IJREAT International Journal of Research in Engineering & Advanced Technology, Volume 2, Issue 2, Apr-May, 2014
ISSN: 2320 – 8791 (Impact Factor: 1.479)
www.ijreat.org

Logo Matching and Recognition in 2D View Based on Context-Dependent Similarities

S.Portchselvi[1], A.Rajalakshmi[2],G.Appasami[3]

[1, 2, 3]**Department of Computer Science and Engineering, Anna University Chennai, Dr. Pauls Engineering College, Villupuram, Tamilnadu, India**

Abstract

In this paper we made a design of a novel variational framework which used for matching and recognizing of multiple reference logos present in image archives. Logos are matched depending on a situation of real world. Here the reference logos and test images are matched by its local features like interest points and region which is seen as a constellation. Matching can be done by minimizing the energy function like: a) fidelity term which makes the exact measures of quality in feature matching, b) the reference and test images which captures the feature geometry, and c) term which is regular will maintain the smoothness of matching. We additionally introduced detection and recognition procedures. Validity of our paper will be explained through MICC-Logos dataset.

Keywords: *Matching, Recognition, Content dependant similarities, logo images, minimizing energy functions, occlusion reduction.*

1. Introduction

In Real world the production of visual data from companies, institution, individuals and the increasing popularity of social system like YouTube, Face book and Twitter for sharing of images and video. There are more urged research in finding effective solution for object detection and recognition for automatic annotation of images and videos [1].Graphic logos are a special class of visual objects which is important to assess the identity of something or someone. Some economical relevant as motivated the active involvements of the companies in soliciting smart image analysis solutions to scan logo archives to find evidence of similar already existing logos. This is because malicious uses of the logos have small variations with respect to the originals. Because of this we can discover improper non-authorized use of their logos.

Logos are graphic productions that may recalls some real world objects or some name, or simply display some abstract signs. Color which plays important role in image process for logo identity. But distinctiveness of logos are measured and carefully studied by the graphic designer, semiologists and experts of social communications. The graphic layout play important role to attract the attention of the customer. Different logos may have similar layout with slightly different spatial disposition of the graphic elements which may differ in their orientation, size and shape. They may differ in few or more traits [see Fig. 1(b)

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.

Logos which appear in images/videos of real world indoor or outdoor scenes superimposed on objects of any geometry for example 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. Occlusion which means hidden images for example: if two persons crossing each other the person on other side is not visible. Such images – and logos

thereafter – have often relatively low resolution and quality. Regions that include logos might be small and contain little information [see Fig. 1(c)]. Logo detection and recognition in these scenarios has become important for a number of applications. Among them it include

automatic identification of products on the web to improve commercial search-engines[4], the verification of the visibility of advertising logos in sports events, the detection of near-duplicate logos and unauthorized uses [2], [3].

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.

A. Paper Contribution and Organization

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 [8], [9]. It 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 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: (a) an energy function which balances a *fidelity* term; (b) a *context* criterion; (c) an *entropy* term. The fidelity term is inversely proportional to the expectation of the Euclidean distance between the most likely aligned interest points of (f_p, f_q). 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.

2. CONTEXT-DEPENDENT SIMILARITY

This process is contributing about the definition of the "Context-Dependent Similarity", function Let $S_X = \{x_1 \ldots x_n\}$, $S_Y = \{y_1, \ldots, y_m\}$ be respectively the list of interest points taken from a reference logo and a test image (the value of n, m may vary with S_X , S_Y). We borrow the definition of context and similarity design from [8], [9], in order to introduce a new matching procedure applied to logo detection. The main differences with respect to [8], [9] 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 [8]; 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 [8], context was isotropic.

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

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 $S_X \times S_Y$ match.

A. Context

The context is defined by the local spatial configuration of interest points in both S_X and S_Y. Formally, in order to take into account spatial information, an interest point $x_i \in S_X$ is defined as $x_i = (\psi_g(x_i), \psi_f(x_i), \psi_o(x_i), \psi_s(x_i), \omega(x_i))$ where the symbol $\psi_g(x_i) \in \mathrm{R}2$ stands for the $2D$ coordinates of x_i while $\psi_f(x_i) \in \mathrm{R}_c$ corresponds to the feature of x_i (in practice c is equal to 128, i.e. the coefficients of the SIFT descriptor [5]). We have also an extra information about the orientation of x_i (denoted $\psi_o(x_i) \in [-\pi, +\pi]$) which is

provided by the SIFT gradient and about the scale of the SIFT descriptor (denoted $\psi_S(x_i)$). Finally, we use $\omega(x_i)$ to identify the image from which the interest point comes from, so that two interest points with the same location, feature and orientation are considered different when they are not in the same image; this is motivated by the fact that we want to take into account the context of the interest point in the image it belongs to. Let $d(x_i, y_j) = \|\psi_f(x_i) - \psi_f(y_j)\|_2$ measure the dissimilarity between two interest point features, where $\|.\|_2$ is the "entrywise" L_2-norm (i.e. the sum of the square values of vector coefficients). The context of x_i is defined as in the following:

$$N_{\theta,\rho}(x_i) = \{x_j : \omega(x_j) = \omega(x_i), x_j \neq x_i \text{ s.t. (i), (ii) hold}\}$$

With

$$\rho - 1/N_r \epsilon_p \leq \|\psi_g(x_i) - \psi_g(x_j)\|_2 \leq \rho/N_r * \epsilon_p \quad (i)$$

and

$$\theta - 1/N_a * \pi \leq \angle(\psi_o(x_i), \psi_g(x_j) - \psi_g(x_i)) \leq \theta/N_a * \pi \quad (ii)$$

Where $(\psi_g(x_j) - \psi_g(x_i))$ is the vector between the two point coordinates $\psi_g(x_j)$ and $\psi_g(x_i)$. The radius of a neighborhood disk surrounding x_i is denoted as ϵ_p and obtained by multiplying a constant value ϵ to the scale $\psi_s(x_i)$ of the interest point x_i. In the above definition, $\theta = 1, \ldots, N_a$, $\rho = 1, \ldots, N_r$ correspond to indices of different parts of that disk (see Fig. 2).

B. Similarity Design

We define k as a function which, given two interest points $(x, y) \in S_X \times S_Y$, provides a similarity measure between them. For a finite collection of interest points, the sets S_X, S_Y are finite. Provided that we put some (arbitrary) order on S_X, S_Y, we can view function k as a matrix $\mathbf{K}$, i.e. $\mathbf{K}_{x,y} = k(x, y)$, in which the "$(x, y)$ –element" is the similarity between x and y. We also represent with $\mathbf{N}_{\theta,\rho}$, $\mathbf{N}_{\theta,\rho}$ the intrinsic adjacency matrices that respectively collect the adjacency relationships between the sets of interest points S_X and S_Y, for each context segment; these matrices are defined as $\mathbf{P}_{\theta,\rho,x,x'} = N_{\theta,\rho}(x, x')$, $\mathbf{Q}_{\theta,\rho,y,y'} = g_{\theta,\rho}(y, y')$ where g is a decreasing function of any (pseudo) distance involving (x, x'), *not necessarily symmetric*. In practice, we consider $g_{\theta,\rho}(x, x') = 1_{\{\omega(x)=\omega(x')\}} \times 1_{\{x' \in N_{\theta,\rho}(x)\}}$, so the matrices $\mathbf{P}$, $\mathbf{Q}$ become sparse and binary. Finally, let $\mathbf{D}_{x,y} = d(x, y) = \|\psi_f(x) - \psi_f(y)\|_2$. Using this notation, the similarity $\mathbf{K}$

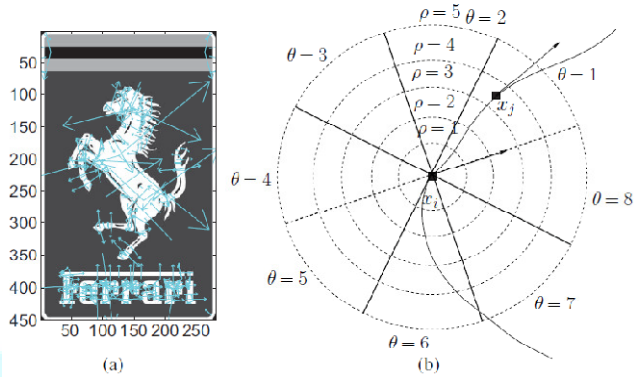

Fig. 2. (a) Collection of SIFT points with their locations, orientations, and scales. (b) Definition and partitioning of the context of an interest point x_i into different sectors (for orientations) and bands (for locations).

Between the two objects S_X, S_Y is obtained by solving the following minimization problem

$$\min_{\mathbf{K}} \text{Tr}(\mathbf{K}\,\mathbf{D}') + \beta\,\text{Tr}(\mathbf{K}\log\mathbf{K}') - \alpha \sum_{\theta,\rho} \text{Tr}(\mathbf{K}\,\mathbf{Q}_{\theta,\rho}\,\mathbf{K}'\,\mathbf{P}'_{\theta,\rho}) \quad (1)$$

Where

$$\mathbf{K} \geq \mathbf{0}, \|\mathbf{K}\|_1 = 1.$$

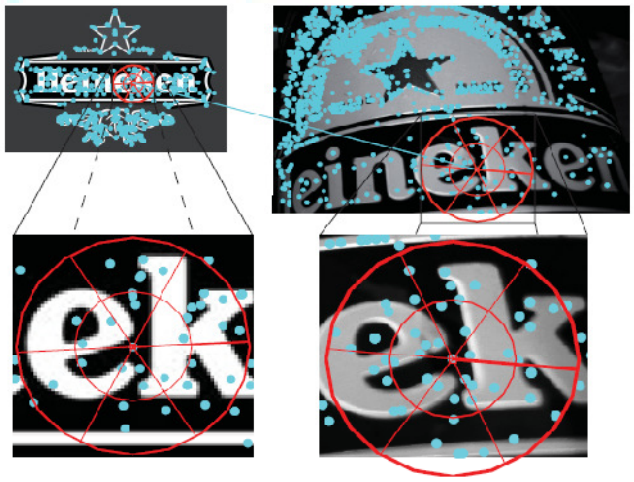

Fig.3. Example of real context definition. The two columns show the partitioning of the context of two corresponding interest points, which belong to two instances of "Heineken." In this example, we consider a context definition, including six sectors and eight bands.

Here $\alpha, \beta \geq 0$ and the operations log (natural), $\geq$ are applied individually to every entry of the matrix (for instance, log$\mathbf{K}$ is the matrix with is the entrywise" L_1-norm (i.e., the sum of the absolute values of the matrix coefficients) and Tr denotes matrix trace. The first term, in the above constrained minimization problem, measures the quality of matching between two features $\psi_f(x)$, $\psi_f(y)$. In our case

this is inversely proportional to the distance, $d(x, y)$, between the 128 SIFT coefficients of x and y. A high value of $\mathbf{D}_{x,y}$ should result into a small value of $\mathbf{K}_{x,y}$ and vice-versa. The second term is regularization criterion which considers that without any a priori knowledge about the aligned interest points, the probability distribution $\{\mathbf{K}_{x,y}: x \in S_X, y \in S_Y\}$ should be flat so the negative of the entropy is minimized. This term also helps defining a direct analytic solution of the constrained minimization problem (1). The third term is a neighborhood criterion which considers that a high value of $\mathbf{K}_{x,y}$ should imply high value in the neighborhoods $N_{\theta,\rho}(x)$ and $N_{\theta,\rho}(y)$.

C. Solution

Let's consider the adjacency matrices $\{\mathbf{P}_{\theta,\rho}\}_{\theta,\rho}$, $\{\mathbf{Q}_{\theta,\rho}\}_{\theta,\rho}$ related to a reference logo S_X and a test image S_Y respectively, each of which collects the adjacency relationships between the image interest points for a specific context segment θ, ρ. It is possible to show that the optimization problem (1) admits a unique solution $\tilde{\mathbf{K}}$, under some constrains.

Proposition 1: Let $\mathbf{u}$ denote the matrix of ones and introduce

$$\zeta = \frac{\alpha}{\beta} \sum_{\theta,\rho} \| \mathbf{P}_{\theta,\rho} \mathbf{u} \mathbf{Q}'_{\theta,\rho} + \mathbf{P}'_{\theta,\rho} \mathbf{u} \mathbf{Q}_{\theta,\rho} \|_\infty$$

Where $\|.\|_\infty$ is the "entrywise" L_∞-norm. Provided that the following two inequalities hold

$\zeta \exp(\zeta) < 1$ (2)

$\| \exp(-\mathbf{D}/\beta) \|_1 \geq 2$ (3)

the optimization problem (1) admits a unique solution $\tilde{\mathbf{K}}$, which is the limit of the *recursive form?*

$$\mathbf{K}^{(t)} = \frac{G(\mathbf{K}^{(t-1)})}{\|G(\mathbf{K}^{(t-1)})\|_1} \quad (4)$$

with

$$G(\mathbf{K}) = \exp\left\{ -\frac{\mathbf{D}}{\beta} + \frac{\alpha}{\beta} \sum_{\theta,\rho} \left(\mathbf{P}_{\theta,\rho} \mathbf{K} \mathbf{Q}'_{\theta,\rho} + \mathbf{P}'_{\theta,\rho} \mathbf{K} \mathbf{Q}_{\theta,\rho} \right) \right\} \quad (5)$$

and

$$\mathbf{K}^{(0)} = \frac{\exp(-\mathbf{D}/\beta)}{\| \exp(-\mathbf{D}/\beta) \|_1}.$$

Besides $\mathbf{K}^{(t)}$ satisfy the convergence property:

$$\|\mathbf{K}^{(t)} - \tilde{\mathbf{K}}\|_1 \leq L^t \|\mathbf{K}^{(0)} - \tilde{\mathbf{K}}\|_1 \quad (6)$$

with $L = \zeta \exp(\zeta)$.

Proof: This solution is a variant of the one found in [9]. The demonstration given in [9] still holds in this case. Notice that at the convergence stage, we omit t in all $\mathbf{K}^{(t)}$ so the latter will simply be denoted as $\mathbf{K}$.

3. LOGO DETECTION AND RECOGNITION

Application of CDS to logo detection and recognition requires establishing a matching criterion and verifying its probability of success.

Let $\mathcal{R} \subset \mathbb{R}^2 \times \mathbb{R}^{128} \times [-\pi, +\pi] \times \mathbb{R}^+$ + denote the set of interest points extracted from all the possible reference logo images (see Section II-A) and X a random variable standing for interest points in R. X and Y are assumed drawn from existing (but unknown) probability distributions. Let's consider $S_X = \{X_1, \ldots, X_n\}$, $S_Y = \{Y_1, \ldots, Y_m\}$ as n and m realizations with the same distribution as X and Y respectively. To avoid false matches we have assumed that matching between Y_J and X is assessed iff

$$\mathbf{K}_{Y_J|X} \geq \sum_{j \neq J}^{m} \mathbf{K}_{Y_j|X} \quad (7)$$

being $\mathbf{K}_{Y|X} = \mathbf{K}_{X,Y} / \left(\sum_{j=1}^{m} \mathbf{K}_{X,Y_j} \right)$.

the reference logo S_X is declared as present into the test image if, after that the match in S_Y has been found for each interest point of S_X, the number of matches is sufficiently large (at least $\tau |S_X|$ for a fixed $\tau \in [0, 1]$, being $1 - \tau$ the occlusion factor tolerated). We summarize the full procedure for logo detection and recognition in Algorithm.

A. Theoretical Foundation of Matching Algorithm

A theoretical lower bound to the probability of finding correct matches using criterion (7) can be obtained from Eq. 5, under the hypothesis of correct matches in $S_X \times S_Y$ (i.e. the reference logo exists in the image). This hypothesis is referred to as $\mathbf{H}_1$. Similarly $\mathbf{H}_0$ (the null hypothesis) stands for the incorrect matches in $S_X \times S_Y$. Assuming without a loss of generality, that all the entries of the left-hand side term of Eq. 5 (i.e., $\exp(-\mathbf{D}/\beta)$) are identical, for a fixed $\tau \in [0, 1]$, it appears clearly that the context term(the right-hand side term inside the exponential) is highly influential and that the probability of finding correct matches s dependent on setting of the parameters α/β and $q = N_a N_r$ (i.e. the fixed number of cells in the context) and also n (i.e. the number of SIFT points in the query image).

Proposition 2: Let $(.)_+$ denote the positive part of any real valued function. For a fixed $\tau \in]0, 1]$, one may show that

$$P\left(\mathbf{K}_{Y_J|X} \geq \sum_{j\neq J}^{m} \mathbf{K}_{Y_j|X}\right) \geq \left(\frac{1-\nu}{1+\nu}\right)_+ \qquad (8)$$

here $\nu = (m-1)\left(\frac{q^2-1+\exp(2\alpha/\beta)}{q^2-\tau q+\tau q\exp(2\alpha/\beta)}\right)^{qn}$ and the probability is w.r.t. $X, Y_1, \ldots, Y_m$, with $(X, Y_J) \in \mathbf{H}_1$, $(X, Y_j) \in \mathbf{H}_0$.

Proof: Fig. 4 compares theoretical expectations with measured performance, as a function of α/β, q, n and shows that with appropriate settings of these parameters; criterion (7) is able to find (almost all) the correct matches while discarding the incorrect ones.

Empirical matching are obtained on a validation set including a subset of "matches" and a subset of "non matches." The two sets were automatically generated (i) by embedding reference logos into test images at random locations so the ground-truth of "matches" and "non-matches" can be *automatically recovered* (these reference logos and test images belong to the MICC-Logos dataset), and (ii) by adding a uniformly distributed noise to the test images. Since logos can be partially occluded, it has been assumed that the reference logo is still detectable even though half-occluded in the test image, so setting $\tau = 0.5$ in the first three curves reported in the figure.

Algorithm 1: CDS Logo Detection and Recognition

Input: Reference logo image: I_X, Test image: I_Y, CDS parameters: $\epsilon, N_a, N_r, \alpha, \beta, \tau$.
Output: A boolean value determining whether the reference logo in I_X is detected in I_Y.

Extract SIFT from I_X, I_Y and let $\mathcal{S}_X := \{x_1, \ldots, x_n\}$, $\mathcal{S}_Y := \{y_1, \ldots, y_m\}$ 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, N_a, N_r;
for $j \leftarrow 1$ *to* m **do**
  Compute the context of y_j, given ϵ, N_a, N_r;
Set $t \leftarrow 1$, $max_t \leftarrow 30$;
repeat
  for $i \leftarrow 1$ *to* n **do**
    for $j \leftarrow 1$ *to* m **do**
      Compute CDS matrix entry $\mathbf{K}^{(t)}_{x_i,y_j}$, given α, β;
  Set $t \leftarrow t+1$;
until *convergence (i.e.,* $\left\|\mathbf{K}^{(t)} - \mathbf{K}^{(t-1)}\right\|_2 \rightsquigarrow 0$*) OR* $t > max_t$;
$\mathbf{K} \leftarrow \mathbf{K}^{(t)}$;
for $i \leftarrow 1$ *to* n **do**
  for $j \leftarrow 1$ *to* m **do**
    Compute $\mathbf{K}_{y_j|x_i} \leftarrow \frac{\mathbf{K}_{x_i,y_j}}{\sum_{s=1}^{m}\mathbf{K}_{x_i,y_s}}$;
    A match between x_i and y_j is declared iff $\mathbf{K}_{y_j|x_i} \geqslant \sum_{s\neq j}^{m}\mathbf{K}_{y_s|x_i}$;
if *number of matches in* $\mathcal{S}_Y > \tau|\mathcal{S}_X|$ **then**
  return *true* i.e. logo detection
else
  return *false*;

Provided that $\tau \gg 1/q$

$$\nu \overset{n\to+\infty}{\longrightarrow} 0 \quad \text{and} \quad P\left(\mathbf{K}_{Y_J|X} \geq \sum_{j\neq J}^{m} \mathbf{K}_{Y_j|X}\right) \overset{n\to+\infty}{\longrightarrow} 1.$$

B. Properties and Considerations

The adjacency matrices $\mathbf{P}_{\theta,\rho}$, $\mathbf{Q}_{\theta,\rho}$ in $\mathbf{K}$ (see Eq. 4 and 5) provide the context and the intrinsic geometry of the reference and the test logos S_X , S_Y . It is easy to see that $\mathbf{P}_{\theta,\rho}$, $\mathbf{Q}_{\theta,\rho}$ are translation and rotation invariant and can also be made scale invariant when the support (disk) of the context (i.e. its radius ϵ_p) is adapted to the scales of $\psi_S(S_X)$ and $\psi_S(S_Y)$ respectively. It follows that the right-hand side of our similarity $\mathbf{K}$ is invariant to any 2D similarity transformation. Notice, also, that the left-hand side of $\mathbf{K}$ may involve similarity invariant features $\psi_f(.)$ (Actually SIFT features), therefore $\mathbf{K}$ – and also our matching criterion (7) – is similarity invariant. The context can also be defined on other supports (rectangles, etc.) and can be made invariant to other transformations including affine and non-linear.

By taking β "not too large," one can ensure that (3) holds. Then by taking "small enough" α, inequality (2) can also be satisfied. Note that $\alpha = 0$ corresponds to a similarity which is not context-dependent (i.e. *context-free*, following our nomenclature). So, in this case, the similarities between neighbors are not taken into account to assess the similarity between two interest points.

Besides our choice of $\mathbf{K}_{(0)}$ is exactly the optimum (and fixed point) for $\alpha = 0$.

One important aspect of the method that has influence on the performance and suits to logo detection/recognition is that the local context is recursively defined. In particular, we assess that *two interest points match if their local neighbors match, and if the neighbors of their local neighbors match too, etc.* To have partitioned the neighborhood into several cells corresponding to different degrees of proximity has lead to significant improvements of our experimental results. On the one hand, the constraint (2) becomes easier to satisfy, for larger α with partitioned neighborhood, compared to [8]. On the other hand, when the context is split into different

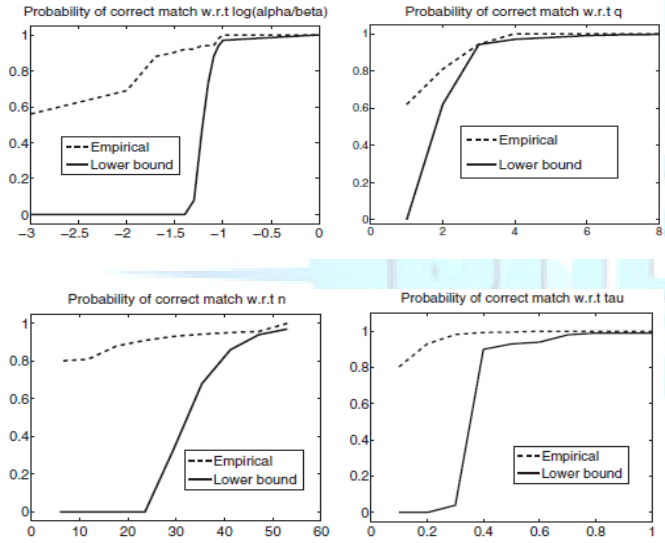

Fig. 4. Evolution of the probability of finding correct matches using criterion (7). Dashed curves correspond to the empirical measures found experimentally, while solid curves correspond to the lower bounds in (8). The evolution of these curves is shown with respect to $\log(\alpha/\beta)$, q, n, and τ, respectively. Settings used are $\alpha/\beta = 1$, $q = 4$, and $\tau = 0.5$; n and m vary with respect to reference and test images, respectively. Note that $q = 1$ corresponds to isotropic context, and $\alpha/\beta \rightarrow 0$ corresponds to context-free setting.

parts, we end up with a context term, in the right-hand side of the exponential (5), which grows slowly compared to the one presented in our previous work [8] and grows only *if similar spatial configurations* of interest points have high similarity values Fig. 5 shows an example of our context dependent matching and detection results (figures on the right) with respect to context-free ones (figures on the left).

From criterion (7) and its theoretical bound (8), several considerations follow. Under the $\mathbf{H}_1$ hypothesis, i.e. the hypothesis that the reference logo exists in the image, the lower bound in (8) increases with respect to n, q, while it decreases with respect to m. Notice that typically n, m and also that this bound is useless when $q = 1$ (i.e., when the context is isotropic) and when $q \rightarrow \infty$ (i.e., when the number of cells in the context is extremely large leading to over fitting).

We cannot know a priori what is the amount of occlusion that we may have in test images, Fig. 6 shows logo detections with different values of τ. Bound (8) shows that performance does not degrade too much when logo structure is different, i.e. some points in reference logo do not have matches in test images.

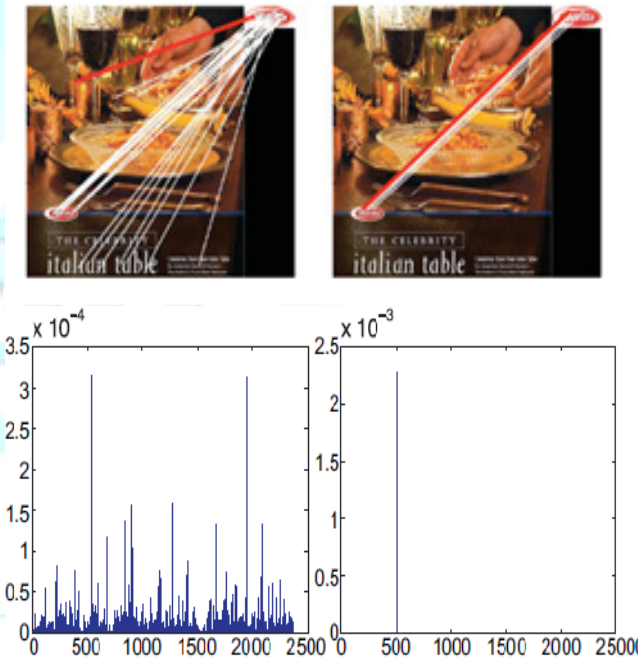

Fig. 5. Comparison of the matching results when using a context-free strategy and our context-dependent matching. The bottom figures show the conditional probability distribution $K.|X$ for a particular interest point X in the reference logo. This distribution is peaked when using context-dependent similarity so the underlying entropy is close to 0 and the uncertainty about possible matches is dramatically reduced. The top figures show the matching results between the reference logo and the test image, which are correct using the context dependent matching framework.

Context remains stable and discriminative. If we want to detect only "exact copies" of logos with only some noise and geometric (similarity) transformations, then we should set τ close to 1 (Fig. 4 also corroborates this aspect showing that the method is very selective without the need of rising the threshold too much). Under $\mathbf{H}_0$, criterion (7) is very strong and difficult to satisfy (i.e. its probability of success is $O(1/m) \rightarrow 0$) and this prevents from creating wrong matches.

4. BENCHMARKING

In order to show the effectiveness of our context dependent matching strategy (i.e., based on CDS) with respect to other approaches, we evaluate the performances of multiple-logo detection on a novel challenging dataset called MICC-Logos, containing 13 logo classes each one represented with 15–87 real world pictures downloaded from the web, resulting into a collection of 720 images (see Fig. 7).1 The image resolution varies from 480 × 360 to 1024 × 768 pixels. Interest points

A. Setting

The setting of β is related to the Gaussian similarity (i.e., exp $(-\mathbf{D}/\beta)$) as the latter corresponds to the left-hand side (and the baseline form) of $\mathbf{K}^{(t)}$, i.e. when $\alpha = 0$. Since the 128 dimensional SIFT features, used to compute $\mathbf{D}$, have a unit $L2$ norm and hence belong to a hyper sphere of radius r ($r = 1$), a reasonable setting of β is $0.1r$ which also satisfies condition (3) in our experiments.

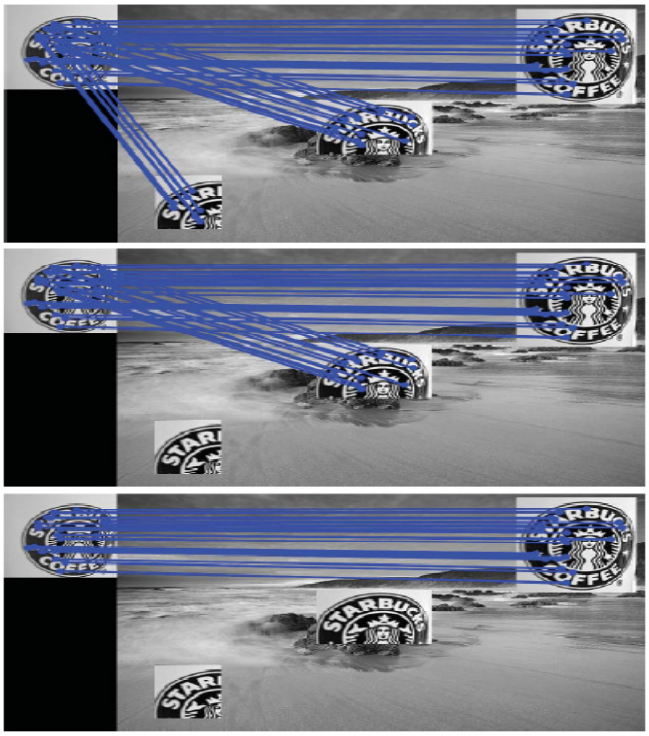

Fig. 6. Examples of logo detections with different parameters of τ (0.25, 0.5, and 0.8, respectively). As τ increases, logo detection is more sensitive to occlusion. In this experiment, $\alpha = \beta = 0.1$ and $Na = Nr = 8$.

Fig. 7. **MICC-Logos** dataset. Logo classes: the number of test images is reported for each class.

The influence (and the performance) of the right-hand side of $\mathbf{K}^{(t)}$, $\alpha \_= 0$ (context term) increases as α increases nevertheless and as shown earlier, the convergence of $\mathbf{K}^{(t)}$ to a fixed point is guaranteed only if Eq. 2 is satisfied. Intuitively, the weight parameter α should then be relatively high while also satisfying condition (2). Following the lower bounds and the empirical measures shown in Fig. 4, it is easy to see that the best matching performance is achieved when $\alpha/\beta = 1$ (in our experiments we set $\alpha = \beta = 0.1$ and $N_r = N_a = 8$) and this setting also guarantees conditions in Eqs. 2,3 and therefore the convergence of CDS to a fixed point.

B. Logo Detection Performance

Logo detection is achieved by finding for each interest point in a given reference logo Sx its best match in a test image SY ; if the number of matches is larger than $\tau \, |Sx|$, then the reference logo will be declared as present into the test image. Different values of τ were experimented and performances are measured using False Acceptance and False Rejection Rates (denoted as FAR and FRR, respectively)

$$\text{FAR} = \frac{\#\text{ of incorrect logo detection}}{\#\text{ of logo detections}}$$

$$\text{FRR} = \frac{\#\text{ of unrecognized logo appearance}}{\#\text{ of logo appearances}}.$$

.Table I reports these FAR and FRR results; setting τ to 0.5 guarantees a high detection rate at the detriment of a small increase of false alarms. Diagrams in Fig. 8 show FAR and FRR for the different classes in the MICC-Logos dataset. We clearly see the out-performance of our context dependent similarity (i.e., $\mathbf{K}^{(t)}$, $t \in \mathrm{N}+$) with respect to the baseline context-free similarity (i.e., $\mathbf{K}^{(0)}$).

C. Comparison 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.

Table I
Performance obtained using cds and different values of τ. Notice that far is a decreasing function of τ while frr is an increasing function

τ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
FAR	0.28	0.22	0.2	0.19	0.18	0.18	0.17	0.17	0.17
FRR	0.1	0.11	0.11	0.12	0.12	0.13	0.13	0.14	0.14

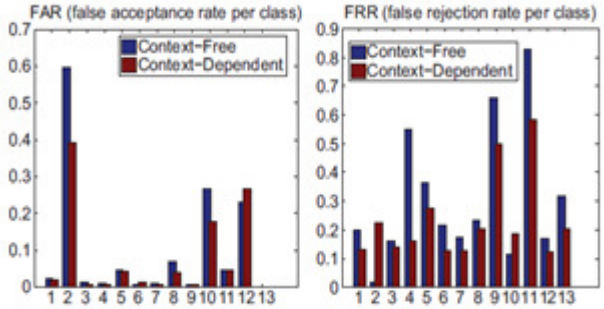

Fig. 8. Comparison of logo detection using our (i) context-dependent similarity and (ii) context-free one (actually Gaussian). FAR and FRR rates are shown for each class. In these experiments, $\beta = \alpha = 0.1$ and $\tau = 0.5$ while n and m vary, of course, with reference logos and test images. Excepting the logos "Apple" and "Mc Donald's" (which contain very few interest points $n < 12$), the FRR errors are almost always significantly reduced while FAR is globally reduced.

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 if Lowe's second nearest neighbor test is satisfied [5].

Secondly, we also compare our CDS logo detection algorithm to two relevant methods that use context in their matching procedure [6]. 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 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 .

Table II
This table shows a comparison of our cds method with respect to sift, ransac, video google and partial spatial Context (psc) matching. The first row reports far values, while each other row the corresponding frr value obtained with each method. In these experiments, cds is computed by setting $\alpha = \beta = 0.1$, $nr = na = 8$ while τ varies in order to have frr for different far

FRR \ FAR	0.299	0.181	0.125	0.094	0.075	0.06	0.051	0.043	0.037
CDS	**0.093**	**0.151**	**0.187**	**0.216**	**0.249**	**0.279**	**0.292**	**0.309**	**0.325**
SIFT	0.264	0.348	0.394	0.452	0.503	0.544	0.571	0.589	0.622
RANSAC	0.253	0.340	0.381	0.407	0.423	0.434	0.444	0.457	0.477
Video Google [25]	0.237	0.304	0.350	0.395	0.427	0.448	0.469	0.508	0.538
PSC matching [37]	0.248	0.330	0.371	0.403	0.433	0.467	0.493	0.524	0.551

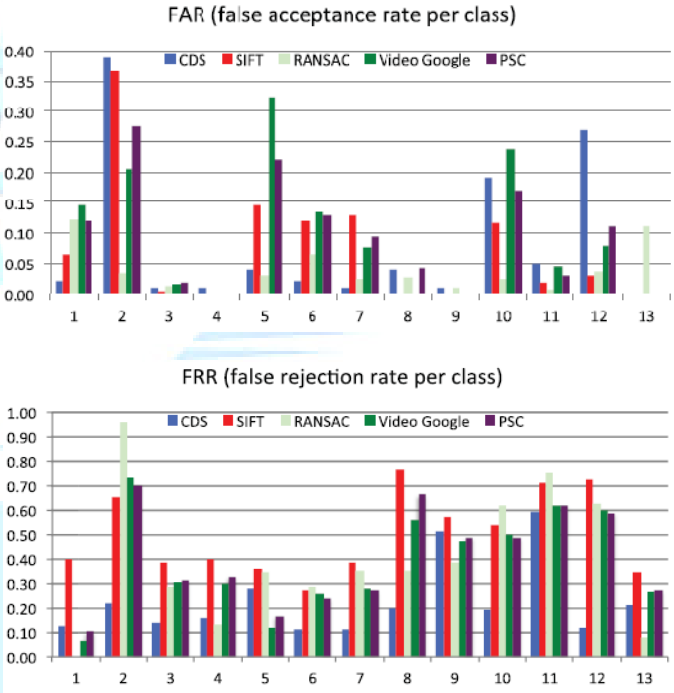

Fig. 9. Comparison of logo detection using our (i) context-dependent similarity, (ii) SIFT, (iii) RANSAC, and (iv) Video Google. FAR and FRR rates are shown for each class. In these experiments, $\alpha = \beta = 0.1$, $Nr = Na = 8$, and $\tau = 0.5$.

Fig:10. Some examples of logo detection results. (a) Examples of matching in case of partial appearance, perspective transformations, and low resolution. (b) Examples of matching in case of deformations. The default parameters used in these experiments correspond to $\alpha = \beta = 0.1$, $Nr = Na = 8$, and $\tau = 0.5$.

Partial Spatial Context (PSC) logo matching [6] 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.

Table II and Fig. 10 show a comparison of the results obtained by the five methods. Table II illustrates the FRR performance for fixed FAR values and clearly shows that our CDS method produces the lowest error rates compared to the other methods. Fig. 9 shows the FAR and FRR errors class by- class on the MICC-Logos dataset.

5. CONCLUSION

This paper contains the work of logo detection and localization on new class of similarities, which is based on context. strength of proposed method is as follows: (i) the inclusion of the spatial configuration in similarity design and visual features, (ii) the control of context and the regularization of solution, (iii) the tolerance to partial occlusions for detecting duplicate logos and some variability in their appearance and (iv) the theoretical method 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.

Further extensions of this work include the application of the method to logo retrieval in videos and also the refinement of the definition of context in order to handle other rigid and non-rigid logo transformations.

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] 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.

[3] 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.

[4] R. Phan and D. Androutsos, "Content-based retrieval of logo and Trademarks in unconstrained color image databases using color edge gradient co-occurrence histograms," *Comput. Vis. Image Understand.*, vol. 114, no. 1, pp. 66–84, 2010.

[5] D. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.

[6] 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.

[7] 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.

[8] H. Sahbi, J.-Y. Audibert, J. Rabarisoa, and R. Kerivan, "Contextdependent kernel design for object matching and recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Anchorage, AK, 2008, pp. 1–8.

[9] 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.

[10] 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.