GRADUAL TRANSITION DETECTION FOR VIDEO PARTITIONING USING MORPHOLOGICAL OPERATORS

VALERY NARANJO[1], JESÚS ANGULO[2], ANTONIO ALBIOL[1], JOSE M. MOSSI[1], ALBERTO ALBIOL[1] AND SOLEDAD GÓMEZ[1]

[1]Departamento de Comunicaciones, Universidad Politecnica de Valencia, Camino de Vera s/n, E-46022 Valencia, Spain, [2]Centre de Morphologie Mathématique, Ecole des Mines de Paris, 35, rue Saint-Honoré, F-77305 Fontainebleau, France
e-mail: vnaranjo@dcom.upv.es, jesus.angulo@ensmp.fr
(Accepted June 12, 2007)

ABSTRACT

Temporal segmentation of video data for partitioning the sequence into shots is a prerequisite in many applications: automatic video indexing and editing, old film restoration, perceptual coding, *etc.* The detection of abrupt transitions or cuts has been thoroughly studied in previous works. In this paper we present a scheme to identify the most common gradual transitions, *i.e.*, dissolves and wipes, which relies on mathematical morphology operators. The approach is restricted to fast techniques which require low computation (without motion estimation and adapted to compressed sequences) and are able to cope with random brightness variations (often occurring in old films). The present study illustrates how the morphological operators can be used to analyze temporal series for detecting particular events, either working directly on the 1D signal or building an intermediate 2D image from the 1D signals to take advantage of the spatial operators.

Keywords: 1D morphological filtering, dissolve detection, geodesic reconstruction opening/closing, video shot segmentation, wipe detection.

INTRODUCTION

Temporal segmentation of video data for partitioning the sequence into shots is a prerequisite in many applications: automatic video indexing and editing, old film restoration, perceptual coding, *etc.* The detection of abrupt transitions or cuts has been widely studied in many previous works (*cf.* Brunelli *et al.*, 1999; Cotsaces *et al.*, 2006 a comparative survey). We have also implemented a new method for cut detection, presented in Albiol *et al.* (2000), which is based on differences between consecutive frames and a morphological set of filters. But the cut detection must be accompanied by specific algorithms for detecting gradual effects such as dissolves (one scene gradually disappearing while another gradually appears) and wipes (one scene gradually entering across the view while another gradually leaves). This paper is only focused on the detection of gradual transitions. Different approaches have been proposed to extract the shots defined by gradual transitions, using different algorithms and models for these phenomena (Meng *et al.*, 1995; Yeo and Liu, 1995; Demarty and Beucher, 1999; Fernando *et al.*, 1999; Lu *et al.*, 1999; Truong *et al.*, 2000; Joyce and Liu, 2006). However, the results are not completely satisfactory even for very complex techniques.

Previous works have already illustrated the usefulness of mathematical morphology to process temporal signals (video sequences, Pardas *et al.*, 1992; Naranjo *et al.*, 2004). In particular, morphological operators have also used temporal segmentation metrics to filter out the video, mainly to detect cuts (Demarty and Beucher, 1999; Llach and Salembier, 1999; Albiol *et al.*, 2000). In particular, the algorithms proposed by Demarty and Beucher (1999); Demarty (2000) dealt with different kind of transition (cuts, dissolves and geometric transitions such as wipes) and were based on the morphological filtering of a metric for dissolves and on the study of the geometry of a local difference image mask between successive frames for wipes.

In this paper we present a scheme to identify the most common gradual transitions, *i.e.*, dissolves and wipes, which also relies on mathematical morphology operators. The algorithm for dissolves is based on the computation of a simple metric between frames, which is morphologically filtered to detect the dissolve effects, in combination with the variance of the frames (the method of variance detection was proposed in Meng *et al.*, 1995) and improved in other contributions such as Yoo *et al.* (2006). The combination of morphological analysis of the evolution a metric with the modeling of variance makes our method more robust than other previous approaches based on a single parameter. The technique for wipes is totally original and uses the orthogonal

projections of the frames, filtered by reconstruction in order to define a "strip image", where the wipe transitions are identified again by morphological filtering.

The approach is restricted to fast techniques which require low computation (without motion estimation), are adapted to compressed sequences (in fact the algorithms are applied to the *dc* image, Meng *et al.*, 1995; Yeo and Liu, 1995) and are able to cope with random brightness variations (often in old films). The algorithms can be used for sequences composed of grey level or color image frames. If the image is color we can use the luminance or the sum of the RGB components to define the metrics of the algorithms. Previous works (Gargi *et al.*, 1995) have evaluated the influence of the chosen color space in the detection of cuts, and the representation luminance, saturation and hue seems to give the best performance. We have tested our algorithms using only the luminance or the luminance together with the hue and the saturation but no improvement is obtained.

The organization of the rest of the paper is as follows. Section "Methods" is decomposed into several parts: first, it is introduced the notation and a brief reminder on morphological operators for temporal series; second, the method for dissolve detection is introduced; then, the approach for wipe detection is presented. The description and the analysis of the experimental results using our transition detector are discussed in section "Results", where we will present not only the results of our detectors of gradual transitions, described in this paper, but also the results of the detector of cuts proposed in Albiol *et al.* (2000). Finally, in section "Discussion", some conclusions and perspectives are given.

METHODS

MORPHOLOGICAL OPERATORS FOR TEMPORAL SERIES

Image and signal lattices

Let $\{f_t(x,y)\}_{t=1}^N$ be a video sequence of *N* frames, where the frame *t* is a grey level or color image $f_t(x,y)$. Assume that $s(t)$ is an equidistant time series (1D signal). We work in this paper on images and time series, therefore we need to precise some notations. Let us consider two complete lattices: $\mathscr{L}_{image}$ and $\mathscr{L}_{signal}$. An image is a function $f(x,y) : E \rightarrow \mathscr{L}_{image}$ where the spatial domain is a discrete set $E \subset \mathbf{Z}^2$, $1 \leq x \leq X, 1 \leq y \leq Y$ (*X* and *Y* are the number of image columns and rows respectively) and the image

lattice is an ordered set of grey levels $\mathscr{L}_{image} \subset \mathbf{Z}$ (or $\subset \mathbf{Z}^3$ for color images). A temporal signal is a function $s(t) : T \rightarrow \mathscr{L}_{signal}$ where $T \subset \mathbf{Z}$ is the discrete time index, *i.e.*, $T = \{1 \leq t \leq N\}$) with real values into the signal lattice $\mathscr{L}_{signal} = \mathbf{R}$.

We also consider that for each morphological operator Ψ (Serra, 1982; 1988; Soille, 1999) we may associate the image mapping $\Psi_B^E : \mathscr{L}_{image} \rightarrow \mathscr{L}_{image}$ (where *B* is the size/shape of flat structuring element) or the signal mapping $\Psi_{\Delta t}^T : \mathscr{L}_{signal} \rightarrow \mathscr{L}_{signal}$ (where Δt is the size or length of the temporal structuring element). We remind in the rest of this section the main morphological operators $\Psi_{\Delta t}^T$ for temporal series.

Temporal erosion and dilation

The basic morphological operators for temporal series are

- Erosion: $\varepsilon_{\Delta t}^T(s(t)) = \{s(y) : s(y) = \wedge[s(z)], z \in \Delta t\}$,

- Dilation: $\delta_{\Delta t}^T(s(t)) = \{s(y) : s(y) = \vee[s(z)], z \in \Delta t\}$,

where Δt is the temporal structuring element, which is typically an odd symmetric centered time window, *i.e.*, $[t_0 - \Delta t/2, t_0 - \Delta t/2 + 1, \cdots, t_0 - 1, t_0, t_0 + 1, \cdots, t_0 + \Delta t/2 - 1, t_0 + \Delta t/2]$.

The erosion and the dilation are increasing operators, *i.e.*, $s_1(t) \leq s_2(t) \Rightarrow \varepsilon_{\Delta t}^T(s_1(t)) \leq \varepsilon_{\Delta t}^T(s_2(t))$, $\forall t$. Moreover, the erosion is anti-extensive, *i.e.*, $\varepsilon_{\Delta t}^T(s(t)) \leq s(t)$; and the dilation is extensive $s(t) \leq \delta_{\Delta t}^T(s(t))$. In practice, the erosion shrinks the positive structures; "peaks of signal" shorter than the structuring element disappear by taking the value of remaining neighboring signal structures. Dilation produces the dual effects, enlarging the positive peaks of signal. Fig. 1b shows two examples of erosion/dilation of sizes 3 and 7 for the same time series.

Temporal opening and closing, and derived operators

The two elementary operations of *erosion* and *dilation* can be composed together to yield a new set of operators having desirable feature extractor properties which are given by

- Opening: $\gamma_{\Delta t}^T(s(t)) = \delta_{\Delta t}^T[\varepsilon_{\Delta t}^T(s)]$,

- Closing: $\varphi_{\Delta t}^T(s(t)) = \varepsilon_{\Delta t}^T[\delta_{\Delta t}^T(s)]$,

The morphological *openings* (*closings*) filter out positive (negative) peaks (1D structures) from the signals according to the predefined length of the temporal structuring element, see the examples given in Fig. 1c using again two sizes of operators. The opening (closing) is an anti-extensive (extensive)

operator and both are increasing and idempotent operators.

The *top-hat transformation* is a powerful operator which permits the detection of contrasted structures or relevant peaks on non-uniform backgrounds. There are two versions,

– White top-hat: The residue of the initial series s and an opening $\gamma_{\Delta t}^{T}(s)$; *i.e.*, $\rho_{\Delta t}^{T,+}(s) = s(t) - \gamma_{\Delta t}^{T}(s(t))$, extracts positive peaks.

– Black top-hat: The residue of a closing $\varphi_{\Delta t}^{T}(s)$ and the initial series s; *i.e.*, $\rho_{\Delta t}^{T,-}(s) = \varphi_{\Delta t}^{T}(s(t)) - s(t)$, extracts negative peaks.

Usually, the top-hat is accompanied by a thesholding operation, in order to binarize the extracted peaks. In addition, the main operator used for the top-hat of size Δt_1 can be preceded by the dual operator of size Δt_2, such as the effect of both operators is taken into account. In Fig. 1d is given a typical example. This temporal operator specially useful for temporal series is applied in the algorithms of the present paper.

A granulometry is the study of the size structure distribution of a time series. Formally, a *granulometry* can be defined as a family of *openings* $\Gamma = (\gamma_{\Delta t_n}^{T})_{\Delta t_n \geq 0}$ such that $\forall \Delta t_n \geq 0$, $\forall \Delta t_m \geq 0$, $\gamma_{\Delta t_n}^{T} \gamma_{\Delta t_m}^{T} = \gamma_{\Delta t_m}^{T} \gamma_{\Delta t_n}^{T} = \gamma_{\Delta t_{\max(n,m)}}^{T}$. Moreover, granulometries by *closings* (or *anti-granulometry*) can also be defined as families of increasing closings $\Phi = (\varphi_{\Delta t_n}^{T})_{\Delta t_n \geq 0}$. Performing the granulometric analysis of a series $s(t)$ with Γ is equivalent to mapping each opening of size Δt_n with a measure $\mathscr{M}(\gamma_{\Delta t_n}^{T}(s))$ of the opened series; where $\mathscr{M}(s)$ is the Lebesgue measure of the time series $s(t)$, *i.e.*, $\mathscr{M}(s) = \sum_{t=1}^{N} s(t)$. The *size distribution* or *pattern spectrum* of $s(t)$ with respect to Γ, denoted $PS_{\Gamma}(s, \Delta t_n)$ or $PS(s, \Delta t_n)$ is defined as the following (normalized) mapping

$$PS(s, \Delta t_n) = \frac{\mathscr{M}(\gamma_{\Delta t_n}^{T}(s)) - \mathscr{M}(\gamma_{\Delta t_{n+1}}^{T}(s))}{\mathscr{M}(s)}, \Delta t_n \geq 0.$$

The pattern spectrum $PS(s, \Delta t_n)$ maps each size Δt_n to some measure of the positive variations with this size (loss of positive peaks between two successive openings). The pattern spectrum $PS(s, \Delta t_n)$ is a probability density function: a large impulse in the pattern spectrum at a given time scale indicates the presence of many peaks at that time scale. It is also possible to use standard probabilistic definitions to compute the moments of $PS(s, \Delta t_n)$. An example of $PS(s(t), \Delta t_n)$ useful to analyze the frequencies of positive peaks is given in Fig. 1e.

Temporal reconstruction

A morphological tool that complements the opening and closing operators for feature extraction (extract the marked particles) is the morphological reconstruction, implemented using the *geodesic dilation*, operator based on restricting the iterative dilation of a function marker $s_m(t)$ by the unitary temporal structuring element Δt_1 to a function reference $s_r(t)$, *i.e.*, $\delta_{s_r,(n)}^{T}(s_m) = \delta_{s_r,(1)}^{T} \delta_{s_r,(n-1)}^{T}(s_m)$, where $\delta_{s_r,(1)}^{T}(s_m) = \delta_{\Delta t_1}^{T}(s_m(t)) \wedge s_r(t)$. The reconstruction by dilation or *opening by reconstruction* is then defined as

$$\gamma^{T-rec}(s_m, s_r) = \delta_{s_r,(i)}^{T}(s_m),$$

such that $\delta_{s_r,(i)}^{T}(s_m) = \delta_{s_r,(i+1)}^{T}(s_m)$ (idempotence).
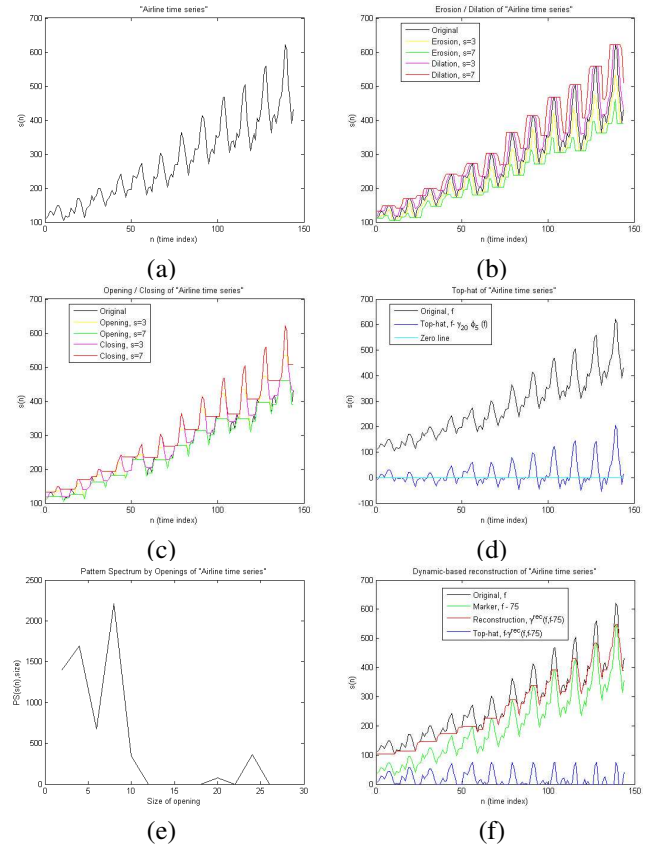

Fig. 1. *Application of temporal morphological operators to the time series "Airline" from Falk et al. (2006): (a) original time series $s(t)$; (b) erosions/dilations, $\varepsilon_{\Delta t}^{T}(s(t))$ and $\delta_{\Delta t}^{T}(s(t))$, $\Delta t = 3, 7$; (c) openings/closings $\gamma_{\Delta t}^{T}(s(t))$ and $\varphi_{\Delta t}^{T}(s(t))$, $\Delta t = 3, 7$; (d) modified white top-hat, $s(t) - \gamma_{\Delta t_2}^{T} \varphi_{\Delta t_1}^{T}(s(t))$, $\Delta t_1 = 5$, $\Delta t_2 = 20$; (e) pattern spectrum, $PS(s(t), \Delta t)$, $1 \leq \Delta t \leq 31$; (f) dynamic-based opening by reconstruction, $\gamma^{T-rec}(s(t) - H, s(t))$, $H = 75$.*

Whereas the adjunction opening $\gamma^T_{\Delta t}(s)$ (from an erosion/dilation) modifies the structures of the series, the associated opening by reconstruction $\gamma^{T-rec}(s_m, s)$ (where the marker $s_m = \varepsilon^T_{\Delta t}(s)$ or $s_m = \gamma^T_{\Delta t}(s)$) is aimed at efficiently and precisely reconstructing the "shape" of the structural peaks which are not totally removed by the marker filtering process (peaks of length Δt). Other useful markers for extracting peaks according to their dynamics correspond to series of type $s_m = s - H$ such as $\gamma^{T-rec}(s_m, s)$ will remove the peak of contrast lower than H, see the example of Fig. 1f.

DISSOLVE DETECTION

Dissolves are the most usual gradual transition between two shots (see Fig. 2 for an example). The blend between the two sequences is usually linear and involves several frames.

Fig. 2. *An example of dissolve (from the film "Torbellino").*

Linear intensity metrics, $s_\rho(t)$

Our method is based on the assumption of the following simple hypothesis: *"The intensity of the pixels in the frames of a dissolve follows a monotonous variation"*. We must then define a new metrics, $s_\rho(t)$, to quantify the monotony of consecutive frames in order to detect dissolves.

Consider the three successive frames $f_{t-1}(x,y)$, $f_t(x,y)$ and $f_{t+1}(x,y)$ of the video sequence $\{f_t(x,y)\}^N_{t=1}$. We define the following two differences:

$$d^-_t(x,y) = (f_t(x,y) - f_{t-1}(x,y))$$

and

$$d^+_t(x,y) = (f_{t+1}(x,y) - f_t(x,y)),$$

for each pixel (x,y). The coefficient of monotonous linearity is given by

$$\rho_t(x,y) = \begin{cases} 1 & \text{If } (|d^-_t(x,y)| > th \text{ and } |d^+_t(x,y)| > th) \\ & \text{and } sign(d^-_t(x,y)) = sign(d^+_t(x,y)) \\ -1 & \text{If } (|d^-_t(x,y)| > th \text{ and } |d^+_t(x,y)| > th) \\ & \text{and } sign(d^-_t(x,y)) \neq sign(d^+_t(x,y)) \\ 0 & \text{otherwise} \end{cases},$$

where th is a threshold to avoid the random variations due to noise (typically $th = 2$ yields satisfactory results). Using this pixel parameter, we can define a metrics for each frame t by computing

$$s_\rho(t) = \frac{\sum^X_{x=1} \sum^Y_{y=1} \rho_t(x,y)}{X\,Y}.$$

If the difference between the pixel (x,y) in the frame f_t and the same pixel in the frame f_{t-1} has the same sign than the difference between the same pixels in frames f_t and f_{t+1}, the luminance of this pixel varies monotonously and we can suppose that this point-to-point evolution is linear. When this situation occurs in most of the pixels of a frame, we obtain high values for s_ρ, which indicates a linear luminance variation in the whole image. Consequently, during a dissolve all its frames present high values for s_ρ. Fig. 3 shows the result of $s_\rho(t)$ calculation using a sequence from the film *Torbellino*. In order to simplify the detection of peaks in $s_\rho(t)$ we propose to carry out a 1D morphological filtering.
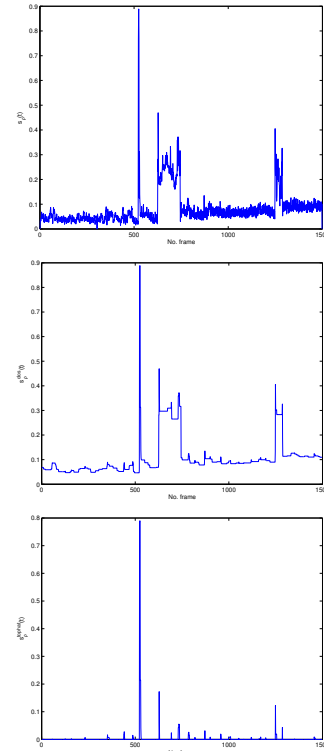

Fig. 3. *Dissolve metrics and corresponding morphological filtering (temporal closing of size 24 and temporal top-hat of size 12) using a sequence from the film "Torbellino".*

Initially a temporal closing of size Δt_1 removes the negative peaks of temporal length less than Δt_1:

$$s^{clos}_\rho(t) = \varphi^T_{\Delta t_1}(s_\rho(t)).$$

In fact, the value of Δt_1 allows us to fix the minimal duration between two dissolves, $e.g.$, $\Delta t_1 = 24$ involves a minimal distance equal to 1 second (frame rate = 24 frames/second). Then, using a top-hat of size Δt_2 the positive peaks are extracted:

$$s_\rho^{tophat}(t) = s_\rho^{clos}(t) - \gamma_{\Delta t_2}^T(s_\rho^{clos}(t)) .$$

In this case the value of Δt_2 defines the maximum duration of a dissolve, $e.g.$, taking $\Delta t_2 = 12$ corresponds to 0.5 second (typical value).

In the $s_\rho^{tophat}(t)$ of Fig. 3, we can observe a peak produced by the dissolve placed in the interval 522–525. Applying a threshold value $u_\rho = 0.15$, the dissolve is detected. However, a false alarm will also be detected in frames 626–628. These false alarms are produced by high motion objects (objects in motion which take up many pixels in the frame). In order to reduce these false alarms, our method combines the information achieved using $s_\rho(t)$ and the information of the parabolic variance evolution in a dissolve.

Detection of a parabolic variance

Let $f_t^1(x,y)$ and $f_t^2(x,y)$ two uncorrelated sequences whose intensity variance are σ_1^2 y σ_2^2, respectively. In a dissolve, the frames are obtained by the weighted average of $f_t^1(x,y)$ and $f_t^2(x,y)$ during the transition interval, in the following way:

$$f_t^{dissolve}(x,y) = f_t^1(x,y)[1 - \alpha(t)] + f_t^2(x,y)\alpha(t),$$

where $t_1 \leq t \leq t_2$ is the dissolve interval.

The weight is given by

$$\alpha(t) = \begin{cases} 0 & t < t_1 \\ (t - t_1)/(t_2 - t_1) & t_1 \leq t \leq t_2 \\ 1 & t > t_2 \end{cases} ,$$

The variance of the dissolve sequence $f_t(x,y)$ is a parabolic curve, such as for each frame t in the dissolve $\sigma^2(t) = (\sigma_1^2 + \sigma_2^2)\alpha^2(t) - 2\sigma_1^2\alpha(t) + \sigma_1^2$.
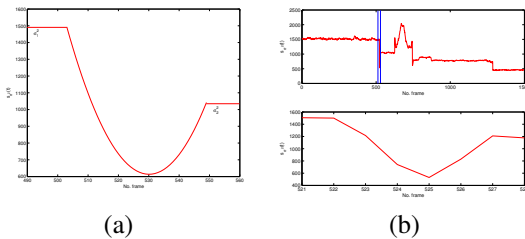

(a)   (b)

Fig. 4. *(a) Curve of variance for an ideal dissolve. (b) Variance of the sequence from the film "Torbellino" ($s_{\sigma^2}(t)$), where we can observe the dissolve placed between the vertical lines. Below, a zoom of the dissolve variance is shown.*

Ideally, in the frames belonging to a shot, the variance remains constant while the variance of the dissolve frames has a parabolic shape (Fig. 4a). In real sequences, the variance signal $s_{\sigma^2}(t)$ in the dissolve region is approximately a parabola, but in a shot the variance could not remain constant and even presents a parabolic curve. This last effect is due to the motion in the scene. Fig. 4b shows the variance of a sequence from the film *Torbellino*. The variance of the dissolve was zoomed in order to observe its shape. We can observe other regions where the variance is also a parabola.

The algorithm

In summary, the steps of the proposed algorithm to detect the dissolves are:

(1) Calculate the signals $s_\rho(t)$ and $s_{\sigma^2}(t)$ for each frame t of the sequence.

(2) Fix Δt_1 and Δt_2 and filter out $s_\rho(t)$ to obtain the signal $s_\rho^{tophat}(t)$. Then, apply a threshold u_ρ. All transitions with a value at $s_\rho^{tophat}(t)$ higher than the threshold will be candidates for a dissolve.

(3) If the difference between the ideal variance model $\sigma^2(t)$ and the obtained variance for the candidate frames $s_{\sigma^2}(t)$ is less than a threshold u_{σ^2}, the candidate transition is detected as a dissolve.

The values selected for the thresholds have been $u_\rho = 0.15$ and $u_{\sigma^2} = 350$. To obtain these values we have achieved a deep study (Angulo, 1999) on a selection of sequences to estimate the probability density functions of the signals $s_\rho(t)$ and $s_\sigma(t)$ for transition and non-transition situations. Fig. 5 shows the probability density functions of $s_\rho(t)$. The threshold is selected as the intersection point between the two curves (pdf for transition and pdf for non-transition), $i.e.$, the hypothesis selected is which originates with higher probability $s_\rho(t)$. Note that this is equivalent to use the maximum likelihood test:

$$P(H_1/x) \gtrless_{H_0}^{H_1} P(H_0/x) ,$$

where $x = s_\rho(t)$, H_0 is the hypothesis that a transition doesn't occur and H_1 is the hypothesis that a transition occurs.

The same study has been carried out for the variance signal, obtaining the intersection between the two curves at the point $s_\sigma^2(t) = 350$, so, this value is selected as threshold u_σ^2.
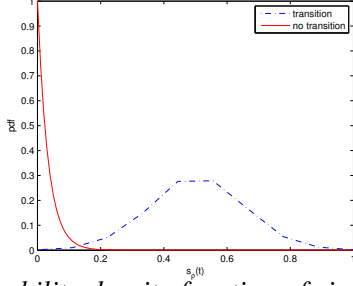
Fig. 5. *Probability density function of signal $s_\rho(t)$ for transitions and non-transitions.*

WIPE DETECTION

A wipe is a gradual transition between two shots where one image belonging to a sequence $f_t^1(x,y)$ is linearly shifted by another $f_t^2(x,y)$, and this effect lasts several frames. In each frame of a wipe, the second image superimposes W pixels on the first image, from left to right (or from right to left) if the wipe is vertical, and from top to bottom (or from bottom to top) if the wipe is horizontal. An example of a vertical wipe is given in Fig. 6.

Fig. 6. *An example of a vertical wipe where the second image replaces the first image from left to right (from the film "Torbellino").*

Fig. 7 shows us the difference $\hat{d}(x,y) = |f_t(x,y) - f_{t+1}(x,y)|$ between two consecutive frames t and $t+1$ belonging to the vertical wipe. This image difference has a width area of W pixels, where the intensity of the pixels is brighter than in all the height of the image.

Orthogonal projections and reconstruction

The first phase of our method consists of determining the position of that area within the width of the image. Using the image difference $\hat{d}(x,y)$, we have to calculate the normalized vertical projection signal by applying the following equation:

$$s_{vp}(x)_t = \frac{1}{Y} \sum_{y=1}^{Y} \hat{d}_t(x,y) \quad 0 \leq x \leq X,$$

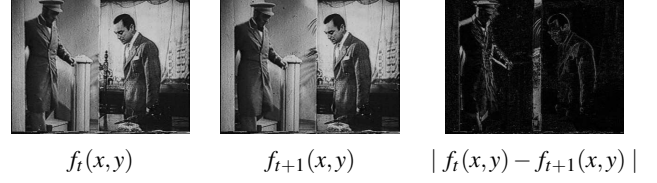where X and Y are the number of image columns and rows, respectively.

$f_t(x,y)$ $f_{t+1}(x,y)$ $| f_t(x,y) - f_{t+1}(x,y) |$

Fig. 7. *Two consecutive frames from the film "Torbellino" and the difference image between both frames.*

The result of the vertical projection, calculated using the difference image of Fig. 7, is the 1D signal shown in Fig. 8. The maximum of this signal belongs to the zone W of the wipe. In order to eliminate the other regional maxima (not associated with the region W), we apply a geodesic reconstruction (Vincent, 1993) using the vertical projection as reference image s_{vp} and a delta signal in the maximum position as a marker s_{vp}^{max}, *i.e.*, $s_{vp}^{rec}(x) = \gamma^{T,rec}(s_{vp}, s_{vp}^{max})$ (see Fig. 8). The amplitude of the maximum obtained for images that do not belong to a vertical wipe is smaller than the amplitude of the maximum obtained when the wipe is present, because the wipe causes a bright area which takes up all the column, yielding a greater maximum.
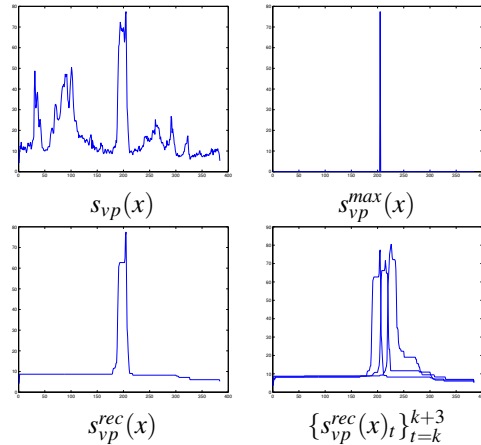

$s_{vp}(x)$ $s_{vp}^{max}(x)$

$s_{vp}^{rec}(x)$ $\{s_{vp}^{rec}(x)_t\}_{t=k}^{k+3}$

Fig. 8. *Vertical projection of the difference image between two consecutive frames and the opening by reconstruction (the marker is the maximum). Last plot, three consecutive reconstructed vertical projections.*

The position of the maximum in $s_{vp}^{rec}(x)$ $(1 \leq x \leq X)$ indicates the spatial position of the wipe. As the wipe passes from frame t to frame $t+1$, the maximum will also move, see Fig. 8. These moving maxima can be used to detect the position of the wipe.

"Strip image" creation

For a sequence of N frames and $X \times Y$ pixels per frame, we propose to create a grey image $f^{strip}(x,y)$ to visualise the consecutive projection signals $s_{vp}^{rec}(x)_t$ for each frame t. The vertical projection image has a size of $(N-1) \times X$ pixels (or $(N-1) \times Y$ pixels for horizontal projections). More precisely, each column $x = i$ (corresponding to the frame i) of the image f^{strip} is the normalised $s_{vp}^{rec}(x)_i$ between 0 and 255. The sequences to be processed have a thousand frames, so the appearance of the projection image is similar to a dark strip of the same length as the sequence ("strip image"). An example of a "strip image" is given in Fig. 9a.

"Strip image" processing

In the "strip image" we can identify several event components:

– Wipes correspond to oblique lines whose slopes indicate the temporal length of the effect. These lines arise from the spatial displacement of the maximums of the projections, and they consist of small vertical lines with a size equal to the number of pixels that the wipe advances in consecutive frames (W pixels).

– Abrupt transitions or cuts correspond to vertical lines, because all the pixels of the difference image in a cut have high intensity levels.

– Areas with large motion produce irregular shaped areas. As the motion increases, the area produced becomes brighter.

The purpose behind the processing of the "strip image" is to eliminate all those event components which do not correspond to a wipe. To achieve this objective some morphological operators are used according to the following algorithm.

(1) In order to eliminate the vertical lines produced by cuts, a vertical top-hat is used, *i.e.*, $\overline{f^{strip}} = f^{strip} - \gamma_{l_1}(f^{strip})$, where the structuring element is a vertical line of size l_1, which must be greater than the small vertical line of the oblique line caused by the wipe (typically $l_1 = 20$), see the result in Fig. 9b. Although this patterns could be used in order to detect the cuts, we would not obtain very good results due to the false alarms produced by the flicker, an artifact which appears frequently in old films. As we will see in section "Results", we have designed and implemented a robust cut detector which was presented in Albiol *et al.* (2000).

(2) Wipes, as we explained above, produce oblique lines which go from the first row to the last. To obtain these lines, eliminating irregular areas, we can compute the opening by reconstruction using $\overline{f^{strip}}$ as reference and placing a marker f_{mrk1}^{strip} in the top image border (all pixels to 0 except the two first rows to 255). So, only those areas which touch the top border will be reconstructed. Then, a second reconstruction using the lower image border as marker f_{mrk2}^{strip}, *i.e.*, $\widetilde{f^{strip}} = \gamma^{rec}(f_{mrk2}^{strip}, \gamma^{rec}(f_{mrk1}^{strip}, \overline{f^{strip}}))$. After these two geodesic reconstructions, we achieve the image $\widetilde{f^{strip}}$ where only those areas which touch both borders are presented (Fig. 9c).

(3) A threshold is applied to the image $\widetilde{f^{strip}}$ in order to obtain only the oblique lines. Finally, the binary image is segmented for labelling each connected line as an independent object (Fig. 9d). In this case, the only connected line which appears is the vertical wipe which exists in this sequence.
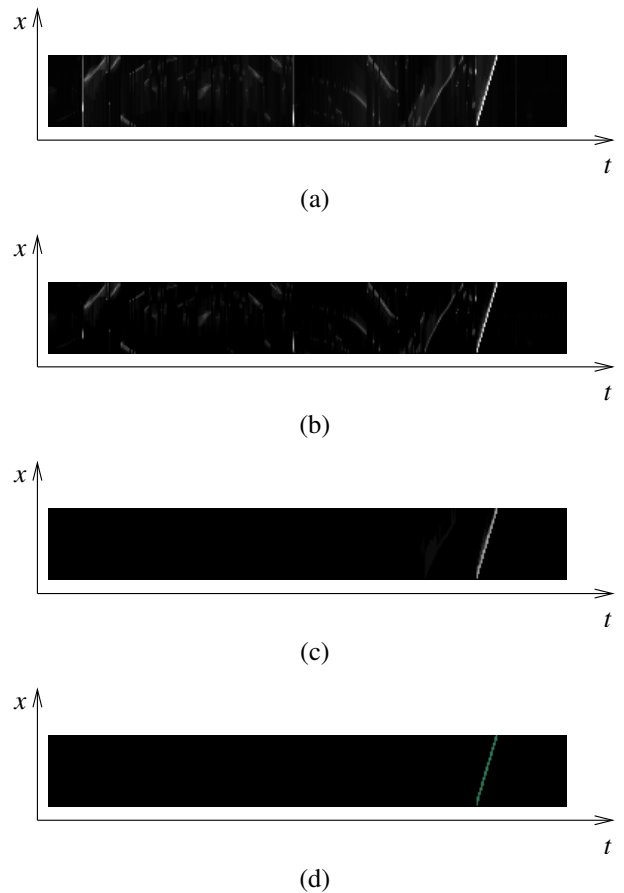

(a)

(b)

(c)

(d)

Fig. 9. *Vertical wipe detection: (a) "Strip image" f^{strip}, (b) residue of the vertical opening of size 20, (c) final processed "strip image", (d) segmentation by thresholding.*

Limitations

The process presented above can also be applied in order to detect horizontal wipes using the horizontal projections. For detecting a wipe in a different direction, the projection must be calculated in the corresponding direction. Obviously, the detection is limited to wipe patterns which change in the same direction. However, the main application of our methods is the old film restoration and, in this case, wipes in a different direction from horizontal or vertical, rarely appear.

Again, this methodology has been applied to our database of video sequences and the detection is satisfactory even for degraded sequences (in our main application), as can be observed in the results presented in section "Results". In spite of these results, some false positives can be produced if the contents of the sequence have the same characteristics than the frames belonging to a wipe. Fig. 10 shows a subsequence which produces a false positive in the vertical wipe detection. The problem arises from a dark vertical object (the door), inserted in a clear background, which moves W pixels per frame in the horizontal direction. The "Strip Image" obtained after processing the sequence which contains the false wipe is presented in Fig. 11.

Fig. 10. *Example of a sequence which produces a false positive in the vertical wipe detection.*
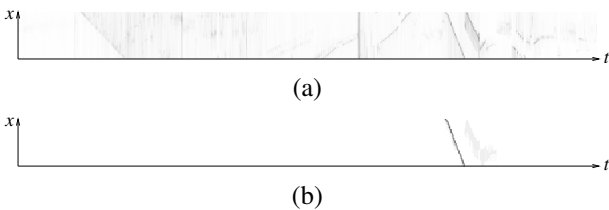

Fig. 11. *Vertical wipe detection of sequence which includes the subsequence shown in Fig. 10: (a) "Strip image" f^{strip}, (b) final processed "strip image" where the false detection is shown.*

RESULTS

The global detector, which we have designed and implemented, consists of the following process:

(1) Abrupt transition or cut detector. It is based on a metric called correlation of modified sign (CMS) which measures if there is a match in the modified sign between two consecutive frames. The equation of the CMS is:

$$CMS(t) = \frac{1}{XY} \sum_{x=1}^{X} \sum_{y=1}^{Y} MS(f_t(x,y)) - MS(f_{t-1}(x,y)),$$
(1)

where X and Y are the frame dimensions, and $MS(f_t(x,y))$ corresponds to the modified sign for frame f_t, which is defined in Eq. 2 and represents the situation of a pixel with respect to the image mean.

$$MS(f_t(x,y)) = \begin{cases} 1 & f_t(x,y) > \mu_t + th \\ -1 & f_t(x,y) < \mu_t - th \\ 0 & \text{otherwise,} \end{cases}$$
(2)

The signal CMS(t) is in the range $-1 \leq CMS(t) \leq 1$. A cut occurs when the CMS(t) is near to -1, which means that a high number of pixels has a different behavior with respect to the mean in frame f_t and f_{t-1}. The cut detector is improved applying a set of morphological filters to the CMS signal, resulting a detector highly robust to the false alarms due to flicker. The whole process is presented in Albiol *et al.* (2000).

(2) Gradual transition detector which consists of the dissolve and wipe detectors presented in sections "Dissolve Detection" and "Wipe Detection", respectively.

PERFORMANCE EVALUATION

In order to evaluate the performance of our detectors, the recall and the precision will be obtained using the following expressions:

$$\text{Recall} = \frac{\text{Detections}}{\text{Detections} + \text{MD's}},$$

and

$$\text{Precision} = \frac{\text{Detections}}{\text{Detections} + \text{FA's}},$$

where "Detections" is the number of detected effects, "MD's" the number of missed detections and "FA's" that of false alarms.

Three different categories of video sequences have been used in order to test the detectors:

– Synthetic sequence: Is the linking, using dissolves of different lengths and wipes of different lengths and directions, of the QCIF sequences: Bridge close view, Bridge far view, Carphone, Claire, Container Ship, Foreman, Grandma, Highway drive, Mother and Daughter, Salesman and Silent, all of which are free available in
http://trace.eas.asu.edu/yuv/qcif.html.
This sequence has 10563 frames and is available in an AVI file
http://personales.upv.es/~vnaranjo/effectfilm.zip

– Real new sequences: Is a set of color sequences of high quality from current television movies and news:

· Sport sequences:

∗ cycling: a fragment from the cycle race around Spain with 1997 frames and 3 dissolves and 0 cuts.

∗ football: a fragment of 7060 frames of a football match with 14 cuts and 13 dissolves.

∗ basket: 6750 frames of a basket match with 15 cuts and 20 dissolves.

· News sequences:

∗ news_a: a piece of a news report (from TVE: Spanish Television) with 1907 frames and several transition effects: some cuts, 3 dissolves and 1 vertical wipe and also several sophisticated edition effects.

∗ news_b: similar sequence to the previous one with 1499 frames and only 1 dissolve.

∗ news_NBC: similar to the previous sequences with 13727 frames, 35 cuts, 19 dissolves and 16 wipes.

· Film sequences:

∗ movie: a fragment, 3010 frames, of the film *La sombra del ciprés es alargada* with many cuts but no transition effects.

∗ drama: a fragment, 3012 frames, of the Spanish TV series *Pepa y Pepe* also with many cuts but no transition effects.

∗ zorro: sequence from the film "The mask of Zorro", 5075 frames, 35 cuts and no gradual transition.

· Other sequences:

∗ cartoon: fragment of a cartoon TV series called *Don Quijote de la Mancha*, 8778 frames, 55 cuts and 29 dissolves.

∗ culture: a documentary about villages from Spain, 14896 frames, 75 cuts and 19 dissolves.

– Real old sequences: Is a set of degraded black and white sequences from several old films:

· malva: a fragment of 1789 frames from the Spanish film *Malvaloca* (1942).

· torbe: 2214 frames from the Spanish film *Torbellino* (1940)

We have analyzed all these sequences, near 70000 frames in all, obtaining the results shown in Tables 1–3, for the detection of cuts, dissolves and wipes respectively. In spite of the high number of processed frames, the number of gradual transition effects evaluated is not very high, due to the rare appearance of this kind of effects in real sequences, in comparison with the appearance of cuts.

Table 1. *Results of cut detection.*

Sequence	Detections	MD's	FA's
Synthetic	4	0	0
Cycling	0	0	0
Basket	23	7	0
Football	21	6	0
news_a	16	0	0
news_b	4	0	1
news_NBC	37	1	2
movie	14	0	0
drama	11	0	0
Zorro	87	5	1
Cartoon	52	3	2
Culture	75	0	0
malva	5	0	0
torbe	3	0	1
Total	352	22	7

Table 2. *Results of dissolve detection.*

Sequence	Detections	MD's	FA's
Synthetic	4	1	0
Cycling	3	0	0
Basket	20	3	2
Football	14	2	5
news_a	1	0	0
news_b	2	0	1
news_NBC 19	1	1	1
movie	0	0	0
drama	0	0	0
Zorro	0	0	0
Cartoon	26	3	5
Culture	19	2	2
malva	1	0	0
torbe	5	0	2
Total	96	12	18

Table 3. *Results of wipe detection.*

Sequence	Detections	MD's	FA's
Synthetic	3	0	0
Cycling	0	0	0
Basket	0	0	0
Football	0	0	0
news_a	2	1	1
news_b	0	0	0
news_NBC	24	0	5
movie	0	0	0
drama	0	0	0
Zorro	0	0	0
Cartoon	0	0	0
malva	1	0	0
torbe	3	0	1
Total	33	1	7

With the detector results presented in Tables 1 (for cut detection), 2 (for dissolve detection) and 3 (for wipe detection), the values of precision and recall are 93.2% and 93.7% respectively. Even in the case of old films, the detector does not obtain a high number of false alarms and misdetections. However, in old very degraded sequences a high number of false positives appears, which are associated to strong intensity degradation throughout several frames. A flicker correction step can be considered in order to improve the results (Naranjo and Albiol, 2000), achieving, after this flicker correction, similar values of recall and precision.

DISCUSSION

We have presented in this paper morphological techniques for detecting dissolves and wipes in video sequences. It is a necessary step, which combined with the detection of cuts, allows the temporal segmentation of sequences into shots. Experimental results have shown the satisfactory performance of our methodology with degraded sequences and still better on sequences in good condition. The developed low complexity algorithms yield to fast implementations and therefore can be adapted for (quasi) real-time applications.

From a methodological viewpoint, the present study illustrates how the morphological operators can be used to analyze time series for detecting particular non periodic events, either working directly on the 1D signal or building an intermediate 2D image from the 1D signals to take advantage of the spatial operators.

Regarding this subject, we can consider other applications of the "strip images", for instance, in video surveillance algorithms, to identify pedestrians (Fig. 10) or other events. As well as the orthogonal projections, we can use other "image parameters" for the non temporal axis of the "strip images", for instance, the luminance histogram to detect abrupt illumination variations, or color images combined with the saturation histogram to detect highlights and shadows, or skin color-centered hue histogram to detect people, *etc.*

REFERENCES

Angulo J (1999). Temporal segmentation of video sequences. Master Thesis, Universidad Politécnica de Valencia, October 1999.

Albiol A, Naranjo V, Angulo J (2000). Low complexity cut detection in the presence of flicker. Proceedings of IEEE International Conference on Image Processing (ICIP'00), Vol. III: 957–60.

Brunelli R, Mich O, Modena CM (1999). A survey on the automatic indexing of video data. J Vis Commun Image R 10:78–112.

Cotsaces C, Nikolaidis N, Pitas I (2006). Video shot detection. A review. IEEE Signal Proc Mag 23:28–37.

Demarty CH, Beucher S (1999). Morphological tools for video indexing. Proceedings of IEEE International Conference on Multimedia Computing and Systems (ICMCS'99), Vol. 2: 991–2.

Demarty CH (2000). Segmentation et Structuration d'un Document Vidéo pour la Caractérisation et l'Indexation de son Contenu Sémantique. Ph.D. Thesis, Centre de Morphologie Mathématique-Ecole des Mines de Paris, January 2000.

Falk M *et al.* (2006). A First Course on Time Series Analysis, by Chair of Statistics. University of Würzburg.

Fernando WAC, Canagarajah CN, Bull DR (1999). Fade and dissolve detection in uncompressed and compressed video sequences. Proceedings of IEEE International

Conference on Image Processing (ICIP'99), Vol. III: 299–303.

Gargi U, Oswald S, Kosiba DA, Devadiga S, Kasturi R (1995). Evaluation of video sequence indexing and hierarchical video indexing. In: Niblack W, Jain RC, eds. Proceedings of SPIE Conference on Storage and Retrieval for Image and Video Databases III, SPIE Vol. 2420, 144–51.

Joyce R A, Liu B (2006). Temporal segmentation of viedeo using frame and histogram space. IEEE Trans Multimedia 8(1):130–40.

Llach J, Salembier Ph (1999). Analysis of video sequences: Table of contents and index creation. Proceedings of International Workshop on Very Low Bitrate Video (VLBV'99).

Lu HB, Zhang YJ, Yao YR (1999). Robust Gradual Scene Change Detection. Proc. of IEEE International Conference on Image Processing (ICIP'99), Vol. III: 304–8 .

Meng J, Juan Y, Chang SF (1995). Scene Change Detection in a MPEG Compressed Video Sequence. In: Rodriguez AA, Safranek RJ, Delp EJ, eds. Proceedings of IST/SPIE Symposium, Vol. SPIE 2419, 14–25.

Naranjo V, Albiol A (2000). Flicker reduction in old films. Proceedings of International Conference of Image Processing 2000 (ICIP'00), 1300–3.

Naranjo V, Albiol An, Mossi JM, Albiol Al (2004). Morphological λ-reconstruction applied to restoration of blotches in old films. In: Villanieva JJ, ed.

Proceedings of the IASTED International Conference on Visualization, Imaging, and Image Processing (VIIP'04), 251–7.

Pardas M, Serra J, Torres L (1992). Connectivity filters for image sequences. In: Gader PD, Dougherty ER, Serra JC, eds. Proceedings of SPIE Symposium on Image Algebra and Morphological Image Processing III, SPIE Vol. 1769:318–29.

Serra J (1982). Image Analysis and Mathematical Morphology, Vol I, Image Analysis and Mathematical Morphology. London: Academic Press.

Serra J (1988). Image Analysis and Mathematical Morphology, Vol II: Theoretical Advances. London: Academic Press.

Soille P (1999). Morphological image analysis. Berlin, Heidelberg: Springer-Verlag.

Truong B T, Dorai C, Venkatesh S (2000). Improved fade and dissolve detection for reliable video segmentation. Proceedings of IEEE International Conference on Image Processing (ICIP'00), Vol III: 961–4.

Vincent L (1993). Morphological Grayscale Reconstruction in Image Analysis: Applications and Efficient Algorithms. IEEE Trans Image Process 2(2):176–201.

Yeo B, Liu B (1995). Rapid Scene Analysis on Compressed Video. IEEE Trans Circ Syst Vid 5(6):533–44.

Yoo HW, Ryoo HJ, Jang DS (2006). Gradual shot boundary dtection using localized edge blocks. Multimed Tools Appl 28:283–300.