Logo Retrieval with Latent Semantic Analysis

Jinqiao Wang Qingshan Liu Jing Liu Hanqing Lu
National Lab of Pattern Recognition
Institute of Automation, Chinese Academy of Sciences, Beijing 100080, China
{jqwang, qsliu, jliu, luhq}@nlpr.ia.ac.cn

Abstract

The frequency and duration of the logo in broadcast are related to the measurement of advertisement effect. In this paper, we propose a novel logo retrieval approach based on latent semantic analysis. The 128 dimensional SIFT logo descriptors are extracted in the logo images to model the invariability against affine transformation, perspective transformations, occlusions and motion blur. The logo descriptors are clustered by k-means method to build the visual logo dictionary, and latent semantic indexing is used for logo retrieval. In addition, temporal continuity and dependency are considered to further improve the performance of logo retrieval. The experiments on the Worldcup 2006 sports video demonstrate the promising results of the proposed method.

Keywords: Logo Retrieval, SIFT descriptor, Latent semantic analysis, Clustering algorithm.

1 Introduction

With the development of TV broadcast, more and more companies spend large amounts of money on propagandizing their products. The duration and frequency of the logo appearing in broadcast are important factors to assess the effect of advertisements, which is directly related to adjustment of the commercial investment. The size, position and deformation of the logo are important to attract customer's attention and evaluate the performance of advertisement. Automatic detection and statistics of the logo can reduce the labor-intensive monitoring work of broadcast TV. Generally speaking, logos are composed of text, graphics and storytelling images. The text has the name of the company and (or) product information of the company. The graphics and images create a more or less abstract, symbolic, or vivid indicator of the product or company.

Previous work mainly focused on logo and brand detection in text documents (Jesus & Facon 2000, Seiden, Dillencourt, Irani, Borrey & Murphy 1997) and logo indexing in large logo databases (Eakins, Boardman & Graham 1998, Soffer & Samet 1998), which just need to consider the logos with front view and in clear backgrounds. Soffer *et al* (Soffer & Samet 1998) represented and matched logo based on positive and negative shape features, and Seiden *et al* (Seiden et al. 1997) extracted a set of grayscale features to construct a suite of rules for classifying the segmentation of the logos. Different with the logos in the database, the brand logos in broadcast videos are almost viewed with perspective and affine transformation, and in complex background. Sometimes the logos are occluded by moving objects and motion blur appears. In (Kovar & Hanjalic 2000) a logo detection and classification method was proposed by using edges and shape moments, but it can only deal with the video stills which contain less information than full video. In (den Hollander & Hanjalic 2003), string matching is used for logo recognition in video. It only considers some simple cases, such as highly contrasted background and non-occluded logo appearance.

In this paper, we focus on the logos in broadcast videos, and propose to use latent semantic analysis (LSA) for the logos retrieval. The logo is modelled by the SIFT logo descriptors which are robust to viewpoint and illumination variations. The descriptors of all the logos are clustered into logo words by k-mean algorithm, and then the logo words are used to build a latent semantic space. Given the queries, latent semantic indexing is applied to search for the logos in the video streams, then the temporal continuity and dependency are considered to further improve the performance of logo retrieval.

The rest of this paper is organized as below. Section 2 briefly introduces logo descriptors using SIFT features. Section 3 presents the logo retrieval algorithm by latent semantic analysis. Section 5 gives the experimental results of logo retrieval and analyzes the effect of cluster number on the retrieval performance. Finally conclusion is drawn in Section 6 and acknowledgement is given in Section 7.

2 Logo descriptor

Due to the complex background, dynamic content and camera motion in broadcast videos, too many variations exist, such as size, rotation, occlusion, illumination, and motion blur, which will influence the performance of logo retrieval. So we should extract the local region descriptors that are robust to these variations. Moreover, a good set of logo descriptors should have sufficient discriminability for retrieval in a large logo database.

The SIFT features developed by Lowe (Lowe 1999) are demonstrated to be superior to steerable filters and orthogonal filters for local feature description. A SIFT descriptor of local region is based on the gradient magnitudes and orientations of its pixels. The region is first split into r * r subregions. As for each subregion, an orientation histogram is then formed by accumulating samples within the subregion, weighted by gradient magnitudes. Concatenating the histograms of all the

subregions forms a SIFT vector. Fig. 1 shows an example of logo descriptors in a frame of sports video, in which the logo template "FUJIFILM" has 78 descriptors and the frame has 1745 descriptors. The number of the matched descriptors is 38.

Figure 1: An example of logo descriptors with SIFT features. The image in the left top corner is the logo template, and the bottom is a frame in a video shot. The matched logo descriptors are connected with blue lines.

The SIFT descriptors are normalized to unit length so as to reduce the effects of illumination changes, so the brightness change in a local region will not affect the gradient values for the normalization. Also, the sensitivity of the SIFT descriptor to affine change was examined in (Lowe 2004), which provided better performance for extreme affine changes.

In this paper, χ^2 distance (Ma & Grimson 2005) is used for the distance measure between the SIFT descriptors instead of Euclidean distance (Lowe 1999). Euclidean distance only cares about absolute differences in histogram bins. If the absolute differences of corresponding bins are small, their Euclidean distance is small, no matter how large the differences are relative to the values in the bins. χ^2 distance considers bin differences relative to bin values to give a better comparison between two histogram distributions.

3 Latent semantic analysis for logo retrieval

Latent semantic analysis (LSA) (Deerwester, Dumais, Furnas, Landauer & Harshman 1990) is widely applied in natural language processing, especially in vectorial semantics to overcome the problems of synonymy and polysemy. Recently, it also achieves a great success in source code analysis and object retrieval. LSA uses a term-document matrix to describe the occurrences of terms in documents, and assumes that some underlying or latent structure exists in word usage, which is partially obscured by variability in word choice. A truncated singular value decomposition (SVD) is used to estimate the structure in word usage across documents. For a queried logo, the visual logo dictionary is built with its logo descriptors that represent the characteristic of logo regions. Cosine similarity is computed to search the matched logos in video streams.

3.1 Building logo word dictionary Heading Level 3

The logo word dictionary is constructed by mapping the similar logo descriptors into clusters. Each cluster represents a logo word in the latent semantic space, which is represented by its cluster centroid. We use the k-means clustering method and χ^2 distance as the similarity measure to build the logo word dictionary.

3.2 Logo word weighting

We take the video frames as a logo document, and use the Okapi BM25 relevance scoring formula (Hawking, Upstill & Toward 2004) generate text vectors.

$$weight_t = tf_s \times \frac{\log\left(\frac{N-n+0.5}{n+0.5}\right)}{k_1 \times ((1-b) + b \times \frac{dl}{avdl}) + tf_s} \qquad (1)$$

where $weight_t$ is the relevance weight assigned to a visual logo document due to query logo term t. tf_s is the term frequency in the visual logo document, $k_1 = 2.0$, $b = 0.75$. N is the total number of logo documents, n is the number of logo documents containing at least one occurrence of logo term t, dl is the length of the logo documents and $avdl$ is the average logo document length. The document-by-terms matrix $A_{t\times s}$ for a whole logo database is created by *words* × *documents*.

3.3 Latent semantic analysis

Considering the video frames as a text document and regarding each cluster of logo descriptors as a logo word, LSA describes the semantic content of the logo by mapping logo words onto a semantic space. Singular value decomposition (SVD) is employed to create the semantic space. The SVD projection is obtained by decomposing the matrix $A_{t\times s}$ of size t words and s contexts into the product of three separate matrices.

$$A_{t\times s} = T_{t\times n} S_{n\times n} (D_{s\times n})^T \qquad (2)$$

Where t is the number of terms, s is the number of logo documents, $n = \min(t,s)$, T and D have orthonormal columns, i.e. $T^TT = D^TD = I$. S_n is a diagonal matrix of size $L = min(t,s)$ with singular values λ_1 to λ_2, where

$$\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_L \qquad S \approx diag(\lambda_1, \lambda_2, \ldots, \lambda_L) \qquad (3)$$

Only the first k ($k < n$) eigenvalues are used create the pseudo-document matrix.

$$\hat{A}_{t\times k} = T_{t\times k} S_{k\times k} (D_{s\times k})^T \qquad (4)$$

Each column in the pseudo-document matrix is the logo descriptors of each logo document. To measure the result of logo retrieval in broadcast video, cosine distance is employed to measure the frame-level logo content similarity. The query logo descriptor q contains the logo words that describe the logo characteristic. The word-to-document logo similarity is

$$\begin{aligned} q^T_{t\times 1} \hat{A}_{t\times k} &= q^T_{t\times 1} T_{t\times k} S_{k\times k} D^T_{s\times k} \\ &= (q^T_{t\times 1} T_{t\times k})(S_{k\times k} D^T_{s\times k}) \end{aligned} \qquad (5)$$

Let $p_q = q_{t\times1}{}^T T_{t\times k}$ and p_j be the jth context of $(S_{k\times k} D_{s\times k}{}^T)$.

$$similarity(p_j, q) = \frac{p_q \cdot p_j}{\|p_q\| \cdot \|p_j\|} \quad (6)$$

The number of singular values k drives the LSA performance. If too many factors are kept, the noise will remain and the detection of synonyms and the polysemy of visual words will fail. Alternatively, if too few factors are kept, important discriminating information will be lost. In our experiment, we empirically keep the dimension of the resulting semantic space k is 60.

4 Temporal continuity

In order to handle the problems of occlusion and motion blur in some broadcast video, temporal continuity and dependency are analyzed to improve the performance of logo retrieval. For the logos appear close together in the retrieved video frames, we first retrieved the frames using the latent semantic analysis, and then refine the results by the temporal continuity and dependency. If the logo is not detected in only a few frames in the middle of the frame sequence, we consider that all the frames in the sequence including the logo. Fig. 2 shows that the logo "HYUNDAI" is occluded by the players in several frames, while the whole shot includes the logo. Fig. 3 shows an occluded logo "MASTERCARD" in a close-up shot, with the frames missed detected is included by the temporal continuity. If the logo is detected in only a few frames in the frame sequence, we treat the frame sequence without the logo.

Figure 2: Logo retrieval with logo occluded by players. The number of matched logo words is small in the occluded frames while other frames are lots of logo words. If the number of occluded frames is small, the frames can be included by temporal continuity analysis.

5 Experiment

To evaluate the performance of the proposed algorithm, we conducted the experiments on Worldcup 2006 sports video data. Four matches FRANCE VS KOREA, BRAZIL VS AUSTRALIA, GERMANY VS ARGENTINA, and FRANCE VS BRAZIL are evaluated. The video is in MPEG-1 format with the frame rate of 29.97 fps and the frame size of 640 × 480. There exit camera motion, occlusion, motion blur and different lightning conditions in the video streams. The ground truth was manually labelled.

For a query logo, the precision is the ratio of the number of relevant frames to the total number of retrieved frames, and recall is the rate of the number of correctly retrieved frames relative to the number of relative frames. Here, we use F1 to evaluate our algorithm for it considers the precision and recall together. F1 is defined as follow:

$$F1 = \frac{2 \times Recall \times Precision}{Recall + Precision} \quad (7)$$

Figure 3: Logo retrieval with partial occlusion and camera motion. Although the logos "MASTERCARD" in the first and second frame have different scales and the second logo is occluded by a people, the frame is included through analyzing the temporal continuity.

Figure 4: Logo retrieval with camera motion. There are 21 logo words in the left frame and 76 logo words in the right frame. The matched logo words between the two frames are 20.

Figure 5: Logo retrieval with perspective and affine transformation. Although the logos "MASTERCARD" in the first and second frame have different scales and position, the matched logo words between the two frames are 15.

Logo retrieval with complex situation such as Camera motion, occlusion, motion blur and different lightning conditions are investigated in our experiments. Fig. 4 gives an example of logo retrieval in a shot with the camera zooming in. The logo "FUJIFILM" is very small in the left frame and the number of detected logo words is 21. Although there are 76 logo words in the right frame that is the end frame of this shot, 21 logo words are enough for the detection of logos.

Fig. 5 shows an example of logo retrieval in different position of the football field, in which the match logo

words are invariant to perspective and affine transformation. The logos "MASTERCARD" in the first and second frame have different scales and about 90 degree change in viewpoint. The number of logo words detected in the first frames is 96, and the number of logo words detected in the second frames is 17. The match logo words between the two frames are 15.

5.1 Comparison with different number of logo words

The number of clusters is important to build the latent semantic space and index logos. Several logo word dictionaries are built to analysis the impact of clusters. Fig. 6 shows the average performance of logo retrieval with 100, 200, and 500 clusters, which reveals the importance of the dictionary size on the performances of logo retrieval in broadcast video. Too small clusters of logo words remove the difference of individual logos, while too many clusters of logo words hide the similarity of individual logos and add the complexity of calculation. As illustrated in Fig. 6, 200 clusters can get a more satisfied result than 100 and 500.

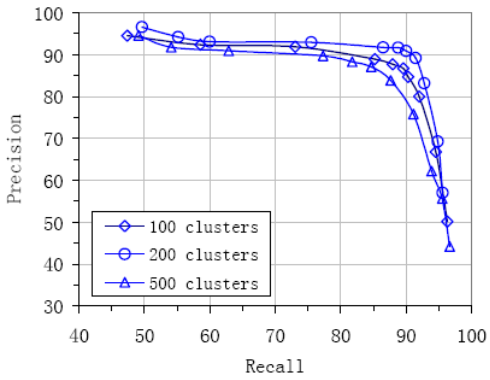

Figure 6: Relationship between the number of cluster and the logo retrieval performance.

Figure 7: Five Logos used in our experiment including "FUJIFILM", "MASTERCARD", "ADIDAS", "HYUNDAI" and "TOSHIBA" . These logos appear in different position in the football field.

5.2 Retrieval performance

With the proposed algorithm, the retrieval results are slightly different for different logo. In our experiment, both simple and complex logos are considered. As shown in Fig. 7, the logos of "FUJIFILM", "MASTERCARD", "ADIDAS", "HYUNDAI" and "TOSHIBA" are used to logo retrieval in Worldcup 2006 video data. The average F1 are 87.76% with 200 logo word clusters as shown in Fig. 8. The performance of logo "HYUNDAI" is low for the size is very small in some frames, and often occluded by the players. Both "FUJIFILM" and "MASTERCARD" are placed on the outside of the touch-line near the mid-field, which have a longer duration and frequency appearing in the broadcast video than other logos.

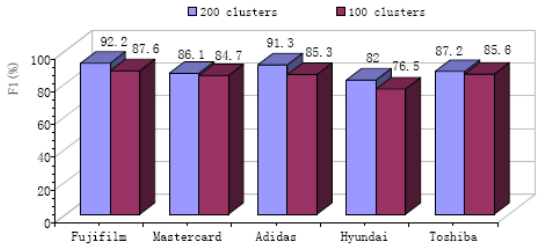

Figure 8: Logo retrieval performance with LSA based algorithm on five logos with different number of clusters.

6 Conclusion

In this paper, a logo retrieval method is proposed by latent semantic analysis. SIFT descriptors are utilized to represent the local characteristic of the logo. The logo descriptors are used to build the visual logo dictionary and the latent semantic indexing is used for logo retrieval. Temporal continuity and dependency is analyzed to improve the performance of logo retrieval. Promising results are obtained in our experiments.

7 Acknowledgement

This work is supported by National Natural Science Foundation of China (Grant No. 60475010 and 60121302).

References

Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K. & Harshman, R. (1990), 'Indexing by latent semantic analysis', *Journal of the Society for Information Science* 41(6), 391-407.

Den Hollander, R. & Hanjalic, A. (2003), 'Logo detection in video by string matching', *Proc. ICIP'03* pp. 517-520.

Eakins, J. P., Boardman, J. M. & Graham, M. E. (1998), 'Similarity retrieval of trademark images', *ACM Multimedia'98* pp. 53-63.

Hawking, D., Upstill, T. & Toward, N. C. (2004), Toward better weighting of anchors, *in* 'Proc. ACM SIGIR'04', pp. 25-29.

Jesus, A. D. & Facon, J. (2000), 'Segmentation of brazilian bank check logos without a priori knowledge', *Proc. International Conference on Information Technology'00* pp. 259-263.

Kovar, B. & Hanjalic, A. (2000), 'Storage and retrieval for media databases', *Logo detection and classification in a sport video: Video indexing for sponsorship revenue control* .

Lowe, D. (1999), 'Object recognition from local scale invariant features', *Proc. ICCV'99* .

Lowe, D. G. (2004), 'Distinctive image features from scale-invariant keypoints', *International Journal of Computer Vision* .

Ma, X. & Grimson, W. E. L. (2005), Edge-based rich representation for vehicle classification, *in* 'Proc. ICCV'05'.

Seiden, S., Dillencourt, M., Irani, S., Borrey, R. & Murphy, T. (1997), Logo detection in document images, *in* 'International Conference on Imaging Science, Systems, and Technology, CISST'97', pp. 446-449.

Soffer, A. & Samet, H. (1998), Using negative shape features for logo similarity matching, *in* 'Proc. ICPR'98', pp. 571-573.