EURASIP Journal on Applied Signal Processing 2004:12, 1943–1953

Flat Zone Analysis and a Sharpening Operation for Gradual Transition Detection on Video Images

Silvio J. F. Guimarães

Laboratoire Algorithmique et Architecture des Systèmes Informatiques, École Supérieure d'Ingénieurs en Électronique et Électrotechnique, 93162 Noisy Le Grand Cedex, Paris, France

Institute of Computing, Pontifical Catholic University of Minas Gerais, 31980-110 Belo Horizonte, MG, Brazil
Email: sjamil@pucminas.br

Neucimar J. Leite

Institute of Computing, State University of Campinas, 13084-971 Campinas, SP, Brazil
Email: neucimar@ic.unicamp.br

Michel Couprie

Laboratoire Algorithmique et Architecture des Systèmes Informatiques, École Supérieure d'Ingénieurs en Électronique et Électrotechnique, 93162 Noisy Le Grand Cedex, Paris, France
Email: coupriem@esiee.fr

Arnaldo de A. Araújo

Computer Science Department, Universidade Federal de Minas Gerais, 6627 Pampulha, Belo Horizonte, MG, Brazil
Email: arnaldo@dcc.ufmg.br

Received 1 September 2003; Revised 28 June 2004

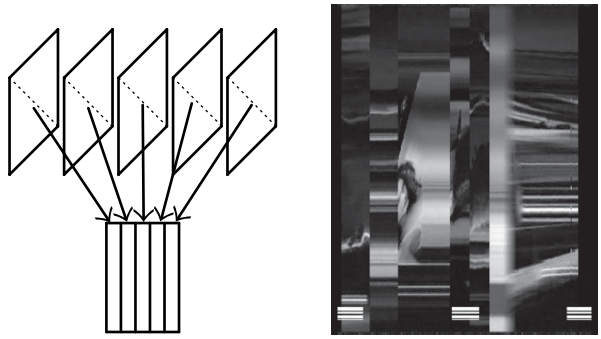
The boundary identification represents an interesting and difficult problem in image processing, mainly if two flat zones are separated by a gradual transition. The most common edge detection operators work properly for sharp edges, but can fail considerably for gradual transitions. In this work, we propose a method to eliminate gradual transitions, which preserves the number of the image flat zones. As an application example, we show that our method can be used to identify very common gradual video transitions such as fades and dissolves.

Keywords and phrases: flat zone analysis, video transition identification, visual rhythm.

1. INTRODUCTION

The boundary identification represents an interesting and difficult problem in image processing mainly if two flat zones, defined as the sets of adjacent points with the same gray-scale value, are separated by a gradual transition. The most common edge detection operators like Sobel and Roberts [1] work well for sharp edges but fail considerably for gradual transitions. These transitions can be detected, for example, by a statistical approach proposed by Canny [2]. Another approach to cope with this problem is through mathematical morphology operators which include the notion of thick gradient and multiscale morphological gradient [3]. From this approach, and depending on the size of the transition and its neighboring flat zones, the gradual transitions cannot be well detected. In this work, we consider the problem of detecting gradual transitions on images by a sharpening process which does not change their original number of flat zones.

As an application example, we consider the problem of identifying gradual transitions such as fade and dissolve on digital videos. Usually, the common approach to this problem is based on dissimilarity measures used to identify the gradual transitions between consecutive shots [4]. In literature, we can find different types of dissimilarity measures used for video segmentation, such as pixel-wise and histogram-wise comparison. If two frames belong to the same shot, then their dissimilarity measure should be small.



(a) (b)

Figure 1: Video transformation: (a) simplification of the video content by transformation of each frame into a column on the visual rhythm representation and (b) a real example considering the principal diagonal subsampling.

Two frames belonging to different shots generally yield a high dissimilarity measure whose value can be significantly affected by the presence of gradual transitions in the shot. In the same way, a dissimilarity measure concerning the frames of a gradual transition is difficult to define and the quality of this measure is very important for the whole segmentation process. Some works on gradual transitions detection can be found in [5, 6, 7, 8, 9]. Zabih et al. [5] proposed a method based on edge detection which is very costly due to the computation of edges for each frame of the sequence. Fernando et al. [6] and Lienhart [7] used a statistical approach that considers features of the luminance signal. This approach presents high precision on long fades. Zhang et al. [8] introduced the twin-comparison method in which two different thresholds are considered. Yeo [9] introduced the plateau method where the computation of the dissimilarity measure depends on the duration of the transition to be detected.

An interesting approach to deal with the problem of identifying gradual transitions is to transform the video images into a 2D image representation, named visual rhythm (VR), and apply image processing tools for detecting patterns corresponding to different video events in this simplified representation. As we will see elsewhere, each frame of the video is transformed into a vertical line of the VR, as illustrated in Figure 1a. This method of video representation and analysis can be found in [10, 11, 12, 13]. In [10], Chung et al. applied statistical measures to detect patterns on the VR with a considerable number of false detections. In [11], Ngo et al. applied Markov models for shot transition detection which fails in the presence of low contrast between textures of consecutive shots. In [12], we proposed a method to identify cuts based on the VR representation and on morphological image operators. In [13], we considered the problem of identifying fades based on a VR by histogram.

This work is an extension of a previous one [14] which introduces the problem of detecting patterns on a VR image by eliminating gradual transitions according to a homotopic sharpening process. Here, we explain in detail some features of the proposed method and illustrate its application and results on a set of video images by taking into account different experiments and variants of the method.

This paper is organized as follows. In Section 2, we give some concepts on digital video and define the visual rhythm transformation. In Section 3, we introduce the approach for transforming gradual into sharp transitions represented by a 1D signal. In Section 4, we consider the problem of identifying fades and dissolves from this signal. In Section 5, we make some comments on the realized experiments. Finally, some conclusions and suggestions of future works are given in Section 6.

2. VIDEO TRANSFORMATION

Let $\mathbb{A} \subset \mathbb{Z}^2$, $\mathbb{A} = \{0,\ldots,\mathrm{H}-1\} \times \{0,\ldots,\mathrm{W}-1\}$, be our application domain, where H and W are the height and the width of each frame, respectively.

Definition 1 (frame). A frame f_t is a function from $\mathbb{A}$ to $\mathbb{Z}$, where for each spatial position (x, y) in $\mathbb{A}$, $f_t(x, y)$ represents the gray-scale value at pixel location (x, y).

Definition 2 (video). A video V, in domain $2\mathbb{D} \times t$, can be seen as a sequence of frames f_t. It can be described by

$$V = (f_t)_{t\in[0,\mathrm{duration}-1]}, \tag{1}$$

where *duration* is the number of frames in the video. In this work, we consider video transitions such as cut, fade, and dissolve. Cut is an event which concatenates two consecutive shots. According to [15], the fade transition is characterized by a progressive darkening of a shot until the last frame becomes completely black (fade-out), or the opposite, allowing the gradual transition from black to light (fade-in). A more general definition of fade is given in [7] where the black frame is replaced by a monochrome frame. This event can be subdivided into fade-ins and fade-outs. Unlike cut, the dissolve transition is characterized by a progressive transformation of a shot P into another shot Q. Usually, it can be seen as a generalization of fade in which the monochrome frame is replaced by the first or last frame of the shot. Figure 2 illustrates these different types of events.

2.1. Visual rhythm

The detection of events on digital videos is related to basic problems concerning, for instance, processing time and choice of a dissimilarity measure. Aiming at reducing the processing time and using 2D image segmentation tools instead of dissimilarity measures only, we consider the following simplification of the video content [10, 11].

Definition 3 (VR). Let $V = (f_t)_{t\in[0,\mathrm{duration}-1]}$ be an arbitrary video, in domain $2\mathbb{D} \times t$. The visual rhythm VR, in domain $1\mathbb{D} \times t$, is a simplification of the video where each frame f_t is transformed into a vertical line on the VR:

$$\mathrm{VR}(t, z) = f_t(r_x * z + a, r_y * z + b), \tag{2}$$

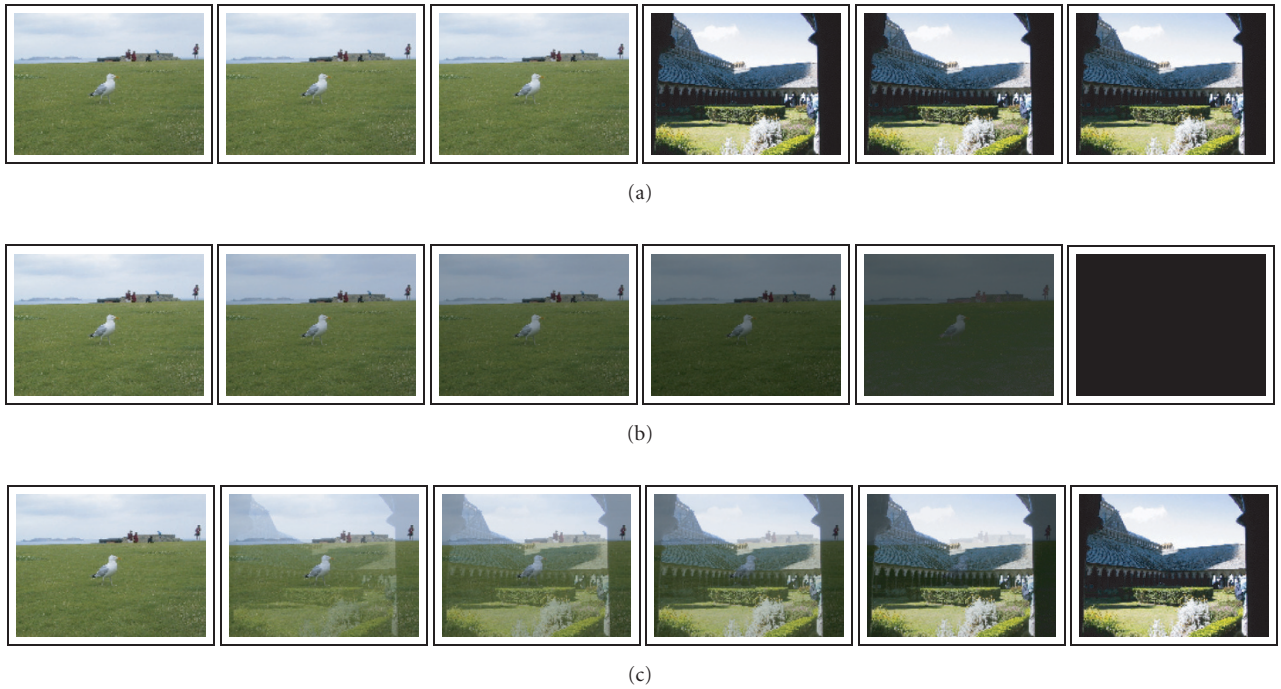


FIGURE 2: Example of cut and gradual transitions: (a) cut, (b) fade-out, and (c) dissolve.

where $z \in [0, \ldots, \mathrm{H_{VR}} - 1]$ and $t \in [0, \ldots, \mathrm{duration} - 1]$, $\mathrm{H_{VR}}$ and duration are the height and the width of the VR, respectively, r_x and r_y are ratios of pixel sampling, and a and b are shifts on each frame. Thus, according to these parameters, different pixel samplings can be considered. For instance, if $r_x = r_y = 1$, $a = b = 0$, and H = W, then we define all pixels of the principal diagonal as samples of the VR.

The choice of the pixel sampling is an interesting problem because different samplings can yield different VRs with different patterns. In [10], the authors analyze some pixel samplings, together with their corresponding VR patterns, and state that the best results are obtained by considering diagonal sampling of the images since it encompasses horizontal and vertical features. In Figure 3, we give some examples of patterns based on the principal diagonal pixel sampling. According to the defined features, we have that all cuts are represented by vertical sharp lines while the gradual transitions are represented by vertical aligned gradual regions. All these features are independent of the type of the frame sampling. Figure 3a illustrates the cut transition. Figures 3b and 3c give examples of fade, and Figures 3d and 3e show some dissolve patterns.

3. SHARPENING BY FLAT ZONE ENLARGEMENT

In a general way, the existence of gradual transitions in an image yields a more difficult problem of edge detection which can be approached, for example, by multiscale and sharpening operations [3]. While the multiscale operations consider gradual regions as edges of different sizes identified at different scales, the sharpening methods try to detect edges by eliminating (or reducing) gradual transition regions. The multiscale operations need the definition of a maximum scale during the processing since the transition detection is associated with this scale parameter.

This work concerns the definition of a sharpening method to identify gradual transitions on video images. As we will see next, we try to transform these transitions, related to events such as fades and dissolves, into sharp regions based on some 1D operations that enlarge the components of the VR image. It is important to remark that the sharp vertical lines representing cuts in the VR will not be modified by this transformation.

Next, we introduce some basic concepts considered in this paper. Let g be a 1D signal represented by a function of $\mathbb{N} \to \mathbb{N}$. We denote by $N(p)$ the set of neighbors of a point p. In such a case, $\mathrm{N}(p) = \{p - 1, p + 1\}$ represents the right and left neighbors of p.

Definition 4 (flat zone, k-flat zone and k^+-flat zone). A flat zone of g is a maximal set (in the sense of inclusion) of adjacent points with the same value. A k-flat zone is a flat zone of size equal to k. A k^+-flat zone is a flat zone of size greater than or equal to k.

Definition 5 (transition). We denote by F the set of k^+-flat zones of g. A transition T between two k^+-flat zones, F_i and F_j, is the range $[p_0 \ \cdots \ p_{n-1}]$ such that $p_0 \in F_i$, $p_{n-1} \in F_j$, for $0 < m < n - 1$, $p_m \notin F_i \bigcup F_j$, for all $l \neq i, j$ $F_l \not\subset [p_0 \ \cdots \ p_{n-1}]$ and for $0 \leq i < n - 1$, $g(p_i) \leq g(p_{i+1})$ (or $g(p_i) \geq g(p_{i+1})$).

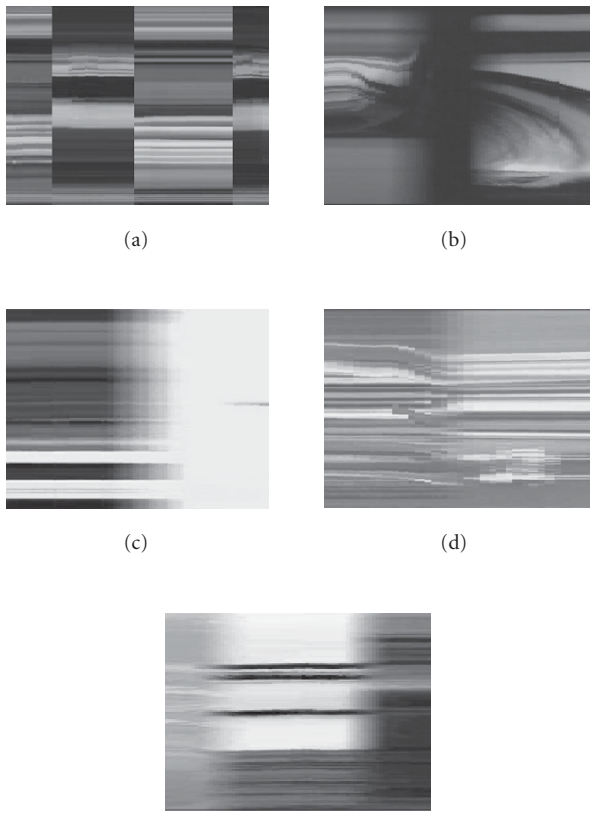


Figure 3: Example of patterns on the visual rhythm associated with cut and gradual transitions: (a) 3 cuts, (b) 1 fade-out followed by 1 fade-in, (c) 1 fade-out, (d) 1 dissolve, and (e) 2 consecutive dissolves.

Figure 4 shows examples of flat zones and transitions. In this work, the analysis of the transition regions is related to the identification and elimination of the neighboring points of these transitions while preserving the number of k^+-flat zones. Next, we define two different types of transition points, namely, constructible and destructible points, as illustrated in Figure 5.

Let $D(p, F)$ be the difference between the gray-scale value of a point p and the value of a flat zone F.

Definition 6 (constructible or destructible transition point). We denote by T the transition between two k^+-flat zones, F_i and F_j. Let $p \in T$, $p-1$, and $p+1$, be a pixel of a 1D signal, g, and its neighbors, respectively. A point p is a constructible transition point if and only if $g(p) \geq \min(g(p-1), g(p+1))$, $g(p) \leq \max(g(p-1), g(p+1))$, and $D(p, F^-) > D(p, F^+)$. A point p is a destructible transition point if and only if $g(p) \geq \min(g(p-1), g(p+1))$, $g(p) \leq \max(g(p-1), g(p+1))$, and $D(p, F^-) < D(p, F^+)$, where F^- and F^+ denote lowest and the highest gray-scale flat zones nearest to p and, $D(p, F^-)$ and $D(p, F^+)$ are the difference of gray-scale values between p and the respective flat zones.

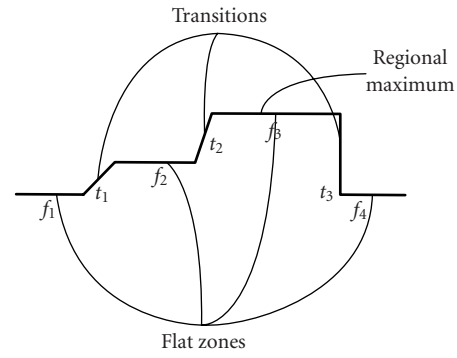


Figure 4: Example of flat zones and transitions.

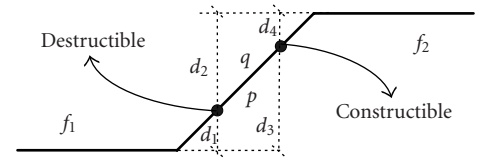


Figure 5: Constructible and destructible points in a transition region.

In Figure 5, we illustrate the identification of constructible and destructible points. In such a case, p is a destructible ($d_1 < d_2$) and q is a constructible point ($d_4 < d_3$). The aim here is to define a homotopic operation which simplifies the image without changing the number of its k^+-flat zones. In other words, we want to change gray-scale values representing transition points in the neighborhood of k^+-flat zones, without suppressing or creating new flat zones. As we will see next, the definition of the sequence of points to be evaluated in the sharpening process is an important aspect to be considered since different sequences can yield different results. Algorithm 1 is used to eliminate gradual transitions of an image by enlarging its original flat zones.

Informally, step (1) identifies all k^+-flat zones of the input VR image. A morphological filtering operation (e.g., a closing followed by an opening with a linear and symmetric structuring element taking into account the minimum duration of a shot) may be considered to reduce small irrelevant flat zones of the original image. We empirically set $k = 7$ as the minimum duration of a shot. For each k^+-flat zone, in step (2), set C represents the neighboring points of the corresponding flat zone.

Steps (3)–(7) deal with the constructible and destructible points related to the transition regions. As stated before, an interesting aspect of these steps concerns the removal of a point from set C which, depending on its removing order, can yield different results. For the purpose of this removal control, we use a hierarchical priority queue to maintain an equidistant spatial relation between the removed points and their neighboring flat zones. To this end,

Input: Visual rhythm (VR) image, size parameter k
Output: Sharpened visual rhythm (VR^e).

For each line L of VR do

For all flat zones of L with size greater than or equal to k do

insert(C, $\{q \mid \exists p \in k^+$-flat zones, $q \in N(p)$, and $q \notin k^+$-flat zones $\}$)

While $C \neq \emptyset$ do

p = extractHighestPriority (C)

q = point in N (p) not yet modified by the sharpening process

$VR^e(L, p)$ = gray scale of p nearest neighboring flat zone

insert (C, q)

Algorithm 1: Algorithm for sharpening by enlarging flat zones.

we define two functions, extractHighestPriority(C) and insert(C, q), which remove a point of highest priority and insert a new point q into set C, according to a predefined priority criterium. A currently removed point presents the highest priority in this queue, where the priority depends on the criterium used to insert new points in this data structure. The gray-scale difference between a k^+-flat zone and its neighboring points is used here as an insertion criterium.

Figure 6 illustrates the data structure representing the set C considered in the sharpening process. In Figure 6a, the k^+-flat zones are represented by letters f and g while the transition points are indicated by a, b, c, and d. Figure 6b shows the first configuration of set C (step (2) of the algorithm), in which points a and d are inserted with the 1 priority corresponding to the gray-scale differences with respect to their nearest k^+-flat zones, f and g, respectively. In Figures 6c and 6d, we illustrate the results of steps (6) and (7) of the algorithm, applied to set C and represented by the priority queue illustrated in Figure 6b. In Figure 6e, we illustrate the results of steps (6) and (7) represented by the new defined queue shown in Figure 6d where the priority of points b and c equals 2. From this example, we have that flat zones f and g were enlarged yielding an elimination of the corresponding gradual transitions between them. This transformation defines a sharpened version of the original signal. Figure 7 gives some examples of the flat zone enlargement (or sharpening) method applied to each line of the original VR representation.

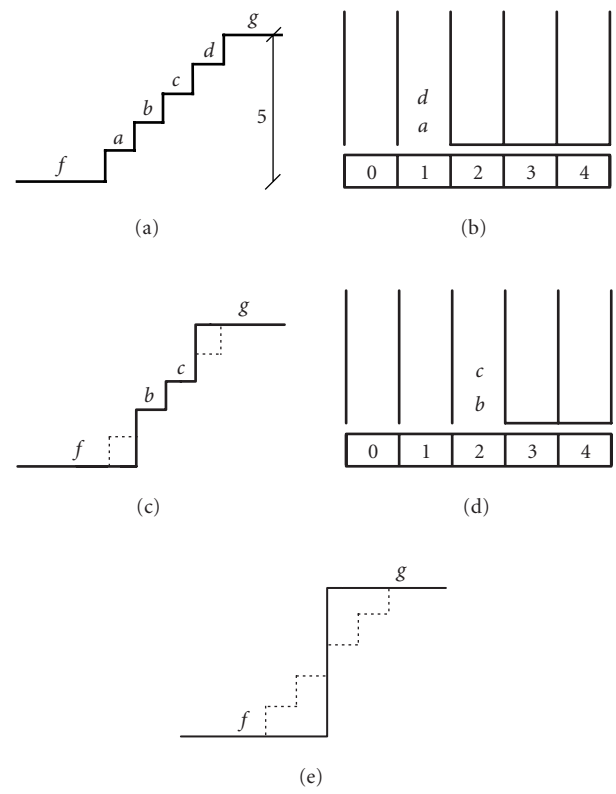


Figure 6: Enlargement of flat zones using a priority queue: (a) original image, (b) initial configuration of the priority queue according to the signal in (a), (c) sharpening after extracting 1 priority points from the priority queue, (d) new configuration of the priority queue according to the signal in (c), and (e) result of the sharpening process.

4. TRANSITION DETECTION

The video segmentation problem is very difficult to consider in the presence of gradual transitions, mainly, in case of dissolves. As described in [11], the gradual transitions are represented by vertically aligned gradual regions in the VR. In Figure 8a, we illustrate a VR of a video containing 4 cuts, 2 fades, and 1 dissolve. In Figure 8b we show the result of our sharpening method applied to the VR image illustrated in Figure 8a. Figures 8c and 8d correspond, respectively, to the line profiles of the center horizontal rows in Figures 8a and 8b. In case of gradual transitions, all lines of the VR present a common feature in a specific range of time, that is, a gray-scale increasing or decreasing regarding the temporal axis.

To detect these gradual transitions, we can simplify the VR by considering the sharpening transformation described in Section 3. As stated before, this transformation preserves the original number of shots in a video sequence since it does not change the number of k^+-flat zones representing them. To reduce noise effects, we can also apply an alternated morphological filter [16, 17] with a linear structuring element of size closely related to the smallest duration of a shot (7, in our case). Further, we consider the following aspects of a gradual transition.

(1) In a gradual transition region, the number of points modified by the sharpening process is high. If the transformation function of the event is linear and the consecutive frames are different from each other, then the number of points in the sharpened visual rhythm (VR^e) modified by the sharpening process equals the

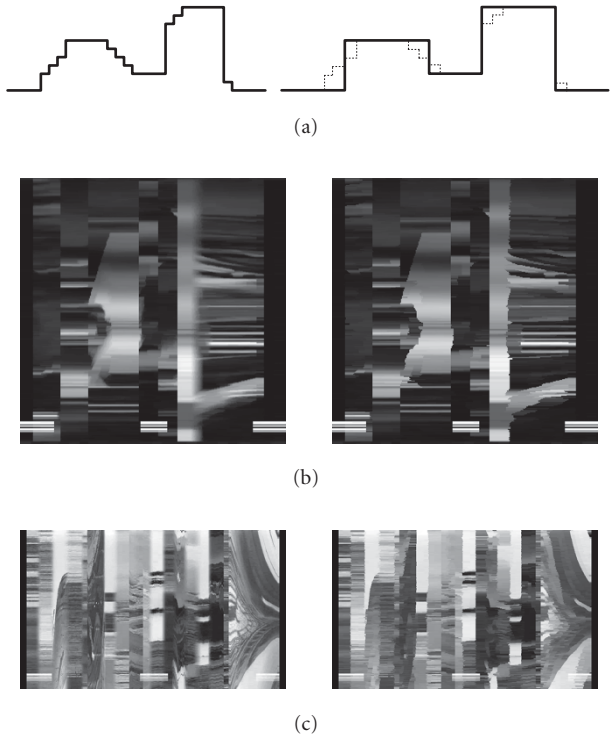


Figure 7: Example of flat zones enlargement: (a) artificial original signal (left) and its sharpened version (right), (b) and (c) original visual rhythms (left) and their corresponding sharpened versions (right).

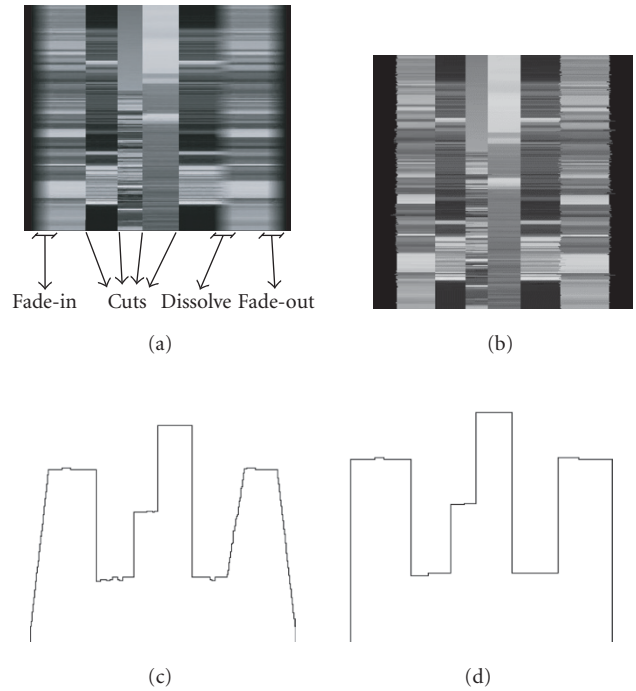


Figure 8: Example of a sharpened image: (a) VR with some events, (b) image obtained after the proposed sharpening process, (c) and (d) the respective line profiles of the center horizontal rows of the images.

height of the original VR. Unfortunately, in real cases, this number can be affected, for example, by the presence of noise and digitization problems.

(2) As we will see next, the regions of gradual transitions will be represented by a specific 1D configuration. Again, if the transformation function of the transition is linear, then the points modified by the sharpening process define a regional maximum corresponding to the center of the transition and given by the highest gray-scale value of the difference between images VR and VR^e.

Now, if we consider both images VR and VR^e, the basic idea of our gradual transition detection method consists in analyzing the VR image by taking into account the number and the gray-scale values of its modified pixels (points of the gradual transitions) in the sharpened version VR^e. Figure 9 summarizes the following steps of the transition detection algorithm.

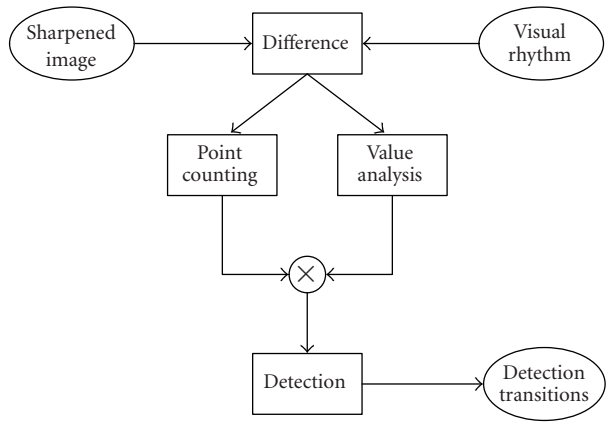


Figure 9: Main steps of the proposed gradual transition detection algorithm for video images.

Difference

This step computes the difference between images VR and VR^e, defining a new image Dif as follows

$$\mathrm{Dif}(x, y) = \left| \mathrm{VR}(x, y) - \mathrm{VR}^e(x, y) \right|. \quad (3)$$

Point counting

This step takes into account the points modified by the sharpening process by counting the number of nonzero values in each column of image Dif. To reduce noise and fast motion influence, we consider a morphological opening with a vertical structuring element of size 3 before the counting

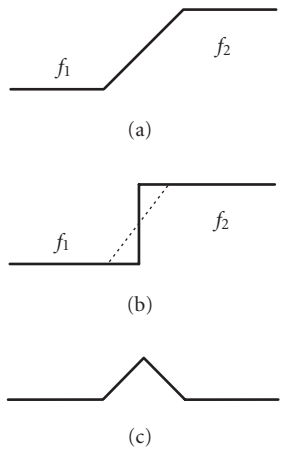


Figure 10: Example of flat zones enlargement: (a) original image, (b) sharpened image, and (c) difference image.

process given by

$$M^p(p) = \sum_{j=0}^{H_{VR}-1} \begin{cases} 1 & \text{if } \mathrm{Dif}(p,j) > 0, \\ 0 & \text{otherwise,} \end{cases} \tag{4}$$

where H_{VR} is the height of the VR image and $p \in [0, \ldots, \text{duration} - 1]$.

Value analysis

This step computes the gray-scale mean of the points modified by the sharpening process. As illustrated in Figure 10, gradual transitions are represented by single domes (Figure 10c) in each row of image Dif, the center of the transitions corresponding to the regional maximum of these domes. Usually, the first and last frames of a gradual transition correspond to the smallest values of these domes. In case of a monotonic transition, we have that the 1D signal increases between the first and the center frames of the event, decreasing from the center of the defined dome until the last transition frames. Furthermore, the duration of each half of the dome is the same if the transformation function of the gradual transition is linear. Before analyzing the domes configuration in image Dif, we compute the mean values in each column of this image, defining a 1D signal, M^v, as follows:

$$M^v(p) = \frac{\sum_{y=0}^{H_{VR}-1} (\mathrm{Dif}(p,y))}{H_{VR}}. \tag{5}$$

To identify a dome configuration (Figure 10c), we decompose the M^v signal into morphological residues by means of granulometric transformations [16, 18, 19]. This multiscale representation of a signal is used here to detect the residues, at a certain granulometric level, associated with the dome configuration of a gradual transition. These residues are defined as follows.

Definition 7 (gray-scale morphological residues [19]). Let $(\psi_i)_{i\geq 0}$ be a granulometry. The gray-scale morphological residues (or simply, morphological residues), R_i, of residual level i are given by the difference between the result of two consecutive granulometric levels, that is,

$$\forall i \geq 1,\ f \in \mathbb{Z}^n, \quad R_i(f) = \psi_{i-1}(f) - \psi_i(f), \tag{6}$$

where f represents gray-scale digital images. The morphological residues represent the components preserved at level $(i-1)$ and eliminated at the granulometric level i. The morphological residues depend on the used structuring element whose parameter i corresponds to its radius (a linear structuring element of radius i has length $(2 \times i) + 1$).

As an illustration of this analysis, we consider two different levels, Inf and Sup. Based on these parameters, we can define the number of residual levels containing a point p as follows:

$$M_{Inf}^{Sup}(p) = \sum_{i=Inf}^{Sup} \begin{cases} 1 & \text{if } R_i(M^v(p)) > 0, \\ 0 & \text{otherwise,} \end{cases} \tag{7}$$

where R_i means the morphological residue at level i (6). A point p corresponding to a regional maximum in M^v represents a candidate frame for gradual transition if $M_{Inf}^{Sup}(p)$ is greater than a threshold l_1. The set of these candidate frames along a video sequence is given by

$$C_{Inf}^{Sup}(p) = \begin{cases} 1 & \text{if } M_{Inf}^{Sup}(p) > l_1, \\ 0 & \text{otherwise.} \end{cases} \tag{8}$$

In this work, the values Inf, Sup, and l_1 were empirically defined as 3, 15, and 3, respectively. The choice of these values is related to the features of the gradual transitions to be detected. For instance, Inf $= 3$ was defined based on the minimum duration of a transition (11 frames on average according to our video corpus) and the maximal number of empty residual levels represented by l_1. Thus, the Inf value corresponds to the radius of the linear used structuring element whose size parameter equals 7 $(2 \times 3 + 1 = 7)$. The value of l_1 concerns the number of odd values between the lowest size parameter (7, in this case), and the minimum duration of a transition (11 frames). If we decrease l_1, the number of missed candidate frames can increase, for example, in cases where the dome configuration is affected by motion and noise. Finally, the parameter Sup concerns the duration of the longest considered gradual transition $(2 \times 15 + 1 = 31$ frames). Note that the configuration of each dome is very important if we want to identify gradual transitions but it does not represent a sufficient criterium. We also need to take into account, for each candidate frame, the number of points modified by the sharpening process as explained next.

Detection operation

This last step of the algorithm combines the information obtained from the point counting and the value analysis steps previously defined. By considering a gradual transition as a specific dome configuration in M^v, represented by candidate

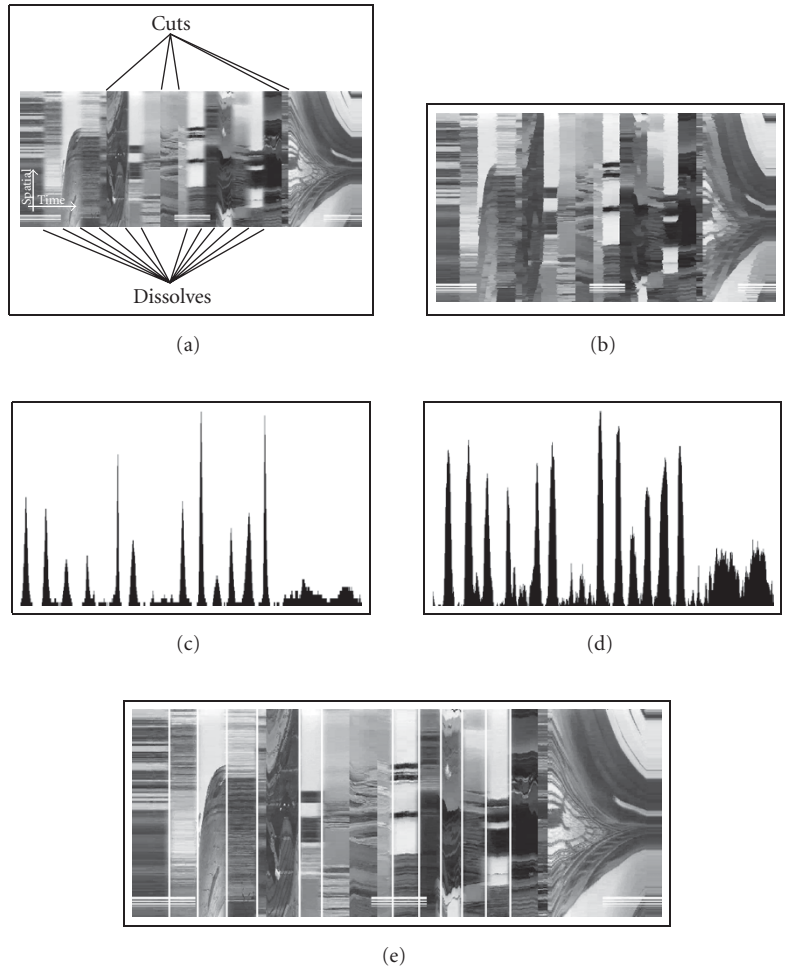


FIGURE 11: Gradual transition detection: (a) original image containing 12 dissolves and 5 cuts, (b) sharpened image, (c) M^v signal, (d) number of modified points, and (e) result of the method without false detection.

frames with a high number of modified points in the sharpening process, we can combine the above steps as follows:

$$M_p^v(p) = \begin{cases} M^p(p) & \text{if } C_{\text{Inf}}^{\text{Sup}}(p) > 0, \\ 0 & \text{otherwise.} \end{cases} \tag{9}$$

This equation takes into account candidate frames p and the corresponding number of values in each column of the VR image modified by the sharpening process. Finally, we can detect a gradual transition at location p through the simple thresholding operation

$$T(p) = \begin{cases} 1 & \text{if } M_p^v(p) > l_2, \\ 0 & \text{otherwise,} \end{cases} \tag{10}$$

where l_2 is a threshold value. Figure 11 illustrates our gradual transition detection method. In this example, we process each horizontal line of the original VR (Figure 11a) containing, among other events, 12 dissolves and 5 cuts. The sharpened version of this image is shown in Figure 11b. The result in Figure 11e (the white vertical bars indicate the detected events) was obtained by defining l_2 as 25% of the maximal value of M_p^v. The relation with this maximal value is important to make the parameter independent from different types of videos (e.g., commercial, movie, and sport videos). Notice that all sharp vertical lines representing cuts in Figure 11a were not detected here.

To evaluate the proposed method, we considered the set of four experiments described next.

5. EXPERIMENTAL ANALYSIS

In this section, we discuss the experimental results concerning the detection of gradual transitions on video images. The choice of the digital videos was guided by the presence of events, such as cut, dissolve, and fades on the sequences. In all experiments, we used 28 commercial video images containing 77 gradual transitions (involving fades and dissolves). To compare the different results, we defined some quality measures [12] demanding a manual identification of the considered events. We denote by Events the number of all events

Table 1: Results of our experiments.

Exp	Gradual	Detected	False	Recall	Precision	Error	Threshold
1	77	75	46	97.5%	62%	60%	2%
2	77	75	52	97.5%	59%	67%	25%
3	77	72	10	93.5%	88%	13%	25%
4	77	57	53	74%	52%	68%	0.5 and 0.1

in the video, by Corrects the number of properly detected events, and by Falses the number of detected frames that do not represent a correct event. Based on these values, we consider the following quality measures.

Definition 8 (recall, precision, and error rates). The recall and error rates represent the ratios of correct and false detections, respectively, and the precision value relates correct to false detections. These measures are given by

$$\alpha = \frac{\text{Corrects}}{\text{Events}} \quad \text{(recall)},$$
$$\beta = \frac{\text{Falses}}{\text{Events}} \quad \text{(error)}, \tag{11}$$
$$\mathcal{P} = \frac{\text{Corrects}}{\text{Falses} + \text{Corrects}} \quad \text{(precision)}.$$

Since we are interested in gradual transitions, Events is related to the gradual transitions satisfying the basic hypothesis in which the number of gradual transition frames is greater than 10. The tests realized in this work concern the following experiments.

Experiment 1. This experiment considers only the gray-scale values of the difference image M^v. In such a case, a transition p is detected if the $M^v(p)$ value is greater than a given threshold T. This value, associated with the M^v regional maximum, was empirically defined as 2% of the maximal possible value (255).

Experiment 2. This experiment takes into account the number of modified points by the sharpening process. If $M^p(p)$ is greater than a given threshold, then the point p represents a transition frame. This analysis is based on the regional maxima of the 1D signal M^v. The threshold value corresponds here to 25% of the VR height.

Experiment 3. This experiment corresponds to our proposed method (Section 3).

Experiment 4. This experiment considers the twin-comparison approach [8] which detects gradual transitions based on histogram information. Two thresholds, T_b and T_s, are defined reflecting the dissimilarity measures of frames between two shots and frames in different shots, respectively. If a dissimilarity measure, $d(i, i+1)$, between two consecutive frames satisfies $T_b < d(i, i+1) < T_s$, then candidate frames representing the start of gradual transitions are detected. For each candidate frame, an accumulated comparison $A(i) = \sum d(i, i+1)$ is computed if $A(i) > T_b$ and $d(i, i+1) < T_s$, and the end frame of a gradual transition is determined when $A(i) > T_s$. Here, we consider $T_b = 0.1$ and $T_s = 0.5$ and since cuts are not considered, a transition is detected only if the video frames are classified as candidates.

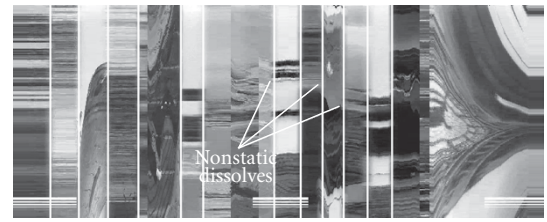


Figure 12: Nonstatic and static gradual transition detection. The white bars indicate the detected transitions (3 nonstatic and 9 static gradual events).

5.1. Analysis of the results

According to Table 1, we can observe that the proposed method (Experiment 3) yields better results when compared to the other experiments. If we take into account only gray-scale values (Experiment 1), the transitions are well identified due to their specific configurations, but this method is very sensitive to differences between two consecutive shots. By considering the modified points only (Experiment 2), some transition frames can be confused with special events and fast motions. Indeed, this method is more sensitive to noise and fast motion. The above features explain why we take into account both the gray scale and the modified point information in Experiment 3 which performs better than the twin-comparison method (Experiment 4) as well.

Some false detections of our approach are due to the identification of transitions whose duration is smaller than 11 frames. These transitions are probably defined by the presence of noise in the VR representation. In case of nonstatic gradual events, their sharpened version is not completely vertically aligned, and the number of modified points may be smaller than the one obtained for static gradual transitions. Due to this some missed detections may have occurred.

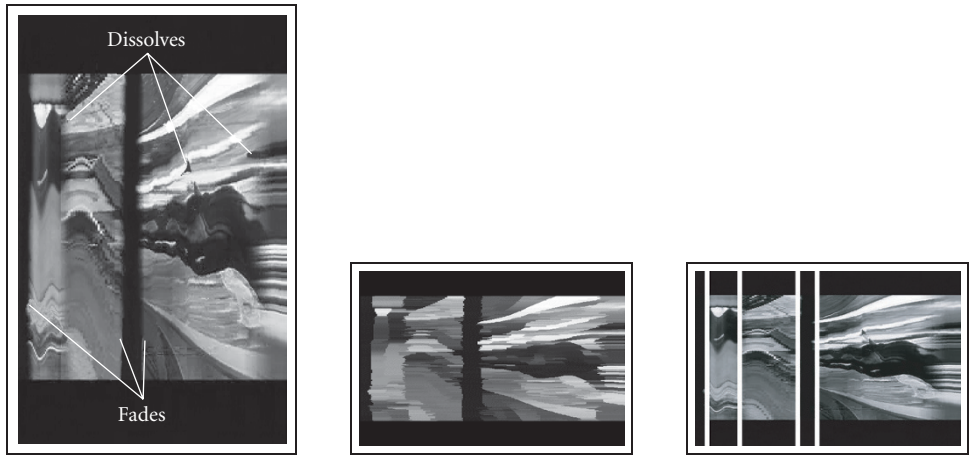


Figure 13: Example of a real video in which 2 dissolves are not detected: (a) visual rhythm which contains 3 dissolves and 2 fades, (b) sharpened visual rhythm, and (c) the detected transitions identified by vertical white bars.

Figure 12 shows an example in which all nonstatic transitions are identified (3 dissolves). Figure 13 shows a VR containing 3 dissolves and 3 fades. This figure illustrates the occurrence of missed detections (2 dissolves) represented mainly by cases in which a gradual transition is combined with other video effects like a zoom in.

Finally, it is important to note that all parameters related to Experiment 3 were defined based on the inherent characteristics of the transitions to be detected.

6. CONCLUSIONS

In this work, we defined a new method for transforming smooth transitions into sharp ones and illustrated its application in the detection of gradual events on video images. The sharpening operator defined here is based on the classification of pixels in the gradual transition regions as constructible or destructible points. This operator constitutes the first step for detecting two very common video events known as dissolve and fade. One of the main features of our approach is that it does not depend on the transition duration, that is, dissolve and fade events with different transition times can be properly recognized. Furthermore, the computational cost of the proposed method, based on the VR representation, is lower when compared to other approaches taking into account all video information. A drawback here concerns the sensitivity to motion which can be avoided through a preprocessing for motion compensation. An interesting extension to this work concerns the analysis of the efficiency of the method, when applied to all video content, and the improvement of the obtained results for nonstatic transitions. Also, the choice of thresholds must be exploited.

ACKNOWLEDGMENTS

The authors are grateful to CNPq, CAPES/COFECUB, the SIAM DCC, and the SAE IC PRONEX projects for the financial support of this work. This work was also partially supported by research funding from the Brazilian National Program in Informatics (decree-law 3800/01).

REFERENCES

[1] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, Prentice Hall, Upper Saddle River, NJ, USA, 2nd edition, 2002.

[2] J. Canny, "A computational approach to edge detection," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 8, no. 6, pp. 679–698, 1986.

[3] P. Soille, *Morphological Image Analysis: Principles and Applications*, Springer-Verlag, Berlin, Germany, 1999.

[4] A. Hampapur, R. Jain, and T. E. Weymouth, "Production model based digital video segmentation," *Multimedia Tools and Applications*, vol. 1, no. 1, pp. 9–46, 1995.

[5] R. Zabih, J. Miller, and K. Mai, "A feature-based algorithm for detecting and classifying production effects," *Multimedia Systems*, vol. 7, no. 2, pp. 119–128, 1999.

[6] W. A. C. Fernando, C. N. Canagarajah, and D. R. Bull, "Fade and dissolve detection in uncompressed and compressed video sequences," in *Proc. IEEE International Conference on Image Processing (ICIP '99)*, vol. 3, pp. 299–303, Kobe, Japan, October 1999.

[7] R. Lienhart, "Comparison of automatic shout boundary detection algorithms," in *SPIE Image and Video Processing VII*, vol. 3656, pp. 290–301, San Jose, Calif, USA, January 1999.

[8] H. Zhang, A. Kankanhalli, and S. Smoliar, "Automatic partitioning of full-motion video," *Multimedia Systems*, vol. 1, no. 1, pp. 10–28, 1993.

[9] B.-L. Yeo, *Efficient processing of compressed images and video*, Ph.D. thesis, Department of Electrical Engineering, Princeton University, Princeton, NJ, USA, January 1996.

[10] M. G. Chung, J. Lee, H. Kim, S. M.-H. Song, and W. M. Kim, "Automatic video segmentation based on spatio-temporal features," *Korea Telecom Journal*, vol. 4, no. 1, pp. 4–14, 1999.

[11] C. W. Ngo, T. C. Pong, and R. T. Chin, "Detection of gradual transitions through temporal slice analysis," in *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '99)*, vol. 1, pp. 36–41, Fort Collins, Colo, USA, June 1999.

[12] S. J. F. Guimarães, M. Couprie, A. de A. Araújo, and N. J. Leite, "Video segmentation based on 2D image analysis," *Pattern Recognition Letters*, vol. 24, no. 7, pp. 947–957, 2003.

[13] S. J. F. Guimarães, A. de A. Araújo, M. Couprie, and N. J. Leite, "Video fade detection by discrete line identification," in *Proc. 16th International Conference on Pattern Recognition (ICPR '02)*, vol. 2, pp. 1013–1016, Quebec, Canada, August 2002.

[14] S. J. F. Guimarães, M. Couprie, N. J. Leite, and A. de A. Araújo, "Video transition sharpening based on flat zone analysis," in *Proc. IEEE-EURASIP Workshop on Nonlinear Signal and Image Processing—NSIP*, Grado-Trieste, Italy, June 2003, IEEE.

[15] A. Del Bimbo, *Visual Information Retrieval*, Morgan Kaufmann Publishers, San Francisco, Calif, USA, 1999.

[16] J. Serra, *Image Analysis and Mathematical Morphology*, vol. 1, Academic Press, London, UK, 1982.

[17] H. J. A. M. Heijmans, *Morphological Image Operators*, Academic Press, Boston, Mass, USA, 1994.

[18] N. J. Leite and S. J. F. Guimarães, "Morphological residues and a general framework for image filtering and segmentation," *EURASIP Journal on Applied Signal Processing*, vol. 2001, no. 4, pp. 219–229, 2001.

[19] G. Matheron, *Random Sets and Integral Geometry*, John Wiley, New York, NY, USA, 1975.

Silvio J. F. Guimarães received the B.S. degree in computer science from Federal University of Viçosa, Brazil, in 1997, the M.S. degree in computer science from State University of Campinas, Brazil, in 1999, and the Ph.D. degree in computer science from Federal University of Minas Gerais, Brazil, and from Université de Marne-la-Vallée, France, in 2003. He is currently an Associate Professor with the Institute of Computing in Pontifical Catholic University of Minas Gerais (PUC Minas), Brazil, where he directs works on image processing and analysis. His main research interests include mathematical morphology, digital topology, image filtering and segmentation, multiscale representation, and content-based video/image analysis/retrieval.

Neucimar J. Leite received the B.S. and M.S. degrees in electrical engineering from Universidad Federal da Paraíba, Brazil, in 1986 and 1988, respectively, and the Ph.D. degree in computer science from Pierre & Marie Curie University, Paris, France, in 1993. He is currently an Associate Professor at the Institute of Computing, State University of Campinas, Brazil, where he directs works on image processing and analysis. His main research interests include mathematical morphology, image filtering and segmentation, multiscale representation, and content-based video/image retrieval.

Michel Couprie received his Ingénieur's degree from the École Supérieure d'Ingénieurs en Électronique et Électrotechnique, Paris, France, in 1985 and the Ph.D. degree from the Pierre & Marie Curie University, Paris, France, in 1988. Since 1988 he has been working in ESIEE where he is an Associate Professor. He is a member of the Laboratoire Algorithmique et Architecture des Systèmes Informatiques, ESIEE, Paris, and of the Institut Gaspard Monge, Université de Marne-la-Vallée. His current research interests include image analysis and discrete mathematics.

Arnaldo de A. Araújo was born in July 1955, Campina Grande-PB, Brazil. He received his B.S., M.S., and D.S. degrees in electrical engineering, from the Universidad Federal da Paraíba (UFPB), Brazil, in 1978, 1981, and 1987, respectively. Arnaldo is currently an Associate Professor at the Computer Science Department (DCC), Universidad Federal de Minas Gerais (UFMG), Belo Horizonte, MG, Brazil since 1990. He was a Visiting Researcher at the Informatics Department, Groupe ESIEE, Paris, France, 1994–1995, a Visiting Professor at DCC/UFMG, in 1989, an Associate Professor at the Electrical Engineering Department (DEE), UFPB, 1985–1989, a Research Assistant at the Rogowski-Institut, RWTH Aachen, Germany, 1981–1985, and an Assistant Professor at DEE/UFPB, 1978–1985. His research interests include digital image processing, computer vision applications to medicine, fine arts, and satellite imagery, content-based image and video retrieval, and multimedia information systems. He has published more than 90 papers, supervised 21 M.S. dissertations, and 5 Ph.D. thesis.