Automatic TV Logo Detection, Tracking and Removal in Broadcast Video

Jinqiao Wang[1], Qingshan Liu[1], Lingyu Duan[2], Hanqing Lu[1], and Changsheng Xu[2]

[1] National Lab of Pattern Recognition
Institute of Automation, Chinese Academy of Sciences, Beijing 100080,China
{jqwang, qsliu, luhq}@nlpr.ia.ac.cn
[2] Institute for Infocomm Research, 21 Heng Mui Keng Terrace, Singapore 119613
{lingyu, xucs@i2r.a-star.edu.sg}

Abstract. TV logo detection, tracking and removal play an important role in the applications of claiming video content ownership, logo-based broadcasting surveillance, commercial skipping, and program rebroadcasting with new logos. In this paper, we present a novel and robust framework using tensor method for these three tasks. First, we use tensor based generalized gradient and the OTSU binarization algorithm to logo detection, and propose a two level framework from coarse to fine to tracking the TV logos. Finally, we extend the regularization PDEs by incorporation of temporal information to inpaint the logo region. Due to the introduction of the structure tensor, the generalized gradient based method can detect the logo region by tracking the change rate of pixels in spatio-temporal domain, and the region of logo removal is well filled in a structure-preserving way. Since temporal correlation of multiple consecutive frames is considered, the proposed method can deal with opaque, semi-transparent, and animated logos. The experiments and comparison with previous methods are conducted on the part of TRECVID 2005 news corpus and several Chinese TV channels with challenging TV logos, and the experimental results are promising.

1 Introduction

TV stations often use a special logo (i.e., TV logo) to distinguish their broadcast video from others, so TV logo plays an important role in claiming video content ownership. Besides this, TV logo detection, tracking and removal can also be used to detect commercial video, monitor the TV signal status, and rebroadcast the programs with a new logo. Several works have been developed for logo detection and removal. Meisinger *et al.* [1] used the difference image between consecutive frames to extract the logo mask with an assumption that the video content changes over time except the logo, and the frequency selective extrapolation technique was employed for logo in-painting. But such an assumption implied that this method could only handle the opaque logo. Yan *et al.* [2] utilized a learning approach (i.e., neural network) to classify candidate logo regions as True or False by using local color and texture features. Overlapping and inpainting methods were used to logo removal, separately. Similar to [1] difference images were used to determine the candidate logo regions. This learning-based approach relies on large amounts of manually labeled samples for training. Albiol *et al.* [3] used the time-averaged gradients of a series of successive images plus morphological operations to extract the coarse mask of a TV logo and only one frame (the last one) is selected from a shot to perform the gradients-based matching within the coarse logo mask. In [4], we proposed a robust logo tracking approach based on generalized gradient, but only the opaque and semi-transparent TV station logos are considered.

In this paper, a framework incorporating logo detection, logo tracking and logo removal using tensor method is proposed, and it can deal with opaque, semi-transparent, and animated TV logos. We first calculate the generalized gradients over a sequence of images to alleviate the noisy edges from the cluttered background and enhance the incomplete contour (i.e., remove partial occlusion from blending) by temporal accumulation. Then the OTSU algorithm [5] is used to locate and extract the mask of a logo from the generalized gradient image, for it is robust to low contrast, variable background intensity and noise. The template matching is utilized to logo detection and a two-level framework is used to logo tracking from coarse to fine. Finally, the regularization PDEs [6] are employed to fill the TV logo regions of video frames with a structure tensor to preserve the edge information and local geometric properties. The proposed algorithm for logo detection, logo tracking and logo removal is analyzed in detail and compared with previous methods.

2 TV logo detection and tracking

Most previous work only focused on the logo detection in static images, which often failed for semi-transparent and animated logos. In this paper, the temporal context is incorporated to enhance logo template modeling and matching for tracking the existence or absence of TV logos in broadcast video. The generalized gradient [4] is referred to the temporal extension of traditional gradient detection from a single image to a sequence of images. Different from simply averaging the gradients of multiple frames over time [3], we employ the technique of tensor gradient of a multi-image [7]. Explicit formulas for the direction along which the rate of change is maximum, and for the maximum rate of change itself, over multiple images in a video, can be derived.

2.1 Generalized gradient

We treat a video segment as a multi-valued image by modeling as an array of ordinary color image. The gradient calculation in a multi-valued image is as follows. Let $I(x_1, x_2) : R^2 \rightarrow R^m$ be a multi-valued image with components $I_i(x_1, x_2) : R^2 \rightarrow R, i = 1, 2, ..., m$. For a color image, there are R,G,B components, namely $(m = 3)$. Assuming a video segment consisting of n frames, each frame having 3 color components. Through integrating temporal information from the time axis, a video segment is expressed as: $I(x_1, x_2) : R^2 \rightarrow R^{3n}$. The image value at a given spatial point (x_1^1, x_2^1) is a vector in R^{3n}. The difference between two image values at the points $M = (x_1^1, x_2^1)$ and $N = (x_1^2, x_2^2)$ is denoted by $\triangle I = I(M) - I(N)$. By dealing with $M - N$ as an infinitesimal displacement, the image value difference becomes the differential $dI = \sum_{i=1}^{2} (\partial I/\partial x_i) dx_i$. For each point $X = (x_1, x_2)^T$, the squared norm dI^2 is called the first fundamental form and is given by:

$$dI^2 = X^T G X \quad where \quad G = \sum_{j=1}^{3n} \nabla I_j \nabla I_j^T \tag{1}$$

where G is a structure tensor, and ∇I_j corresponds to the spatial gradient of the jth color component. dI^2 is a measure of the rate of change in the multi-value image. For a spatial point (x_1, x_2), the orthogonal eigenvectors $\phi_\pm$ of tensor G provide the direction of maximal and minimal change in the spatio-temporal domain, and the eigenvalues $\lambda_\pm$ of tensor G are the corresponding rates of change in the temporal and spatial domain. For a sequence of images, the resulting edge is not simply given by the rate of maximal change λ_+, but by comparing λ_+ with λ_- in the form $f = f(\lambda_+ - \lambda_-)$ [8]. Since $f(\lambda_+ - \lambda_-)$ is the analog multi-spectral extension of $f = f(\|\nabla I\|^2)$ for single gray images($i.e., m = 1$), it reduces to the gradient-based edge detector.

By jointly considering R, G, B color components and employing temporal accumulation, the generalized gradients enhance the persistent edges belonging to a TV logo. It helps remove or weaken the noisy edges from changing background video content, as those edges are instable over time. As illustrated in Fig. 1, the energy distribution of enhanced edges at the CCTV-4 channel logo is stronger than that at the background area in the generalized gradient image.

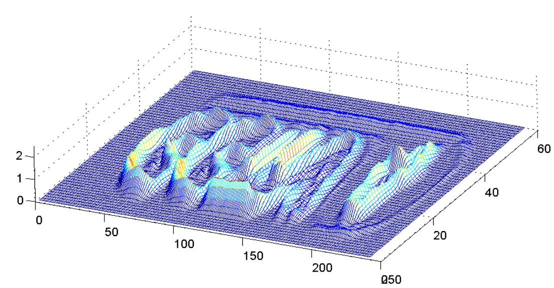


Fig. 1. Energy distribution of edges for CCTV-4 logo using generalized gradients

2.2 Logo mask extraction

TV logos are often fixed and composed of very few image pixels, such as several hundred pixels for CIF image size in MPEG-1 video streams. In order to make full use of the mask information, we consider the logo in sub-pixel space. We first enlarge the generalized gradient image by triple using bilinear interpolation. and then the OTSU algorithm [5] is used to automatically binarize the image with an optimal threshold. Given an image represented by L gray levels $[1, 2, 3, ...L]$. The number of pixels at level i is n_i and the total number of pixels is N. The gray-level histogram is normalized and regard as a probability distribution $p_i = \frac{n_i}{N}$. The OTSU algorithm aims to maximize discriminant measure variable of the gray image. By utilizing the zeroth cumulative moment $\omega(k) = \sum_{i=1}^{k} p_i$ and the first cumulative moment $\mu(k) = \sum_{i=1}^{k} ip_i$ of the histogram, the optimal threshold k^* can be obtained by the discriminative criterion: $\sigma^2(k^*) = \max_{1 \le k \le L} \left(\frac{(\mu(L)\omega(k) - \mu(k))^2}{\omega(k)(1-\omega(k))} \right)$.

The OTSU algorithm is originally used in the binarization of document images. Like document images, TV logo regions show similar visual characteristics and pose similar challenges such as low contrast, variable background and noise. Moreover, the OTSU method is nonparametric and unsupervised, so it is very efficient and effective for automatic extracting the logo mask. Fig. 2 illustrates the mask extraction of a semi-transparent CCTV logo.

Fig. 2. TV logo mask extraction (Left to right: a frame from the original sequence, the generalized gradient image, and the logo's binary mask by OTSU algorithm)

2.3 TV logo detection

In consideration of the constant position of a TV logo, we calculate the generalized gradients over a sequence of images and apply the OTSU binarization algorithm to build the logo template. For animated logo with n frames, we separate it into several static frames by the difference of histogram to build templates. Each template is generated by calculating the generalized gradients every n frames. Such temporal accumulation helps to reduce background noises and to produce a clear contour of the TV logo. A logo template database including opaque, semi-transparent and animated logo is built for logo detection. Template matching is employed to logo detection with the matching criteria as below. Eq. 2.

$$C(I, T) = \max_{k} \left\{ \sum_{T_k(i,j)=1} I(i, j) \right\} \tag{2}$$

where $I(i, j)$ is the binary image derived from the generalized gradients of consequent frames, and $T_k(i, j)$ is the matching template. k is the number of templates in the database. For opaque and semi-transparent logo, we compute the generalized gradients from 40 frames with the sampling rate of one frame out of five consecutive frames. As for animated logo with different periods $p_1, \ldots, p_n$, we compute the generalized gradients from 40 frames with the sampling rate of one frame out of $p_i (i = 1, \ldots, n)$ frames separately, then the gradient image is binalized with OSTU method and matched with the templates. If $C(I, T) \geq Threshold$, the TV logo is existent; otherwise, the TV logo is absent.

2.4 TV logo tracking

TV logo tracking is to compensate false logo detections in the broadcast and refine the true position of logo appearance, disappearance or change. Accordingly, a two-level logo tracking scheme is proposed. At both levels, for a detected TV logo with template T (for animated logo, there are several templates $T_1, \ldots, T_n$), the existence or absence of the TV logo is decided by Eq. 2. At the first level, a coarse resolution (i.e., more adjacent frames, say 20 frames) is used to roughly determine the boundaries of segments in the absence or the existence of a logo. At the second level, a fine resolution (say 1 frame) is used to precisely locate the transition points by shifting windows backwards and forwards around current time stamp. Twin thresholds are consequently applied.

When a new logo appears in the video streams, the logo detection algorithm is first used to extract the logo mask. If no matching template is find in the logo database, we use the logo mask as a new template and tracking the logo in the following frames. If the duration of the new logo is longer than 200 minutes, we consider we detect a new logo and add it to the database. For animated logo is few used and the detection of unknown animated logo is very time-consuming for searching the period of the logo. in this paper, we mainly consider the opaque and semi-transparent logos for the new logo detection.

3 TV logo removal

In TV program rebroadcasting or other applications, the TV logo needs to be removed from the video after detection. Since TV logo as an indicator is usually small and appears in the corner of the images without occluding the content of TV programs, we consider image inpainting technique in multi-value image for logo removal, i.e., filling the logo region with its neighbor data. Here, we present a multi-valued image regularization with PDEs (partial differential equation) [6] for inpainting the logo regions, and the structure tensor G in Eq. 1 is employed to preserve the local geometry of the multi-value image discontinuities. Different from computing the tensor G in logo tracking in video segment in which the position of TV logos is fixed, the inpainting of the logo region with the neighborhood requires calculating the tensor G around the logo region using 5 frames. We extend the image inpainting method in [6] to video inpainting by computing the structure tensor in spatio-temporal domain. a multi-valued regularization PDE that respects the local geometric properties is defined as, $\frac{\partial I_i}{\partial t} = trace(TH_i), (i =$

$1, \ldots, n)$ where H_i designates the Hessian matrices of I_i and T is the tensor field defined as, $T = f_-(\sqrt{\lambda_+^* + \lambda_-^*})\phi_-^* \phi_-^{*T} + f_+(\sqrt{\lambda_+^* + \lambda_-^*})\phi_+^* \phi_+^{*T}$. $\lambda_\pm^*$ and $\phi_\pm^*$ are the spectral elements of the Gaussian smoothed structure tensor G in Eq. 1. f_+ and f_- are given as, $f_+(x) = \frac{1}{1+x^2}, f_-(x) = \frac{1}{\sqrt{1+x^2}}$.

The numerical implementation of the PDE is based on the local filtering interpretation. For each point (x_1, x_2), a spatially varying gaussian smoothing mask $G^{(T,t)}$ is applied, $trace(TH_i) = \sum_{k,l=-1}^{1} G^{(T,dt)}(k,l) I_i(x-k, y-l)$. In our experiments, 5×5 convolution kernels are selected. The masks of opaque, semi-transparent, and animated TV logos are obtained from the logo detection template after morphology operation . Examples of TV logo removal are shown in Fig. 3, in which the CCTV logo in (a) is semi-transparent. We can see that the edge information and local geometric properties are better preserved in Fig. 3.

(a) (b) (c) (d)

Fig. 3. Examples of TV logo removal based on the regularization PDE. (a) and (c) are original frames with TV logos. (b) and (d) are frames with removed TV logos.

4 Experiment

Our experimental video data is extensively collected from TRECVID'05 news video corpus and several challenging Chinese TV channels. The video is in MPEG-1 format with the frame rate of 29.97 fps and the frame size of 352×240.

Referring to Fig. 4, our approach is compared with two previous algorithms: edge-based matching and pixel-wise difference computing. For each approach, the results by using different number of neighbouring frames are also compared. For edge-based matching, Canny edge detector is applied and the resulting edge images are time-averaged. In order to reduce false alarms, the OTSU method is employed to derive the final edge mask instead of morphology operators [3]. For pixel-wise difference computing, the gray-level difference images are temporally accumulated. The OTSU method is also applied to get the final mask. As illustrated in Fig. 4, our approach produces a solider and clearer contour with 40 frames to calculate generalized gradient than the other approaches. When the neighbouring frames are 150, all the three approaches get a clear contour. For relatively fewer neighbouring frames, our approach generally delivers better results. The number of neighbouring frames is decided by an application as it affects the temporal resolution.

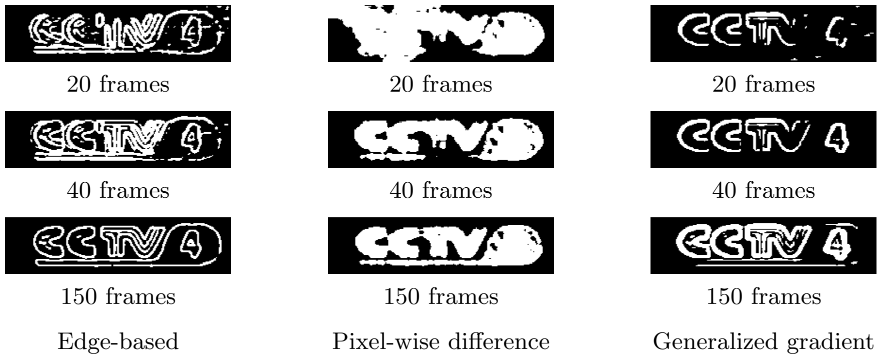


Fig. 4. Comparison of three TV logo detection algorithms

Table 1 lists the tracking results of different channels including opaque, semi-transparent, and animated logos with the proposed two level tracking framework. The duration of each channel is around 30 mins.

Table 1. Experiment results of logo tracking. TV channels including opaque (MSNBC, NTDTV, LBC, CNN and NBC), semi-transparent (CCTV-4), and animated logos (CQTV, GZTV, HEBTV) are analyzed. The false alarm rate(FAR), the false reject rate(FRR) and F1 are used to evaluate the results.

Logo types	Tne name of TV channels	FAR(%)	FRR(%)	F1(%)
Station logo	MSNBC	0.4	0	99.80
	NTDTV	2.2	0.61	98.59
	LBC	1.74	0.24	99.00
	CNN	1.03	0.24	99.36
	NBC	2.42	0.2	98.68
	CCTV-4	3.9	0.52	97.76
Program logo	CQTV	2.4	0.32	98.63
	GZTV	2.11	0.84	98.52
	HEBTV	1.8	0.62	98.79

In order to evaluate the performance of the proposed PDE based TV logo removal method, we compare it with the image inpainting method proposed by Oliveira in [9]. Fig. 5 shows some experimental results. We can see that the proposed method preserves the edge information and the homogenerity of regions better than the Oliveira's method with the same iteration $iter = 30$. An example of logo removal with the regularization PDEs based method is given in Fig. 6, in which four templates are used to model the animated logo. The local geometric properties are better preserved in the frames with the animated logo removed. As shown in Fig. 3, the TV station logo "CNN" and the TV program logo "ELECTORAL VOTES" appear in the same video segment. Although the logo "ELECTORAL VOTES" is a little bigger TV logo, the proposed method also can give a satisfied performance.

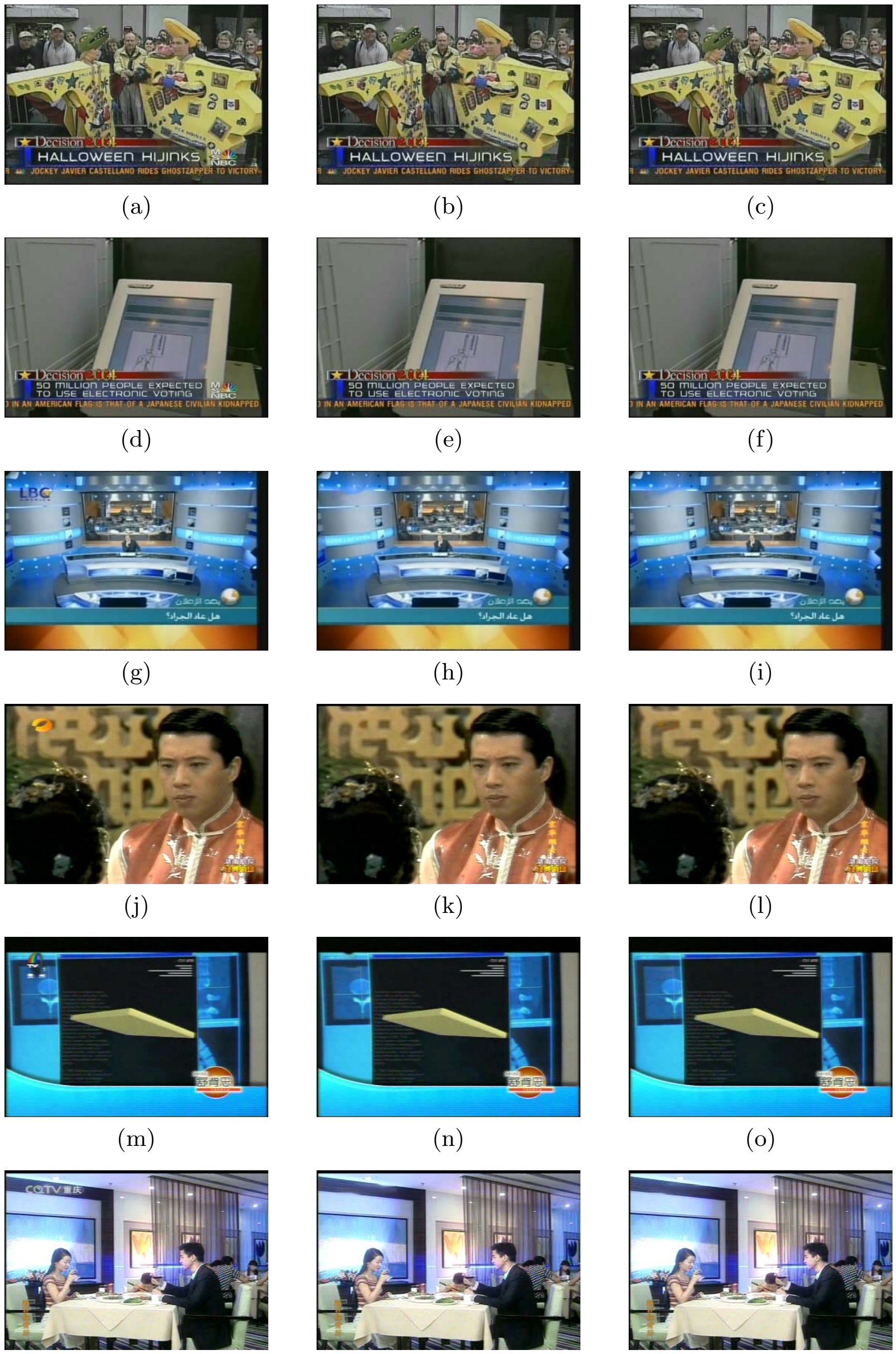
All the experiments are run in P4 3.0 GHz desktop PC with 1 GB memory. The initial TV logo detection with 40 frames with the sampling rate of one frame out of five consecutive frames takes 105 milliseconds. The average time of logo tracking is 55 milliseconds for a coming frame. Since the frame rate in the broadcast video is 29.97 fps with the frame size of 352×240, the proposed logo detection and tracking algorithm is fast enough to catch up with the real time broadcast for processing one every three frames. The average time of the proposed logo removal with the regularization PDEs is 760 milliseconds with 20 iterations for one frame, which can not reach a realtime requirement. Next step, we will improve the speed of the algorithm.

5 Conclusion

In this paper, we proposed an automatic TV logo processing framework integrating detection, tracking, and removal. The experimental results indicate the feasibility and effectiveness of our approach. Compared with edge-based matching and pixel-wise difference, our approach can use small number frames to get more satisfactory detection and tracking results. The regularization PDEs based logo removal can preserve the local geometric properties and outperform the Oliveira's approach. In the further work, we will consider classification of TV logos, and insert virtual logos in the proper position in video streams.

References

1. Meisinger, K., Troeger, T., Zeller, M., Kaup, A.: Automatic tv logo removal using statistical based logo detection and frequency selective inpainting. Proc. ESPC'05 (2005)
2. Yan, W.Q., Wang, J., Kankanhalli, M.S.: Automatic video logo detection and removal. ACM Trans. on Multimedia System (2005)
3. Albial, A., Fullà, M.J.C., Albial, A., Torres, L.: Detection of tv commercials. Proc. ICASSP'04 (2004)
4. Wang, J., Duan, L., Li, Z., Liu, J., Lu, H., Jin, J.S.: A robust method for tv logo tracking in video streams. Proc. ICME'06 (2006)
5. Otsu, N.: A threshold selection method from gray-level histograms. IEEE Trans. Systems, Man and Cybernetics **9** (1979) 62–66
6. Tschumperlè, D., Deriche, R.: Vector-valued image regularization with pdes: A common framework for different applications. IEEE Trans. Pattern Analysis and Machine Intelligence **27** (2005)
7. Zenzo, S.D.: A note on the gradient of a multi-image. Comput. Vision Graphics Image Processing **33** (1986) 116–125
8. Sapiro, G.: Vector (self) snakes: a geometric framework for color, texture, and multiscale image segmentation. Proc. ICIP'96 **1** (1996) 817–820
9. Oliveira, M.M., Bowen, B., McKenna, R., Chang, Y.S.: Fast digital image inpainting. Proceedings on the International Conference on Visualization (2001)



(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

(j) (k) (l)

(m) (n) (o)

(p) (q) (r)

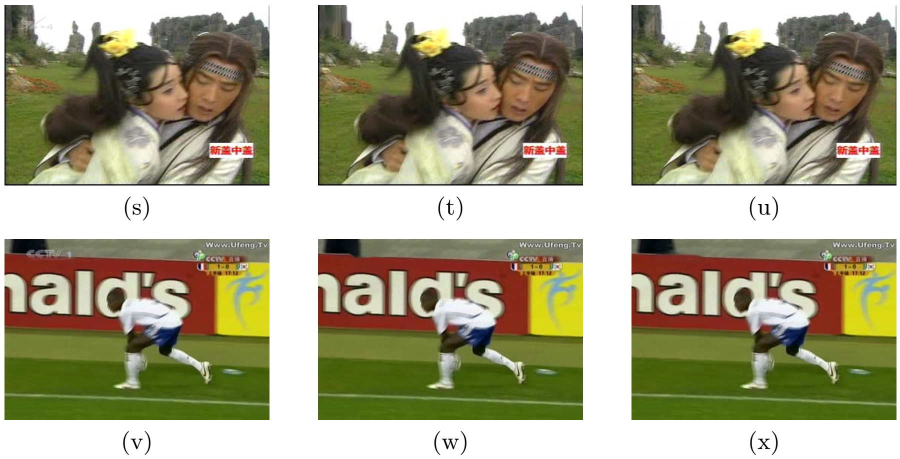


Fig. 5. Comparison between the regularization PDEs based approach and the Oliveira's approach. (a), (d), (g), (j), (m), (p), (s) and (v) are original frames with TV logos from different TV channels. (b), (e), (h), (k), (n), (q), (t) and (w) are frames with removed TV logo using the Oliveira's approach. (c), (f), (i), (l), (o), (r), (u) and (x) are frames with removed TV logo using the regularization PDEs based approach. From the comparison, we can see that logo removal with the tensor based regularization PDEs can better preserve the edge information and local geometric properties than using Oliveira's approach.

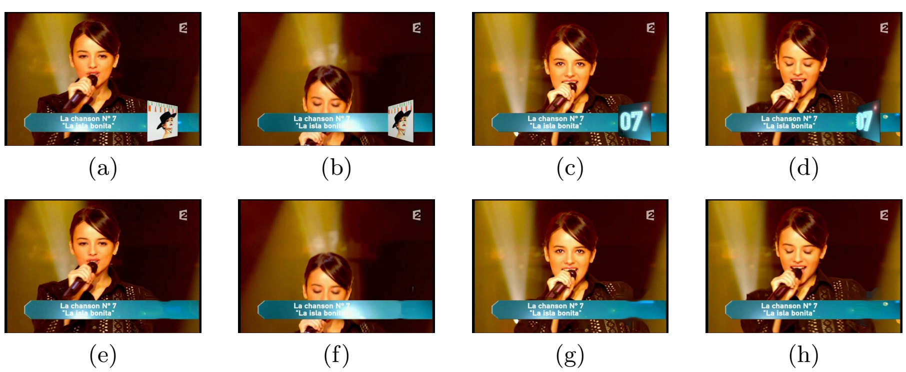


Fig. 6. Examples of TV logo removal for animated logos. Four logo templates that are obtained by difference of logo images are used to model the animated logo. (a), (b), (c) and (d) are frames with the animated logo. (e), (f), (g) and (h) are frames with removed logos with the regularization PDEs based approach.