Available online at www.sciencedirect.com
ScienceDirect
Signal Processing: *Image Communication* 22 (2007) 669–679
SIGNAL PROCESSING: IMAGE COMMUNICATION
www.elsevier.com/locate/image

Logotype detection to support semantic-based video annotation

J.R. Cózar[a,*], N. Guil[a], J.M. González-Linares[a], E.L. Zapata[a], E. Izquierdo[b]

[a]*Department of Computer Architecture, University of Málaga, Spain*
[b]*Multimedia and Vision Research Group, Electronic Engineering Department, Queen Mary, University of London, UK*

Received 29 May 2007; accepted 30 May 2007

Abstract

In conventional video production, logotypes are used to convey information about content originator or the actual video content. Logotypes contain information that is critical to infer genre, class and other important semantic features of video. This paper presents a framework to support semantic-based video classification and annotation. The backbone of the proposed framework is a technique for logotype extraction and recognition. The method consists of two main processing stages. The first stage performs temporal and spatial segmentation by calculating the minimal luminance variance region (MVLR) for a set of frames. Non-linear diffusion filters (NLDF) are used at this stage to reduce noise in the shape of the logotype. In the second stage, logotype classification and recognition are achieved. The earth mover's distance (EMD) is used as a metric to decide if the detected MLVR belongs to one of the following logotype categories: *learned* or *candidate*. *Learned logos* are semantically annotated shapes available in the database. The semantic characterization of such logos is obtained through an iterative learning process. *Candidate logos* are non-annotated shapes extracted during the first processing stage. They are assigned to clusters grouping different instances of logos of similar shape. Using these clusters, false logotypes are removed and different instances of the same logo are averaged to obtain a unique prototype representing the underlying noisy cluster. Experiments involving several hours of MPEG video and around 1000 of *candidate* logotypes have been carried out in order to show the robustness of both detection and classification processes.

Keywords: Semantic video classification; Minimal luminance variance; Non-linear diffusion filter; Earth mover's distance; Logotype classification

1. Introduction

Logotypes convey information that can be crucial to infer semantics implicit in broadcasted videos. It is a common practice for broadcasters and specific TV programs, e.g., news, talk shows, advertisements, etc., to superimpose a specific logotype on the broadcasted material. Usually such logotypes refer to the actual video content or the content creator and, as such, they can be used to support automatic semantic-based video annotation. Logotypes extracted from broadcasted videos can be annotated in a database by indicating their shape and shots or scenes where they appear. Then, video retrieval tools can be used to search for a particular TV program or to group different pieces of video

*Corresponding author. Tel.: +34 95 2132822; fax: +34 95 2132790.
E-mail addresses: julian@ac.uma.es (J.R. Cózar), nico@ac.uma.es (N. Guil), gonzalez@ac.uma.es (J.M. González-Linares), ezapata@ac.uma.es (E.L. Zapata), ebroul.izquierdo@elec.qmul.ac.uk (E. Izquierdo).

0923-5965/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.image.2007.05.006

material with related contents. As a consequence, accurate logotype detection can be efficiently exploited for semantic-based video classification, video retrieval, aggregation and summarization.

In this paper, a framework to support semantic-based video classification and annotation is described. The backbone of the proposed framework is a technique for logotype extraction and recognition. The aim is to accurately extract and recognize opaque (or semitransparent) static logos embedded in a video sequence and to use this information in a video cataloguing system to annotate sequences containing the same logo. In practical applications, new logotypes are continuously created as new channels and TV programs are broadcasted. Thus, an efficient and robust tool for logotype detection should be able to continuously update the metadata base in an iterative process in which new logos are learned as they are detected and classified.

Many works related to logo detection in document analysis have been reported in the literature [4]. However, few approaches consider logo detection in conventional video sequences. Logo detection techniques have been used to differentiate advertisements from TV programs in [1]. This approach assumes that a logo exists if an area with stable contours can be found in the image. The authors claim that their approach does not require any supervised training and can be easily used for any type of logos without human interaction. In [5], a neural network is trained using two sets of logo and non-logo examples to detect a transparent logo. It obtains a good detection rate at the expense of a rather large training set. In [7], color outliers are used to detect pixels different from the background. No temporal information is used, thus many false detections can arise. The work presented in [13] shows an application for logo removing. The logotype is detected by exploiting frame differences in video sequence. This procedure fails in video with low motion activity. In this case, however, the authors propose to use a logo database and to search for them using a Bayesian approach. The detection accuracy is improved by assuming that the probability of the logos appearing in the four corners of the video frames is higher than in the center. This prior knowledge is combined with a neural network-based classifier.

Other works argue correctly that logos can provide a helpful visual cue for finding related news stories. Usually, a logo is defined as the small graphic or picture that appears behind the anchor person on the screen. In [8], it is assumed that each broadcasted TV channel contains some representative semantic objects, including the channel logotype, that is displayed only during news programs. Channel logotype detection and tracking is performed to automatically classify news events in conventional broadcasting material. The approach relies on the use of logotype models stored in a database. Information about logo position and scale helps to identify the channel and the type of news. In [6], the detection of logos is used to mark news stories, as an alternative approach for tracking them. Here, from each logo, three sets of 2D Haar coefficients are computed (one for each of the RGB channels). The logo's feature vector is formed by selecting the coefficients representing the overall averages and the low frequency coefficients of the three color channels. However, in all these works, logos must be known a priori.

There are related works that try to identify known brand logotypes in video data. In [9], certain variability in the logotype appearance must be allowed. In practice, due to the high computational cost of this method, only a few logotypes can be identified simultaneously. Similarly, in [3], a search for specific instances of brand logos is performed. Logo detection is achieved by exploiting homogeneously colored regions surrounding large intensity frame differences.

Contrasting these and other approaches from the literature, this paper presents a multistage method for video cataloguing based on logo detection and recognition. The proposed approach consists of two processing stages. The first stage detects the minimal luminance variance region (MLVR) in every frame by applying temporal segmentation to the luminance variance image (LVI). Since noise affects the correct extraction of potential logos, spatial segmentation is performed to filter noise and discard false positives. Detected MLVRs are then compared to an available database of previously annotated or *learned* logotypes. If a match is found, the video annotation process is triggered. Otherwise, the MLVR is labeled as a *candidate* logotype. *Candidate* logotypes are defined as potential logos that need to be validated. The validation is performed during the classification process. Here, similar *candidate* logotypes are grouped together. Specific logotypes are then promoted to *learned* logotypes, if specific conditions are satisfied. Otherwise, the corresponding shapes are declared as false positives and discarded. Fig. 1 outlines the proposed system and