VIDEO CATALOGING BASED ON ROBUST LOGOTYPE DETECTION

J.R. C´ozar, N. Guil, J.M. Gonz´alez-Linares and E.L. Zapata

Dept. of Computer Architecture, University of Málaga.

ABSTRACT

In this paper a technique for video cataloging based on logo detection is shown. No a priori knowledge about shape or spatial-temporal location of logos is assumed. The method implements a new algorithm for on-line logo detection based on temporal and spatial segmentation of broadcasted videos. Temporal segmentation identifies constant luminance regions within video frames while spatial segmentation helps to refine previous segmented regions. In a final step, identified logos are searched in a database and classified into candidate or learnt logotypes. Learnt logos can be directly tracked through the video. Candidate logotypes are assigned to a cluster of similar logos. After a promotion process, all the candidate logos belonging to the same cluster are used to create a new learnt logotype.

Index Terms— Video signal processing, multimedia databases, image segmentation, pattern clustering methods.

1. INTRODUCTION

A logotype is an icon that appears superimposed in broadcasted videos making some references to the content of these videos. Thus, it can identify a TV channel, a specific program or the brand in an advertisement. Hence, logo detection is an useful tool to be used as a part of an automatic video cataloging system. The aim of our work is to identify any kind of logo included in a video and use this information in order to group sequences containing the same logo in a video cataloging system. No assumptions about the logos shape neither their spatial-temporal locations will be made.

There are many works regarding to logo detection in document analysis [1]. However, not so many are available in video processing. In [2] logo presence detection is performed to differentiate advertisements from TV programs. In [3] it is assumed that each broadcasted TV channel has some representative semantic objects, including the channel logotype, that is displayed only during news programs. They perform channel logotype detection and tracking to specify a frame as a news frame if it includes the channel logotype. They use models for several logotypes stored in a database. Information about its position and scale helps to identify the channel and the type of news. Logo images used by channels to mark news stories are used as an alternative approach for tracking news stories in [4].

Other similar works look for static images inside video sequences. For example, in [5] brand logotypes from video data are identified. They allow for certain variability in the logotype appearance, but they must be previously stored in a database. In practice, due to the high computational demands of the method, only a few logotypes can be identified simultaneously.

In this work a multistage method for video cataloging based on logo detection is presented. First stage is based on detecting Minimal Luminance Variance Regions (MLVR) in every frame through time, as it is shown in section 2. Noise can affect the correct extraction of these zones, thus a spatial segmentation stage is mandatory (see section 3) to filter this noise and even discard false logotype detections. The main novelty of our system is that it requires no previous information about the logotypes to be found. Once a logotype is detected, it is automatically searched in the database as described in section 4. Thus, when a new video is processed, if a known logotype is detected it is identified and tracked as shown in section 5. Otherwise, it is added to the database with the other known logotypes. Finally, section 6 shows results obtained with real videos. Figure 1 summarizes the whole algorithm that will be shown next.

2. TEMPORAL SEGMENTATION

A broadcasted video is a set of ordered frames through time. These frames can include a superimposed logo (or any number of them) as time varies. In our work we will refer to a sequence as a set of consecutive frames containing the same logo layout. Figure 2 shows an example of a video consisting in three sequences: 1) logo ‘smiley’ ($f_1$ frames), 2) no logo ($f_2$ frames) and 3) logo ‘heart’ ($f_3$ frames).

The objective of the temporal segmentation is to find in a video the areas in the frame with a low intensity variation through time. The fact that either the sequences starting frame or their duration are unknown complicates the detection process.

Logotypes are usually placed at any of the four corners of the frame. Moreover, their size is limited to not disturb the normal visualization of the rest of the image. Therefore, in most cases we can
reduce the computational cost without any loss of precision by analyzing only the four corners (our ROIs) instead of the whole frame. On the other hand, no significant changes occur from one frame to the next. Thus, analysis of consecutive frames can be skipped without loss of accuracy.

Minimal Luminance Variance Region (MLVR) computation is carried out by recording the maximum and minimum luminance value taken by each pixel through time. The luminance values range depends, mainly, on the opacity of the logotype and the noise introduced by the video codec as shown in figure 3. In this figure it can be seen the histograms of the ranges of luminance variation of pixels in a set of synthetic test videos. The aim of the temporal segmentation is to detect the peak at the left of the histogram, corresponding to the logo area. Pixels are considered independent through time in this study because our method only analyze distant frames (one frame per second) and with an uniform distribution of luminance in the range 0-255. These videos are built by using a theoretical model and two modes of logotype overlaid: one fully opaque (top row) and another with a transparency level of 25% (bottom row). Every column corresponds to different compression levels in MPEG-4 (Microsoft WMV format): none, 4 Mbps (low) and 100 Kbps (high). Leftmost column (no compression) corresponds to the theoretical model. It can be observed that in this case and with a semitransparent logotype, there is a peak at the lower values of the histogram due to the logotype transparency. This peak is not a delta in 0 (like with the opaque logotype), but a smaller version of the peak that corresponds to the background (the higher values of the histogram). On the other side, as the compression level grows the peaks get wider and they can even overlap when there are too much compression losses.

2.1. Logo detection in a sequence

A sequence, as defined before, can present two possible states along all its duration: presence or absence of logotype(s). In order to identify the state corresponding to a given sequence, the frames are segmented depending on the luminance variability of their pixels through time. Thus, we conclude that there is logotype presence if there are two well-defined frame regions, one with a high variability corresponding to the background, and another with a low variability corresponding to the logotype (as shown in the previous section). On the other side, we decide that there is logotype absence if the variability is enough high in almost every pixel through time.

Figure 4 shows an example of the evolution of a temporal logo segmentation in a real sequence through time. Frame regions with low luminance variance –corresponding to logotypes– are shown in black and regions with high luminance variance –corresponding to the background– have light colors. Initially, all the ROI is considered as a null variation area (black). The pixels maximum and minimum values are updated as the sequence frames are processed and the pixel luminance variation areas are modified accordingly. The last image corresponds to a detected logo as the addition of new frames do not change the segmentation results.

In this stage it is necessary to consider the minimum and maximum area a valid logotype should have. During the segmentation process, while the number of low variability pixels exceeds the maximum area (black area in the figure 4), no state is assumed as not enough changes through the frames have occurred yet. On the other hand, if this number is below the minimum area, logotype presence can be discarded in the ROI. Thus, we can consider that a logotype is detected when the number of low variability pixels does not change with new frames and remains between these minimum and maximum limits.

Two kind of errors can arise when there are not enough differences in the frame contents along the whole sequence. These errors are caused because high variations areas do not fully emerge and can not be removed from logos’ shape. If a real logo is present in the video stream, the detection is incomplete and we can consider that a Type I error (added pixels) occurs. If there is a logotype absence, false positives could be detected. These errors are represented in figure 5. The following spatial segmentation stage, introduced in section 3, will try to correct these errors.

2.2. Logo detection in a video

In a video broadcast, logotype presence and absence sequences of unknown duration alternate. These transitions introduce more complexity in the process of logotype detection, as follows:

- Logotype—non-logotype transition. The current logo detection process should be stopped before it disappears in the next sequence, otherwise a false negative error will take place.
- Non-logotype—logotype transition. It could be a problem only if there are high variability pixels in any frame region where the logotype appear previous to the true logotype ap-
Fig. 5. Examples of error in logotype detection: (a) False positive, (b) Type I error (inside gray circle), (c) Type II error (inside gray circle), (d) codification.

Fig. 6. Examples of similarities obtained using equation 1. Top-left image is the reference image.

The filtered binary image is analyzed and connected components are found. Several properties of these like area, compactness or aspect ratio of their bounding box are also used to remove those that their values are not in a valid range [6]. After this filtering, many of the erroneous logos and false positives will be discarded.

As a result of this stage, a binary image of the shape is obtained. This image will be searched in the logotypes database in the following stage.

4. LOGOTYPES DATABASE SEARCH

The aim of this work is the cataloging of video sequences based on logotype information. Thus, the logo detected in the previous stage must be searched in a logotype information database. This database is created from scratch and updated as new logos appear in the videos. Database building is implemented by a robust mechanism that considers two possible states of the logos: candidate and learnt.

Different detections of the same logo can have different appearances caused by errors during the segmentation processes. In order to alleviate this problem, a new logo is not added definitively to the database until several similar instances of the same logo are detected.

New logos are assigned to a logo cluster that contains similar logos (or a new cluster is created if it was its first detection). Clusters are represented by the mean average binary shape of their members, \( L_c \). The expression used to calculate logos likelihood \( L(d, c) \) between a detected logo \( L_d \) and a logo cluster \( L_c \) is given by:

\[
L(d, c) = \frac{c_{dc}}{A_m} \cdot \left( 1 - \frac{D_{dc}}{A_m} \right) \text{ if } D_{dc} \leq A_m
\]

where \( c_{dc} \) is the number of pixels present at the same location in \( L_d \) and \( L_c \) simultaneously; \( A_m \) is the mean area of the logos; and \( D_{dc} \) is the absolute difference of their areas (zero if the difference is above the mean area). Figure 6 shows an example of the similarities obtained using equation (1). First logos in the ranking belong to the same cluster.

The logo’s bounding boxes will be used as a first approximation of their shapes to discard more different logo clusters. Then, the search will be refined by comparing the shapes at pixel level with (1). Temporal information is also used to speed up the search of the cluster a logo belongs, because when a video is being analyzed, it is very probable to find similar logos in consecutive order from the same sequence.

Once a cluster has enough similar logotypes, its mean average is labeled as learnt logotype. Figure 7 shows an example of this sit-
Fig. 7. Cluster of candidate logotypes (first three) promoting to learnt logotype, with noise removed after combination (right).

Pixels appearing only in one logo (o) are considered as not belonging to the final logo. Pixels appearing in all instances (s) are considered as part of the combined logo. For the rest of the pixels (n), the more times they appear, the more probably they are no noise. Thus, a SNR for the combined logo can be estimated by: 

$$SNR = \frac{s}{n+o}.$$

Detected logos after the segmentation process are searched in the logotype database. If the found logo corresponds to a learnt logo, the video is annotated and a logo tracking process is started up. Otherwise, this logo is added to the database as candidate in the corresponding logo cluster. In this situation, the validity of this cluster is reinforced, and it can even promote to a learnt logotype.

5. TRACKING LOGOTYPES

When a logotype is identified, it is necessary to check its presence until it disappears. If the logotype disappears all the process is restarted, although it is convenient to consider that the next logotype will probably be the same. Information about the time of appearance or disappearance, duration time, etc. can be used to enrich semantic information of cataloged videos.

6. EXPERIMENTAL RESULTS

In order to test the validity of our system we have analyzed 5400 frames (four 64x64 ROIs at their corners at a rate of 1 frame per second) of broadcasted video from nine different European TV channels of CIF size. In these videos, 175 possible logotypes were detected in the temporal segmentation stage. After spatial segmentation and filtering 52 images were rejected: 15 corresponding to wrongly detected logos (transitions, etc.) and 37 originated by sequences of low variability frames (false positives). The rest of the images are different instances of 14 valid logos and 19 low variability frames sequences.

Table 1 summarizes the most important results for the logo database processing stage. Second column of table 1 shows the number of instances of each logo stored in the candidates database. The number of logos belonging to the same cluster as a function of the similarity threshold –obtained by equation (1)– and using as reference logo its first detection is depicted in columns 3 to 5 of this table. Using a low threshold (50%) allows to correctly cluster almost all the logos in the database, including those having a big number of noisy pixels. Lowers values introduces misclassification. Finally, the last column represents the SNR computed using expression described in section 4 for the clusters of candidates logos that promote to learnt. No promotion (np) occurs for clusters with a low number of members or a low SNR value.

Visual aspect of learnt logos can be checked in figure 8. It can be seen that noise is removed after promoting the logotypes.

7. CONCLUSIONS

A new process for video cataloging based on logo detection has been introduced. It uses a multistage process to apply temporal and spatial segmentation that increases the accuracy of the segmentation process. Additionally, detected logos are classified as learnt or candidate depending on if previous similar logos were detected or not. In the first case, logo is available for tracking process. In the second case, logo is stored and its validity is checked for promotion with previous and possibly subsequent candidate logos. Experiments show the robustness of the proposed method.

8. REFERENCES