AUTOMATIC TV LOGO DETECTION AND CLASSIFICATION IN BROADCAST VIDEOS

Nedret ÖZAY, Bülent SANKUR

Bogazici University
Bebek, 34342, Istanbul, Turkey
phone: + (90) 212 359 6414, fax: + (90) 212 287 24 65 , email: {nedret.ozay, bulent.sankur}@boun.edu.tr
web: http://www.busim.ee.boun.edu.tr

ABSTRACT

In this study[1], we present a fully automatic TV logo identification system. TV logos are detected in static regions given by time-averaged edges subjected to post-processing operations. Once the region of interest of a logo candidate is established, TV logos are recognized via their subspace features. Comparative analysis of features has indicated that ICA-II architecture yields the most discriminative with an accuracy of 99.2% in a database of 3040 logo images (152 varieties). Online tests for both detection and recognition on running videos have achieved 96.0% average accuracy. A more reliable logo identifier will be feasible by improving the accuracy of the extracted logo mask.

1. INTRODUCTION

Television (TV) broadcast channels are recognized by their logo; in turn, logos provide the clue for automatic identification of the channel. Several applications of automatic recognition of TV logos can be envisioned:

- TV Commercial Detection: TV logos partly or fully disappear during commercials. This can be exploited to segment broadcast videos into commercial or non-commercial (normal program) parts. For instance, Personal Video Recorder (PVR) devices can be programmed to store only the normal program.
- TV Logo Removal: Whenever a video content owned by a TV channel is rebroadcast by another TV channel, then two different logos appear on the screen. To eliminate this viewing displeasure the rebroadcaster may opt to remove the source logo before rebroadcast.
- Audience Measurement (Rating Measurement): The rating measurements have become ever more important for commercial channels in planning their broadcasting schedule and content. TV logo identification is instrumental for non-proprietary audience measurement.
- Archival Search: For navigation in large video archives TV logos form an obvious tool.
- Broadcast monitoring: There are some legal regulations for broadcasting (e.g. duration of commercials) that TV broadcast channels must conform to. An automatic broadcast monitoring system can make use of TV logos.

The apparent simplicity of the problem is, in fact, a bit deceiving, as the automatic TV logo identification problem must overcome several subtle problems. The logo identification problem consists of the stages: logo detection and logo classification stages. Logo detection entails sensing the presence of the logo and localizing its mask or its region of interest; logo classification implies recognizing the candidate object under the mask as one of logos in the database. The novelty of our work consists of:

- Exploration of several subspace techniques for feature extraction.
- Performance analysis in terms of logo detection time.
- An integrated detection and classification scheme and its analysis.
- Classification experiments on a large logo database (152 TV Channels).

In Section 2, we summarize the related work in the literature. We describe our logo detection scheme in Section 3, and the features for classification are presented in Section 4. TV logo identification system is explained in Section 5. Experimental results are given in Section 6, and conclusions drawn in Section 7.

2. RELATED WORK

Logo detection algorithms are almost always based on the static character of the logo region vis-a-vis regions where typical program material appears. For example, Albiol et al. [1] use time-averaged gradients and the subsequent thresholded stable contours that emerge. The stable contours are morphologically operated and converted to logo masks, provided they persisted for a given duration. A similar algorithm is introduced by Wang et al. [8], who use color gradient information.

An alternative logo detection approach [2, 3, 5] in the literature exploits pixel-wise differences. For example, in Yan et al. [2], the two-step algorithm first uses pixel-wise frame differencing to extract stable pixels. Then, an Artificial Neural Network (ANN) is trained to detect logolets (12x12 image regions which are parts of logos) in a frame, and finally, logolets are combined to construct the logo. In [3], one starts with frame difference images which includes initial rough masks and then applies a contour relaxation method to refine the rough masks. Cozar et al. [5] exploit both temporal and spatial features. Temporal segmentation is enabled by spotting the minimum variance regions in the luminance channel based on pixel-wise differencing. Spatial segmentation refines this area via connected components and control of bounding-box area and its aspect ratio. Ekin et al. [6] partition the screen on a 3:5:3 scale horizontally and vertically. Then they calculate scene models for each corner, such that outlier pixels qualify as logo pixels. Duffner and Garcia [7] address the transparent logo issue. They trained an ANN with positive and negative examples of a specific transparent logo. An image pyramid is used to feed pixels to ANN in different scales, and location of logo is determined according to results of the ANN.

3. TV LOGO DETECTION

Our TV logo detection method also uses stationarity characteristics of the logo regions, and similar to [1], uses time-averaged edges.

[1]This project was supported by Bogazici University Project 08A202 and TUBITAK Project 107E001.

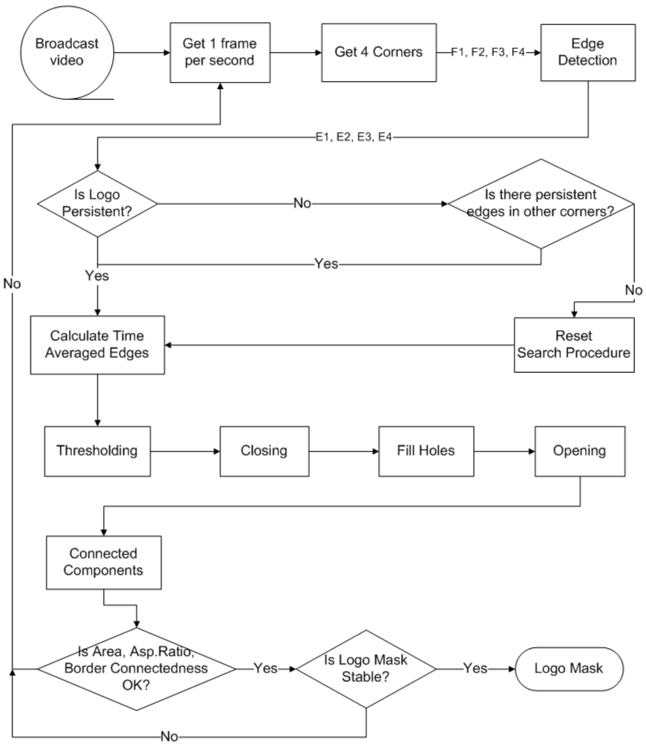


Figure 1: Flowchart of logo detection

(a) (b) (c)

Figure 2: a) Edge field of a frame, b) Logo mask, c) Time averaged edge field

As detailed in flowchart (Figure 1), we subsample frames in time at a rate of one frame per second from a broadcast video. Since logos are expected to be found in one of the frame corners, edge field is calculated only in corner regions, averaged over time, and thresholded. Thus stable contours are extracted, and smoothed via morphological operations (i.e. closing, hole filling, opening), and finally they are checked against shape constraints (area, aspect ratio, etc.). Corner regions that satisfy the constraints are promoted to logo candidates. Details of each step of the algorithm are as following:

- Corner regions: Each frame is subdivided in 3:5:3 proportions horizontally and vertically into nine regions [6] and the four corner regions are selected.
- Edge Detection: We opted for Canny edge detection method [10] especially due to its two-tier thresholding. Canny's method uses two thresholds to detect strong and weak edges, and includes the weak edges in the output only if they are connected to strong edges.
- Logo Persistence: The presence of a logo is corroborated if the edge persist from frame to frame. To this effect, a given percentage (say, 0.70) of the edge pixels comprised in the mask region at time $t-1$ should survive at time t. This persistence check step is is illustrated in Figure 2.
- Time-averaged Edges: The edge-field is time averaged as following,

$$S_i = \alpha S_{i-1} + (1-\alpha)E_i : \begin{cases} \alpha = \frac{i-1}{i}, & \text{if } i \leq n_{ref} \\ \alpha = \frac{n_{ref}-1}{n_{ref}}, & \text{otherwise} \end{cases}$$

where i is the frame index, E_i is edge-field, S_i is time-averaged edge field, n_{ref} is logo mask refresh parameter, which guarantees that the weight of current frame would not fall to a negligible level. We have used $n_{ref} = 20$ in our experiments that would enable refreshing of the logo in a reasonable time, in less than one minute.
- Thresholding: The time-averaged edge-field is binarized via hysteresis thresholding method. First strong edges are obtained with a high threshold value, then weak edges are included provided they are connected to strong edges.
- Morphological Operations: We apply closing to merge neighboring pixel groups, hole filling to prevent deformation of logo mask after opening, and finally opening to remove noise in the background. We have found useful to apply two-stage morphological operations differing in size of the structuring elements. First, bigger structuring elements are used in closing and opening, and if no logo mask is found then the operations are repeated with smaller structuring elements. For example, in our experiments we used disk structuring elements with radius r=5 as big structuring element and r=3 as small structuring element.
- Shape Constraints: TV logos possess typical shape characteristics, the basic ones being the limited ranges of their *area* and *aspect ratio*. These constraints are used to eliminate improbable shapes. Furthermore logos should be sufficiently distanced from frame boundaries.
- Logo Mask Stability: The final check consists in the stability of the logo, which means that the candidate mask should not change beyond a tolerance in area, in its coordinates and in the size of the bounding box throughout the logo search sequence.

All logo regions of interest under the detected mask are then subjected to the recognition scrutiny, hence fed to the classifier.

4. TV LOGO CLASSIFICATION

To elaborate on the feature types and classification scheme we represent the (logo) image as $I(i,j)$. The logo images are lexicographically ordered and they constitute the data matrix X. For example, for logo images of size 32x32 and the 2280 logo images, the X matrix measures 1024x2280.

We have comparatively assessed the performance of the following feature sets, namely, grid descriptors(GD), Principal Component Analysis (PCA) features, Independent Component Analysis (ICA) features (both architecture I and II), Nonnegative Matrix Factorization (NMF) features, and finally Discrete Cosine Transform(DCT). In all experiments, we have used Support Vector Machines (SVMs) as classifier, and used LibSVM [11] software package. LibSVM uses one-versus-one method for multi-class classification, which constructs C(C-1)/2 SVMs for classification of C logo classes and the winner is decided by max-wins voting mechanism. The SVMs are trained in linear mode.

4.1 Grid Descriptors(GD)

The object image(pixels within the bounding box of the logo mask) is projected onto a fixed-size grid as illustrated in Figure 3. The features are simply the average pixel values in each grid cell, essentially, macro-pixels corresponding to grid cell size. These macro-pixels are computed as

$$\theta_{kl} = \frac{1}{|C_{kl}|} \sum_{(i,j) \in Cell_{kl}} I(i,j) \qquad k = 1 \ldots K, \quad l = 1 \ldots L$$

Figure 3: A grid example

where $|C_{kl}|$ is the number of pixels in each cell, and number of cells is KxL. The GD feature vector becomes the lexicographic listing of the M macro-pixels. For color images, the macro-pixels will be listed as follows:

$$\theta_{grid} = [\theta_{R,11} \ldots \theta_{R,KL} \quad \theta_{G,11} \ldots \theta_{G,KL} \quad \theta_{B,11} \ldots \theta_{B,KL}]$$

Notice that one may need to interpolate the logo region to match it with the grid dimensions. Finally, reducing any logo region onto a given macro-pixel vector size will be instrumental in data-driven algorithms (PCA, ICA etc.) that necessitate fixed sized vector representation.

4.2 Principle Component Analysis (PCA)

The sample covariance matrix was estimated using logo training data and the principal components calculated. The number r of eigen-logo images kept represented 95% of the variance. The PCA features were obtained by projecting the grid features onto the matrix of eigen-images, so that $\theta_{PCA} = P_{PCA}\theta_{grid}$

4.3 Non-Negative Matrix Factorization (NMF)

While PCA admits both additive and subtractive components, NMF factorizes data matrices under the constraint of non-negative basis vectors and coefficients. This leads to parts-based representation of images.

The factorization of $V \approx WH$ yields basis vectors W and coefficient vectors H. The logo image is then represented as $\theta_{NMF} = (W^T W)^{-1} W^T \theta_{grid}$. The dimensionality r of the Mxr W matrix must be chosen to satisfy the discrimination ability.

4.4 Independent Component Analysis (ICA)

ICA decomposes a data matrix, X, into statistically independent source W, or mixing coefficient, H, vectors according to the chosen architecture. In ICA architecture I, X has logo images in its rows, and statistically independent, spatially local and sparse basis vectors are obtained. In ICA architecture II, statistically independent coefficients are obtained. For architecture II, the logo feature vectors become $\theta_{ICA} = W^{-1}\theta_{grid}$, where W^{-1} is the un-mixing matrix. Notice that ICA algorithm is often preceded by the vector size reduction operation via PCA; however, this step is not explicitly shown in the above formula.

4.5 Discrete Cosine Transform (DCT)

Feature vectors for logos are extracted by considering the r low-index coefficients via zigzag scan that take place in the upper left corner of the transform matrix. These coefficients are taken over several sub-blocks that partition the detected region. The corresponding feature vector consists of the concatenation low-indexed DCT coefficients belonging to the sub-blocks.

5. TV LOGO IDENTIFICATION SYSTEM

Figure 4 details the steps of the logo identification from a given video input. The scheme first extracts the logo mask followed by feature extraction and multi-class SVM classification. For each tested frame we get the p highest scoring candidates from SVM and store their scores in the likelihood vector $\lambda_c(t); c = 1, \ldots, C$, where t is the tested frame index. In the decision system, scores for T frames are accumulated and decision of the system is determined according to the highest score, as following

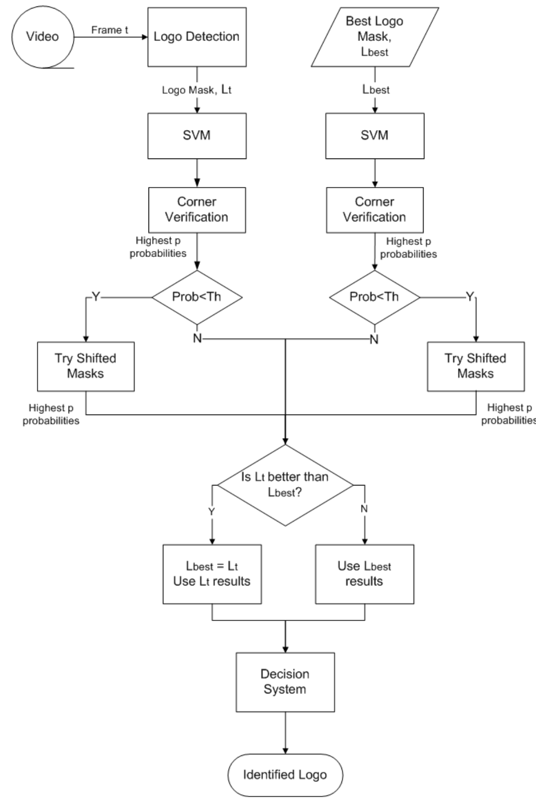


Figure 4: Flowchart of logo identification system

$$c = arg\ max \left[\sum_{t=1}^{T} \lambda_c(t) \right]$$

There are several measures to render this decision robust.

- First, since expected corner for each TV logo is known, a corner verification is applied to SVM predictions in order to filter out inconsistent predictions.
- Second, we consider several variations of the mask, that is, we shift the current mask Δk pixels (up, down, left, and right), enlarge it by Δk pixels (up, down, left, right, horizontally, vertically, and both directions) and shrink it by Δk pixels (up, down, left, right, horizontally, vertically, and both directions). However, we can skip testing all the variations if the score that the edge mask yields is sufficiently high.
- Thirdly, we preserve the highest scoring logo mask from previous frames (i.e. L_{best}), and make queries to SVM using both the best mask from the past and with the current mask.
- Fourth, the decision of the system (i.e. identified logo) should persist for n_{repeat} consecutive frames.
- Fifth, cumulative SVM probability value for T frames should be above a certain threshold value.

Image Size	GD	PCA, NMF, ICA1, and ICA2	DCT
8x8	64	27	64
16x16	256	61	128
32x32	1024	100	128
8x8	192	43	192
16x16	768	82	384
32x32	3072	121	384

Table 1: Sizes of feature vectors for gray-scale (first 3 rows) and colour images (last 3 rows)

Image Size	GD	PCA	NMF	DCT	ICA1	ICA2
8x8	95.39	93.68	86.71	91.45	91.45	95.26
16x16	97.24	96.71	91.58	96.84	94,34	97.37
32x32	97.76	97.63	94.74	97.24	96.05	97.37
8x8	96.32	95.13	92.37	95.66	95.66	97.50
16x16	98.03	97.50	97.50	97.37	96.97	98.68
32x32	98.29	98.29	97.63	97.50	97.63	**99.21**

Table 2: Classification Test Results for gray-scale (first 3 rows) and colour images (last 3 rows)

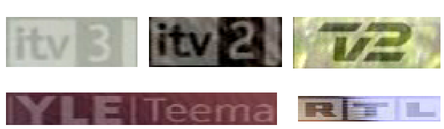
6. EXPERIMENTS AND RESULTS

6.1 Best Feature Set

A first group of experiments is conducted on TV logo images in order to determine best feature set for classification of logo images. For this purpose, we considered logos from 152 different channels and collected 20 samples for each TV channel, in total 3040 logo images. Logo database(DB) includes many similar (Figure 5.a) and transparent logos (Figure 5.b) as shown in Figure 5. Out of 20 samples, we used 15 of them for training and 5 for testing. We have used both gray-scale and rgb images to evaluate the contribution, if any, of color information.

We first investigated the feature set, namely, GD, PCA, NMF, ICA1, ICA2, and DCT features. The PCA features were determined on the basis of 95% energy content. The sizes of the NMF, and ICA feature vectors were selected the same as that of the PCA for a fair comparison. DCT is applied to 8x8 subblocks and the feature vectors are formed by the concatenation of low-indexed DCT coefficients of each subblock. Sizes of feature vectors are listed in Table 1, where leftmost column shows the size of logo images (i.e. macro-pixels or grid size). The multi-class SVM was trained with 2280 samples and tested with 760 images. Results of the experiments are shown in Table 2. Comparative analyses indicate that the best performance is obtained by using ICA2 feature on 32x32 colour logo images with an accuracy of 99.21% (i.e. 754/760).

(a)



(b)

Figure 5: Difficult TV logo cases a) Similar looking logos, b) Transparent logos

6.2 Identification Experiments

A second group of experiments is conducted on broadcast videos in order to test the proposed logo detection and identification method. In total 240 video samples were collected from 12 different TV channels. Each sample is one minute duration and is in CIF format (i.e. 352x288). Logo identification becomes more challenging problem for shorter video durations. While many works in literature have performed experiments over long-duration video records (e.g. 30 minutes or longer) we experimented with much shorter video records to achieve a more realistic application. In addition we addressed several challenging situations. These include absence of motion in the scene, overlapping TV logos, and program name text and program logos in other corners.

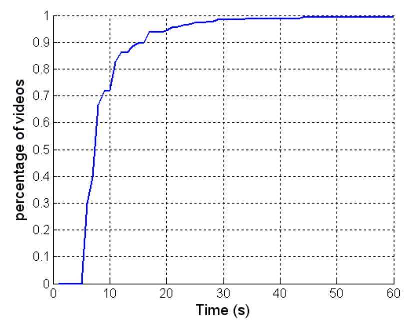


Figure 6: Cumulative logo detection histogram

Since logo detection is a time-critical operation, we measured the time of the first detection of logos in each video sample. The logo detection performance of the proposed system is shown in Figure 6 as the cumulative histogram, where the horizontal axis is the epoch from the start of the logo search process. It can be seen that in 90 percent of videos a logo is detected within 15 seconds, and that 99.17% (238/240) of them found within 1 minute. Logo identification results for various TV channels are given in Figure 7. Notice that these accuracies reflect both logo mask detection performance and that of recognition within this mask. While half of the channels score almost 100%, the average accuracy remains at 96.03%. There are two main cases for the poor performance instances. The first case is when there is no motion in the video and the background is very complex, confounding mask extraction (Figure 8a). The second case is the presence of the alternate static contours (e.g. text lines, etc.) in the proximity of the actual logo (Figure 8b). In this case, edges of those static contours become connected to the edges of TV logo and lead to deformations in the logo mask.

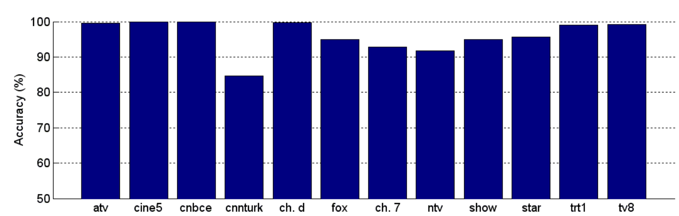


Figure 7: Logo identification results

(a) Cnnturk17

(b) Show10

(c) Details of Cnnturk17

(d) Details of Show10

Figure 8: Images a, and b shows video examples that our algorithm fails. Images c, and d shows output of each step in algorithm for each video example(Images Left to Right: Corner region of frame, After Thresholding, After Morphological Operations).

7. CONCLUSIONS

In this work, we have developed a fully automatic TV logo identification system which consists of TV logo detection and TV logo classification parts. The logo detection algorithm utilizes *time-averaged edges* to detect stable regions (TV logos) in broadcast videos. For the logo classification by using ICA2 features of 32x32 macro-pixel logo images, we achieve 99.21% accuracy when tested over a logo database. The logo identification, which entails the coupling of mask detection followed by recognition, achieves 96.03% average accuracy when tested over actual video sequences. Static video scenes with complex backgrounds and proximal text blocks remain the main impediments to fully reliable logo identification.

This work can be advanced in two avenues:

- Comparisons with competitor algorithms: It would have been desirable to give comparative performance results from alternative schemes in the literature. However we have noticed that such a comparative assessment was handicapped due to several factors and that a fair comparison was impossible. The main impediments were:
    1. The method of classification is not explicitly stated by the author, hence results are not reproducible;
    2. The term success is not consistently defined; for example, the first frame that the logo is correctly recognized is indicated success, but without any hint to its persistence;
    3. In one case, only static frames are considered, but not the video;
- We have indicated a number of difficult cases where logo detection becomes near impossible.
  We believe a scheme based on logo-ness measure can help. Thus a template can be scanned in the focus areas and the locations where the probability of logo appearance can be marked for further testing. This logo search template can be either built using a neural network to exploit nonlinear pixel dependencies or via an Adaboost scheme. The Adaboost scheme, however, would be much less laborious as compared to the Viola-Jones face detector since logos appear in one scale and only in certain focus areas.

REFERENCES

[1] Albiol, A., M. J.C. Fulla, A. Albiol, and L. Torres, "Detection of TV Commercials", Proc. IEEE ICASSP, Montreal, Canada, May 2004.

[2] Yan, W.Q., J. Wang, and M.S. Kankanhalli, "Automatic video logo detection and removal", Multimedia Systems 10(5): 379-391, 2005.

[3] Meisinger, K., T. Troeger, M. Zeller, and A. Kaup, "Automatic TV Logo Removal Using Statistical Based Logo Detection and Frequency Selective Inpainting", Proc. European Signal Processing Conference, September 2005.

[4] Santos, A.R. and H.Y. Kim, "Real-Time Opaque and Semi-Transparent TV Logos Detection", Proc. 5th Int. Information and Telecommunication Technologies Symposium (I2TS), Cuiab, 2006.

[5] Cozar, J.R., N. Guil, J.M. Gonzlez-Linares, and E.L. Zapata, "Video Cataloging Based On Robust Logotype Detection", IEEE International Conference on Image Processing (ICIP 2006).

[6] Ekin, A. and R. Braspenning, "Spatial detection of tv channel logos as outliers from the content", Proceedings of SPIE – Volume 6077 Visual Communications and Image Processing 2006.

[7] Duffner, S. and C. Garca, "A neural scheme for robust detection of transparent logos in TV programs", Lecture Notes in Computer Science – II vol. 4132, pp. 14–23, Springer, Berlin 2006.

[8] Wang, J., L. Duan, Z. Li, J. Liu, H. Lu, and J. Jin, "A Robust Method for TV Logo Tracking in Video Streams", icme,pp.1041–1044, 2006 IEEE International Conference on Multimedia and Expo, 2006.

[9] Wang, J., Q. Liu, L. Duan, H. Lu and C. Xu, "Automatic TV Logo Detection, Tracking and Removal in Broadcast Video", MMM 2007, LNCS 4352, Part II, pp.63–72, 2007.

[10] Canny, J., "A Computational Approach to Edge Detection," IEEE Transactions on Pattern Analysis and Machine Intelligence, 1986. Vol. PAMI-8, No. 6, pp. 679–698.

[11] Chih-Chung Chang and Chih-Jen Lin, LIBSVM: a library for support vector machines, 2001. Software available at http://www.csie.ntu.edu.tw/ cjlin/libsvm