Context Based Similarity Matching

Ishrath Jahan C[1], Abitha E[2]

[1]M.Tech (ECE) Student, MEA Engineering College,

[2]Assistant Professor, MEA Engineering College,

Abstract: ***Through this work a new method is introduced for the similarity matching of logos. Logos appear in images of real world indoor or outdoor scenes super imposed on objects of any geometry. In most of the cases these images are subjected to perspective transformations and often corrupted, which have low resolution and quality. Based on context dependent similarity kernel that directly incorporates the spatial local features, a novel solution for logo detection and recognition is attained. 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. This solution is proved to be highly effective and responds to the requirements of similarity matching of logos in real world videos.***

Keywords: similarity matching, context, logo matching, context dependent similarity, context dependent kernel.

1. Introduction

Logo or trademark recognition has been a well-studied subject for decades since it arises in many practical scenarios of modern marketing, Advertising & trademark registration. When people look to natural scenes, although they are harder to detect and image on the shirt of a soccer player for example can vary in shape although the generic object recognition and close-duplicate detection two related problems that largely has been studied in the last decades are Natural scenes logo recognition must fall under any category. On the one hand, they generally are simple geometric shapes and text and most planar surfaces appear to help detect can provide some useful prior knowledge logo. On the other hand, they have a very wide range of near duplicates are comparing and take many different forms, or variants can be. Global color or shape descriptors commonly recognized logo in clean environment that is used when It comes natural to images, because it is mainly they are extremely sensitive to background clutter, however, such descriptors have been successful. Here the problem consists of: a logo to the associated sections of the annotated brand and a query image or video frame, given the task to ascertain if brands to appear in one or more query. Context of the peculiarities of the geometry and even with acknowledgement to object localized in one scene recognition of individual things is important, and therefore difficult to address the requirements of hand needed. Relevant data image, full-grade image classification approach for spatial pyramid as apparently not justified.

2. Literature Survey

Logos are special class of graphic objects that directly connected with the identity of a particular company's product and service. Sometimes an image convey messages more than words can do. Due to this everywhere using of logos become common. As the popularity increases non authorised use is also increase. There are lot of image matching techniques are introduced for the detection and retrieval of logos.

Wei et al. [2] proposed a solution, where logos were described by global Zernike moments, local curvature and distance to centroid. This approach commences with the extraction of edges using the Canny edge detector, performs a shape normalization procedure, and then extracts the global and local features. The global features capture the gross essence of the shapes while the local features describe the interior details of the trademarks. This method 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 use of global descriptors for logo detection in real world images has been proposed by several authors [4], [5]. Phan et al. [5], considered pairs of color pixels in the edge neighbourhoods and accumulated differences between pixels at different spatial distances into a Color-Edge Co-occurence Histogram [4]. This paper present a logo and trademark retrieval system for general, unconstrained, color image databases, extending the Color Edge Co-occurrence Histogram (CECH) object detection scheme. This introduce more accurate information to the CECH, by virtue of incorporating color edge detection using vector order statistics. This produces a more accurate representation of edges in color images, in comparison to the simple color pixel difference classification of edges as seen in the CECH. This method is thus reliant on edge gradient information, thus call it the Color Edge Gradient Co-occurrence Histogram (CEGCH).

In this paper by J.Luo [4] present a robust algorithm designed to detect a class of compound colour objects given a single model image. A compound colour object is defined as having a set of multiple, particular colours arranged spatially in a particular way, including flags, logos, cartoon characters, people in uniforms, etc. This approach is based on a particular type of spatial-colour joint probability function called the colour edge co-occurrence histogram. In addition, this algorithm employs perceptual colour naming to handle colour variation, and pre-screening to limit the search scope (i.e., size and location) for the object. Experimental results demonstrated that the proposed algorithm is insensitive to object rotation, scaling, partial occlusion, and folding, outperforming a closely related algorithm based on colour co-occurrence histograms by a decisive margin.

J. Matas, O. Chum, M. Urban, and T. Pajdla,[10] explains in their paper about the wide-baseline stereo problem, i.e. the problem of establishing correspondences between a pair of images taken from different viewpoints is studied. A new set of image elements that are put into correspondence, the so called extremal regions, is introduced. Extremal regions possess highly desirable properties: the set is closed under 1.continuous (and thus projective) transformation of image coordinates and 2.monotonic transformation of image intensities. An efficient (near linear complexity) and practically fast detection algorithm (near frame rate) is presented for an affinely-invariant stable subset of extremal regions, the maximally stable extremal regions (MSER). A new robust similarity measure for establishing tentative correspondences is proposed. The robustness ensures that invariants from multiple measurement regions (regions obtained by invariant constructions from extremal regions), some that are significantly larger (and hence discriminative) than the MSERs, may be used to establish tentative correspondences.

D.Lowe [9], presents a method for extracting distinctive invariant features from images that can be used to perform reliable matching between different views of an object or scene. The features are invariant to image scale and rotation, and are shown to provide robust matching across a substantial range of affine distortion, change in 3D viewpoint, addition of noise, and change in illumination. The features are highly distinctive, in the sense that a single feature can be correctly matched with high probability against a large database of features from many images. This paper also describes an approach to using these features for object recognition. The recognition proceeds by matching individual features to a database of features from known objects using a fast nearest-neighbour algorithm, followed by a Hough transform to identify clusters belonging to a single object, and finally performing verification through least-squares solution for consistent pose parameters. This approach to recognition can robustly identify objects among clutter and occlusion while achieving near real-time performance.

G. Carneiro And A. Jepson [6], in their paper describes, a key step for the effective use of local image features (i.e., highly distinctive and robust features) for recognition or image matching is the appropriate grouping of feature matches. Spatial constraints are important in this grouping because, during a recognition process, they allow for the reduction of the number of hypotheses that must be verified and also reduce the number of false positives present in each of these hypotheses. A common choice for this grouping task is to use the Hough transform on the global spatial transformation parameters of the hypothesized matches. Here, instead, use semi-local spatial constraints which allow for a greater range of shape deformations

M. Merler, C. Galleguillos, And S. Belongie[3], in their paper Recognizing Groceries In Situ Using In Vitro Training Data explains the problem of using pictures of objects captured under ideal imaging conditions (here referred to as in vitro) to recognize objects in natural environments (in situ) is an emerging area of interest in computer vision and pattern recognition. A. D. Bagdanov[11] describe a system for detection and retrieval of trademarks appearing in sports videos. In the paper named Efficient Mining of Frequent and Distinctive Feature Configurations, Explains a novel approach to automatically find spatial configurations of local features occurring frequently on instances of a given object class, and rarely on the background is present. The approach is based on computationally efficient data mining techniques and can find frequent configurations among tens of thousands of candidates within seconds.

H. Sahbi[8] in his paper Context Dependent Kernel Design for Object Matching and Recognition explains the success of kernel methods including support vector networks (SVMs) strongly depends on the design of appropriate kernels. While initially kernels were designed in order to handle fixed-length data, their extension to unordered, variable-length data became more than necessary for real pattern recognition problems such as object recognition and bioinformatics. This paper focus on object recognition using a new type of kernel referred to as “context-dependent”. Objects, seen as constellations of local features (interest points, regions, etc.), are matched by minimizing an energy function mixing (1) a fidelity term which measures the quality of feature matching, (2) a neighborhood criteria which captures the object geometry and (3) a regularization term. The fixed-point of this energy is a “context dependent” kernel (“CDK”) which also satisfies the Mercer condition.

3. Context Dependent Similarity Matching

A. Context Dependent Similarity

Let $S_X = \{x_1, \dots \dots \dots x_n\}$, $S_Y = \{y_1, \dots \dots \dots y_m\}$ 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).The definition of context and similarity design in order to introduce a new matching procedure applied to logo detection.

The use of context for matching

Context is used to find interest point correspondences between two images in order to tackle logo detection. 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.

The similarity diffusion process

Resulting from the definition of context, similarity between interest points is recursively and anisotropically diffused.

The interpretation of the model

This designed similarity may be interpreted as a joint distribution (pdf) which models the probability that two interest points taken from S_X , S_Y 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 S_X , S_Y.

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 $x_i = (\varphi_g(x_i), \varphi_f(x_i), \varphi_o(x_i), \varphi_s(x_i), w(x_i))$, where the symbol $(\varphi_g(x_i) \in R^2$ stands for the 2D coordinates of x_i while $\varphi_f(x_i) \in R^c$ corresponds to the feature of x_i (in practice c is equal to 128, i.e. the coefficients of the SIFT descriptor). We have also an extra information about the orientation of x_i (denoted , $\varphi_o(x_i) \in [-\pi, +\pi]$ which is provided by the SIFT gradient and about the scale of the SIFT descriptor (denoted $\varphi_s(x_i)$). Finally, we use $w(x_i)$ 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 $d(x_i, y_i) = \| \varphi_f(x_i) - \varphi_f(y_j) \|_2$ 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:

$$N^{\theta,\rho}(x_i) = \{x_j : w(x_j) = w(x_i), x_j \neq x_i\} \quad (1)$$

with

$$\frac{\rho-1}{N_r}\varepsilon_p \leq \|\varphi_g(x_i) - \varphi_g(x_j)\|_2 \leq \frac{\rho}{N_r}\varepsilon_p \quad (2)$$

and

$$\frac{\theta-1}{N_a}\pi \leq \angle(\varphi_o(x_i), \varphi_g(x_j) - \varphi_g(x_j)) \leq \frac{\theta}{N_a}\pi \quad (3)$$

Where $(\varphi_g(x_j) - \varphi_g(x_i))$is the vector between the two point coordinates $\varphi_g(x_j)$ and $\varphi_g(x_i)$. The radius of a neighborhood disk surrounding x_i is denoted as p and obtained by multiplying a constant value to the scale $s(x_i)$ of the interest point x_i . In the above definition,$\theta = 1 \ldots\ldots, N_a, \rho = 1 \ldots\ldots\ldots, N_r$ correspond to indices of different parts of that disk. In practice Na and Nr correspond to 8 sectors and 8 bands. The definition of neighborhoods $\{N^{\theta,\rho}(x_i)\}_{\theta,\rho}$ reflects the co-occurrence of different interest points with particular spatial geometric constraints. Fig. 2 shows an example taken from two different images containing the same logo; the figure reports the context definition for two corresponding keypoints, showing a similar spatial configuration. All the definitions about interest points in S_Y and their context are similar toS_X.

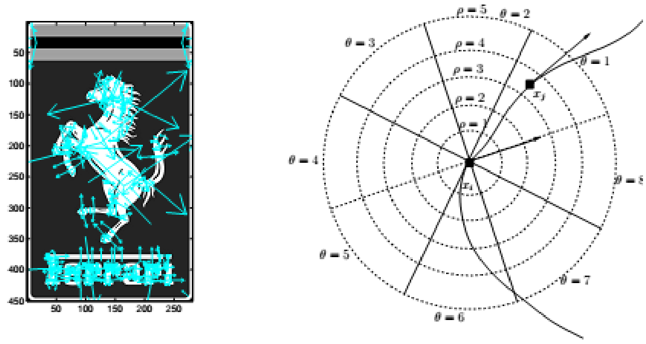

Figure 1: Collection of SIFT points

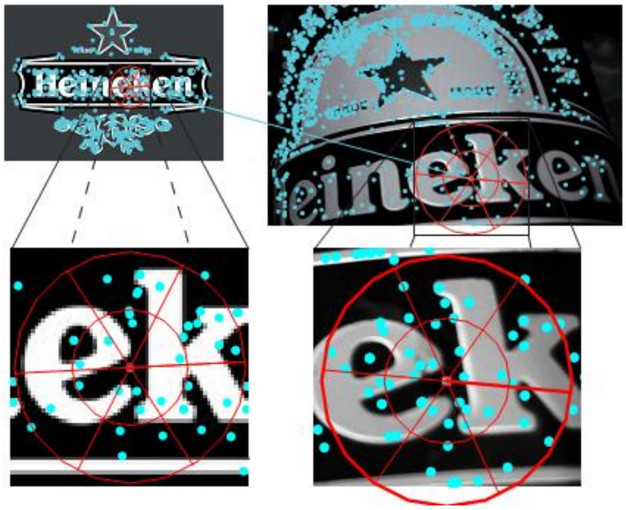

Figure 2: Example of real context

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 K, i.e. $K_{x,y} = k(x, y)$, in which the "(x, y) - element" is the similarity between x and y. We also represent with $P_{\theta,\rho}$, $Q_{\theta,\rho}$ 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 $P_{\theta,\rho,x,x'} = g_{\theta,\rho}(x, x')$, $Q_{\theta,\rho,y,y'} = g_{\theta,\rho}(y, y')$ where g is a decreasing function of any (pseudo) distance involving (x, x'), not necessarily symmetric. In practice, we consider $g_{\theta,\rho}(y, y') = 1_{\{w(x)=w(x')\}} \times 1_{\{x' \in N^{\theta,\rho}(x)\}}$, so the matrices P, Q become sparse and binary. Finally, let $D_{x,y} = d(x, y) = \| \varphi_f(x) - \varphi_f(y) \|_2$. Using this notation, the similarity K between the two objects S_X,S_Y is obtained by solving the following minimization problem

$$\min_K Tr(K\,D') + \beta\, Tr(K \log K') - \alpha \textstyle\sum_{\theta,\rho} Tr(K\, Q_{\theta,\rho}\, K'\, P'_{\theta,\rho}) \quad (4)$$

$$\text{s.t} \quad \begin{cases} K \geq 0 \\ \|K\|_1 = 1 \end{cases} \quad (5)$$

Here α,β ≥ 0 and the operations log (natural), ≥ are applied individually to every entry of the matrix (for instance, log K is the matrix with $(log\ K)_{x,y} = \log k(x,y)$, $\|\ .\|_1$ is the "entrywise" L1-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 f (x), 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 $D_{x,y}$ should result into a small value of $K_{x,y}$ and vice-versa. The second term is a regularization criterion which considers that without any a priori knowledge about the aligned interest points, the probability distribution $\{K_{x,y} : x \in S_x, : y \in S_y\}$ 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. The third term is a neighborhood criterion which considers that a high value of $K_{x,y}$ should imply high values in the neighborhoods $N^{\theta,\rho}(x)$, $N^{\theta,\rho}(y)$ .This criterion also makes it possible to consider the spatial configuration of the neighborhood of each interest point in the matching process. This minimization problem is formulated by adding an equality constraint and bounds which ensure a normalization of the similarity values and allow to see K as a probability distribution.

4. Logo Detection

There are two inputs one is the reference logo which is the one to detect from the real world video. The second input is a real world test video which contains lot of images logos etc. The test input is a real world video that means taken without any isolation. A real world video or image consists of many limitations like background cluttering, change in illumination, less quality, low resolution, lighting effects, partial occlusion etc. due to this the logo detection become a really challenging task.

The two inputs are accepted by the program. The next step is to convert the video into frames. The real world video is converted to frames as per the given number of frames. This is for the ease of calculation. This converted frames as well as the reference image are taken to the next step called feature extraction and context dependent algorithm. With the help of this we get the probability of matching. In feature extraction the local features of both images are extracted. In the case of test video each frame is considered as one image and the features of each image is taken. With the help of feature extraction the keypoints and descriptors are extracted from each figure.
The extracted features are then used to calculate the context and the adjacency matrix is created with the help of context dependent similarity algorithm. According to this a probability of keypoint matching is calculated. If this value exceeds the threshold value then the logo is detected in that particular frame. This is repeated for each frame and calculated the number of frames containing the logo images.

5. Results

As we had described two inputs are there, one is an image and the second one is real world video.

Figure 3: input logo image

The above figure represents the input logo image. This image is matched with the each and every frame of the real world video that given as the second input.

Figure 4: matching with the frames that does not contain logo

Figure 5: matching with the frames that contains logo

The real world video frames may or may not contain logo image. There are frames that contain logo and does not contain logo. Each and every frame matched with the input and some of example output for both the frames contain and does not contain logo are shown above.

6. Conclusion

This context based similarity detection introduces a novel logo detection and localization approach based on a new class of similarities referred to as context dependent similarity. The companies use logos for conveying their messages about the product or service. Searching a particular logo from a group of images present in the real world video is extracted using a context dependent algorithm. The strength of this method resides in several aspects: the inclusion of the information about the spatial configuration in similarity design as well as visual features, the ability to control the influence of the context and the regularization of the solution, 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.

References

[1] Hichem Sahbi, Lamberto Ballan, Giuseppe Serra, and Alberto Del Bimbo, ‘‘Context-Dependent Logo Matching and Recognition’’, 2013, pp. 1-13.
[2] 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 Recognition, vol. 42, no. 3, pp. 386–394, 2009.
[3] M. Merler, C. Galleguillos, and S. Belongie, “Recognizing groceries in situ using in vitro training data,” in Proc. of IEEE CVPR SLAM Workshop, Minneapolis, MN, USA, 2007, pp. 1–8.
[4] J. Luo and D. Crandall, “Color object detection using spatial-color joint probability functions,” IEEE Transactions on Image Processing, vol. 15, no. 6, pp. 1443–1453, 2006
[5] R. Phan, J. Chia, and D. Androutsos, “Unconstrained logo and trademark retrieval in general color image database using color edge gradient cooccurence histograms,” in Proc. of IEEE ICASSP, Las Vegas, NV, USA 2008, pp. 1221–1224.
[6] G. Carneiro and A. Jepson, “Flexible spatial models for grouping local image features,” in Proc. of IEEE CVPR, vol. 2, Washington, DC, USA, 2004, pp. 747–754.
[7] J. Kleban, X. Xie, and W.-Y. Ma, “Spatial pyramid mining for logo detection in natural scenes,” in Proc. of IEEE ICME, Hannover, Germany, 2008, pp. 1077–1080.
[8] H. Sahbi, J.-Y. Audibert, J. Rabarisoa, and R. Kerivan, “Context dependent kernel design for object matching and recognition,” in Proc. of IEEE CVPR, Anchorage, AK, USA, 2008, pp. 1–8.
[9] D. Lowe, “Distinctive image features from scale-invariant keypoints,” International Journal of Computer Vision, vol. 60, no. 2, pp. 91–110, 2004
[10] J. Matas, O. Chum, M. Urban, and T. Pajdla, “Robust wide-baseline stereo from maximally stable extremal regions,” Image and Vision Computing, vol. 22, no. 10, pp. 761–767, 2004.
[11] A. D. Bagdanov, L. Ballan, M. Bertini, and A. Del Bimbo, “Trademark matching and retrieval in sports video databases,” in Proc. of ACM MIR, Augsburg, Germany, 2007, pp. 79–86.