formally, the CDS function is defined as the fixed-point of three terms: (i) an energy function which balances The fidelity term is inversely proportional to the expectation of the Euclidean distance between the most likely aligned interest points. The context criterion measures the spatial coherence of the alignments. The "entropy" term acts as a smoothing factor, assuming that with no a priori knowledge, the joint probability distribution of alignment scores is flat.

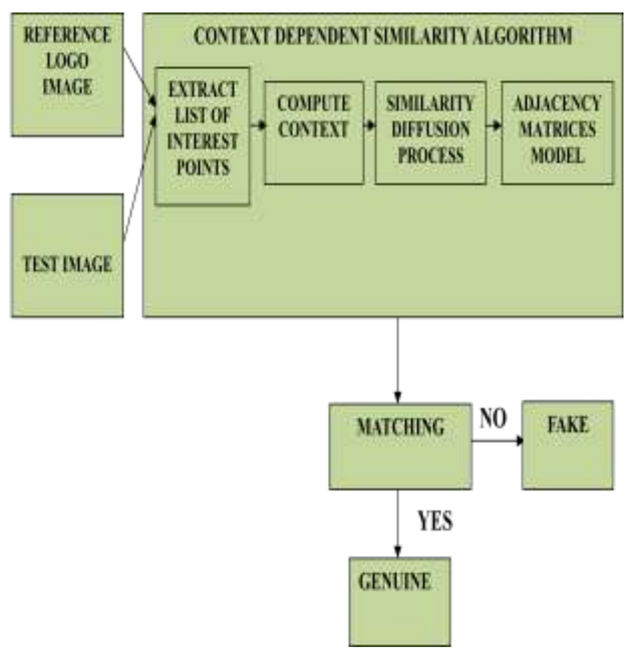


Fig: Proposed System Block Diagram.

In this paper, we present a novel solution for logo detection and recognition which is based on the definition of a "Context Dependent Similarity" (CDS) kernel that directly incorporates the spatial context of local features. The proposed method is model-free, i.e. it is not restricted to any a priori alignment model. Context is considered with respect to each single SIFT keypoint and its definition recalls shape context with some important differences: given a set of SIFT interest points X, the context of $x \in X$ is defined as the set of points spatially close to x with particular geometrical constraints. Formally, the CDS function is defined as the fixed-point of three terms: (i) an energy function which balances a fidelity term; (ii) a context criterion; (iii) an entropy term. The fidelity term is inversely proportional to the expectation of the Euclidean distance between the most likely aligned interest points. The context criterion measures the spatial coherence of the alignments: given a pair of interest points (f_p, f_q) respectively in the query and target image with a high alignment score, the context criterion is proportional to the alignment scores of all the pairs close to (f_p, f_q) but with a given spatial configuration. The "entropy" term acts as a smoothing factor, assuming that with no a priori knowledge, the joint probability distribution of alignment scores is flat. It acts as a regularizer that controls the entropy of the conditional probability of matching, hence the uncertainty and decision thresholds so helping to find a direct analytic solution. Using the CDS kernel, the geometric layout of local regions can be compared across images which show contiguous and repeating local structures as often in the case of graphic logos. The solution is proved to be highly effective and responds to the requirements of logo detection and recognition in real world images.

FEATURE SELECTION AND EXTRACTION

Extracting robust and generic features that can be detected reliably is essential for matching as logos often appear as complex mixtures of graphics and formatted text. We extract corner features from detected logos as follows. We first extract the object contours from the edge image computed by the Canny edge detector [4] and fill in the gaps along the contours. We then use the corner detector of He and Yung [10]. It has shown excellent performance in applications involving real-world scenes compared to other popular feature detectors. It identifies an initial set of corner candidates from local curvature maxima and uses adaptive local thresholds and dynamic support regions to eliminate false corners .Measures of Shape Dissimilarity Several measures of shape dissimilarity have demonstrated success in object recognition and retrieval.

CONTEXT-DEPENDENT SIMILARITY

Let $SX = \{x1, \dots, xn\}$, $SY = \{y1, \dots, ym\}$ be respectively the list of interest points taken from a reference logo and a test image (the value of n, m may vary with SX , SY). We borrow the definition of context and similarity design from, in order to introduce a new matching procedure applied to logo detection. The main differences with respect to reside in the following.

1) *The use of context for matching:* Context is used to find interest point correspondences between two images in order to tackle logo detection while in , context was used for kernel design in order to handle object classification using support vector machines.

2) *The update of the design model:* Adjacency matrices are defined in order to model spatial and geometric relationships (context) between interest points belonging to two images (a reference logo and a test image). These adjacency matrices model interactions between interest points at different orientations and locations resulting into an anisotropic context, while in , context was isotropic.

3) *The similarity diffusion process:* Resulting from the definition of context, similarity between interest points is recursively and anisotropically diffused.

4) *The interpretation of the model:* Our designed similarity may be interpreted as a joint distribution (pdf) which models the probability that two interest points taken from $SX \times SY$ match. In order to guarantee that this similarity is actually a pdf, a partition function is used as a normalization factor taken through all the interest points in $SX \times SY$ (and not over all the objects in a training database as in [40]).

LOGO DETECTION AND RECOGNITION

Application of CDS to logo detection and recognition requires establishing a matching criterion and verifying its probability of success. Let $R \subset \mathrm{R}^2 \times \mathrm{R}^{128} \times [-\pi,+\pi] \times \mathrm{R}^+$ denote the set of interest points extracted from all the possible reference logo images and X a random variable standing for interest points in R. Similarly, we define $T \subset \mathrm{R}^2 \times \mathrm{R}^{128} \times [-\pi,+\pi] \times \mathrm{R}+$ as the set of interest points extracted from all the possible test images (either including logos or not) and Y a random variable standing for interest points in T . X and Y are assumed drawn from existing (but unknown) probability distributions. Let's consider $SX = \{X_1, \dots, X_n\}$, $S_Y = \{Y_1, \dots, Y_m\}$ as n and m realizations with the same distribution as X and Y respectively. To avoid false matches we have assumed that matching between YJ and X is assessed iff

$$K_{YJ/X} \geq \sum_{i \neq J}^{m} K_{Y_j}/X$$

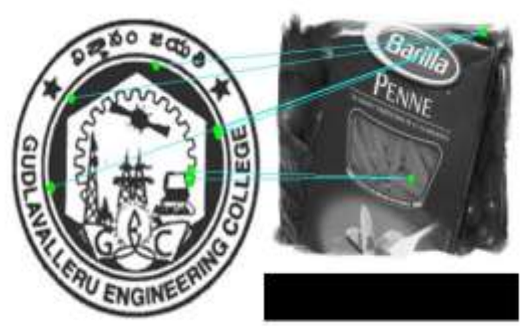
The intuition behind the strong criterion above comes from the fact that when $\mathbf{K}_{YJ\ |X} >> \mathbf{K}_{Yj}\ |X$, the entropy of the conditional probability distribution $\mathbf{K}.|X$ will be close to 0, so the uncertainty about the possible matches of X will be reduced. The reference logo SX is declared as present into the test image if, after that the match in SY has been found for each interest point of SX , the number of matches is sufficiently large (at least $\tau\ |SX|$ for a fixed $\tau \in [0, 1]$, being $1 - \tau$ the occlusion factor tolerated). We summarize the full procedure for logo detection and recognition .

RESULTS AND DISCUSSION

Firstly, we compare our proposed CDS matching and detection procedure against nearest-neighbor SIFTS matching and nearest-neighbor matching with RANSAC verification. SIFT based logo detection follows the idea in [26] where a reference logo is detected, in a test image, if the overall number of SIFT matches is above a fixed threshold. SIFT matches are obtained by computing for each interest point in S_X its Euclidean distance to all interest points in S_Y , and keeping only the nearest-neighbors. RANSAC based logo detection follows the same idea but it introduces a model (transformation) based criterion not necessarily consistent in practice. This criterion selects only the matches that satisfy an affine transformation between reference logos and test images. The (iterative) RANSAC matching process, is applied as a "refinement" of SIFT matching (a similar approach is used in. In both cases a match is declared as present iff Lowe's second nearest neighbor test is satisfied . Secondly, we also compare our CDS logo detection algorithm to two relevant methods that use context in their matching procedure . The Video Google approach is closely related to our method as it introduces a spatial consistency criterion, according to which only the matches which share similar spatial layouts are selected. The spatial layout (context) of a given interest point includes 15 nearest neighbors that are spatially close to it. Given $X \in S_X$, $Y \in S_Y$, points in the layouts of X and Y which also match casts a vote for the final matching score between X and Y . The basic idea is therefore similar to ours, but the main difference resides in the definition of context in Video Google which is strictly local.2 In our method the context is also local but recursive; two interest points match if their local neighbors match, and if the neighbors of their local neighbors match too, etc, resulting into a recursive diffusion of the similarity through the context. Partial Spatial Context (PSC) logo matching relies on a similar context definition. Given a set of matching interest points, it formulates the spatial distribution for this set (i) by selecting a circular region that contains all these points, (ii) by computing the scale and orientation of the set as the average value of, respectively, all the scales and orientations of the points, (iii) by partitioning the distribution of these points in 9 cells. Starting from this context definition, PSC histograms are computed for both reference logos and test images. A PSC histogram is defined as the number of matches lying in each cell, and logo matching is performed by computing the similarity between two PSC histograms. This schema is efficient and quick to be computed, but its spatial (context) definition is rough and is very sensible to outliers.

(a)

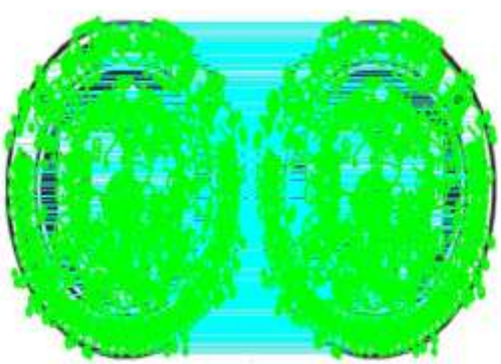
(b)



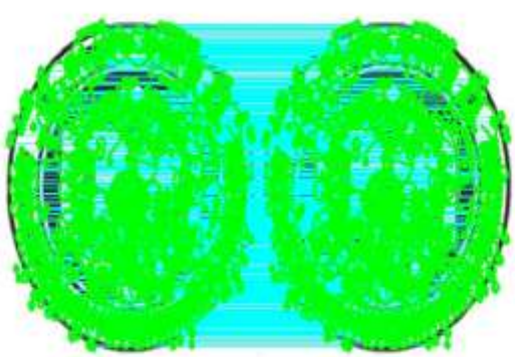
(c)

(d)

(e)



(f)



(h)

Fig. 2(a,b,c,d,e,f,g,h). Some examples of logo detection results.

Computational Cost of our logo detection procedure is mainly dominated by CDS evaluation. In particular, the key part of the algorithm is the computation of the context term. Assuming $\mathbf{K}(t-1)$ known for a given pair of points (x, y), the complexity is $O(\max(N2, s))$; here s is the dimension of $\psi f\ (x)$ (i.e. 128 since we use SIFT features) and N is given by the $\max_{x,\theta,\rho} \#\{N^{\theta,\rho}(x)\}$ (i.e. the max number of points in all the neighborhoods). When $N < \sqrt{s}$, evaluating our CDS is equivalent to efficient kernels such as linear or intersection. In worst cases $N >> \sqrt{s}$ and the evaluation of CDS should be prohibitive. In practice it may only happen when the context is too large. Anyway, using the same setting for CDS used in the previous experiments, our method is able to process images and checks for the existence of a reference logo in less than 1 s. This running time is achieved, on average on our MICC-Logos dataset, on a standard 2.6 GHZ PC with 2 GB memory.

CONCLUSION

We introduced in this work a novel logo detection and localization approach based on a new class of similarities referred to as context dependent. The strength of the proposed method resides in several aspects: (i) the inclusion of the information about the spatial configuration in similarity design as well as visual features, (ii) the ability to control the influence of the context and the regularization of the solution via our energy function, (iii) the tolerance to different aspects including partial occlusion, makes it suitable to detect both near-duplicate logos as well as logos with some variability in their appearance, and (iv) the theoretical groundedness of the matching framework which shows that under the hypothesis of existence of a reference logo into a test image, the probability of success of matching and detection is high.

REFERENCES

[1] A. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, “Contentbased image retrieval at the end of the early years,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 12, pp. 1349–1380, Dec. 2000.

[2] 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.

[3] L. Ballan, M. Bertini, A. Del Bimbo, L. Seidenari, and G. Serra, “Event detection and recognition for semantic annotation of video,” *Multimedia Tools Appl.*, vol. 51, no. 1, pp. 279–302, 2011.

[4] Y. Jing and S. Baluja, “Pagerank for product image search,” in *Proc.*
WWW, Beijing, China, 2008, pp. 307–316.

[5] 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.

[6] 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.
[7] C. Constantinopoulos, E. Meinhardt-Llopis, Y. Liu, and V. Caselles, “A robust pipeline for logo detection,” in *Proc. IEEE Int. Conf. Multimedia Expo*, Barcelona, Spain, Jul. 2011, pp. 1–6.
[8] 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.
[9] C.-H. Wei, Y. Li, W.-Y. Chau, and C.-T. Li, “Trademark image retrieval using synthetic features for describing global shape and interior structure,” *Pattern Recognit.*, vol. 42, no. 3, pp. 386–394, 2009.
[10] 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.
[11] J. Sivic and A. Zisserman. Video Google: a text retrieval approach to object matching in videos. International Conference on Computer Vision, 2003.
[12] H. Wolfson and I. Rigoutsos. Geometric hashing: An overview. IEEE Computational Science & Engineering, 4(4), 1997.
[13] Z. Wu, Q. Ke, M. Isard, and J. Sun. Bundling features for large scale partial-duplicate web image search. In IEEE Conference on Computer Vision and Pattern Recognition, 2009.

