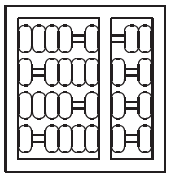


INSTITUTO DE COMPUTAÇÃO
UNIVERSIDADE ESTADUAL DE CAMPINAS

Identification of video transitions by multi-scale gradient analysis

S. J. F. Guimarães *N. Leite* *A. A. Araújo* *M. Couprie*

Technical Report - IC-03-013 - Relatório Técnico

May - 2003 - Maio

Identification of video transitions by multi-scale gradient analysis

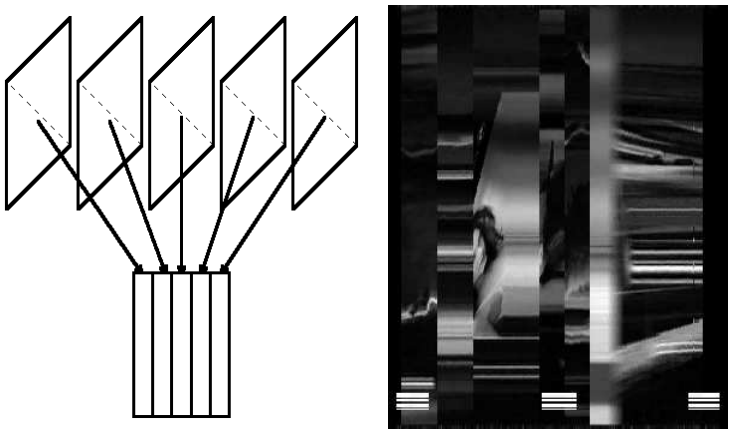
Silvio Jamil Ferzoli Guimarães - UFMG
Neucimar Jerônimo Leite - UNICAMP
Michel Couprie - ESIEE
Arnaldo de Albuquerque Araújo - UFMG

Abstract

The video segmentation problem consists in the identification of boundaries between two consecutive shots. These boundaries can be subdivided in cut and gradual transitions. The first one is characterized by a concatenation of two shots and the second one corresponds to the gradual transformation of a shot into another. The simplest gradual transitions are fade and dissolve. The common approach to solve the identification problem of these features in a video is based on dissimilarity measures between frames. The possibility of representing the video transitions by specific patterns in a 2D image, called visual rhythm, and to apply image processing tools on this image is an interesting alternative to cope with this problem. In this work, we consider the morphological multi-scale gradient operators to identify cut and gradual transitions at the same time on the visual rhythm representation. Three different morphological operators are considered: a multi-scale gradient proposed by Soille and two variants of this operation based on ultimate erosion and thinning concepts. Also, the Canny's filter is used to compute the gradient values.

1 Introduction

The segmentation of the video structures can be considered as a problem of dissimilarity between images (or frames). Usually, the common approach to cope with this problem is based on dissimilarity measures that are used to identify the boundary between consecutive shots. The simplest transitions between two consecutive shots are cut and gradual transitions [10]. A cut is a concatenation of two consecutive shots. When there is a gradual transition between two shots, new frames are created from these shots [10]. Usually, there are specific methods for each kind of transition. Methods to identify the cuts can be found in [17, 2, 9]. Methods for detecting fade and dissolve can be found in [18, 6, 12, 8, 19]. Most of these methods consider dissimilarity measures to characterize the video transition. If two frames belong to the same shot, then their dissimilarity measure should be small. Two frames belonging to different shot present high dissimilarity value, which can be affected by the occurrence of different effects such as, zoom, pan, tilt, flash, and so on. Thus, the choice of a good measure is essential for the quality of the segmentation results.

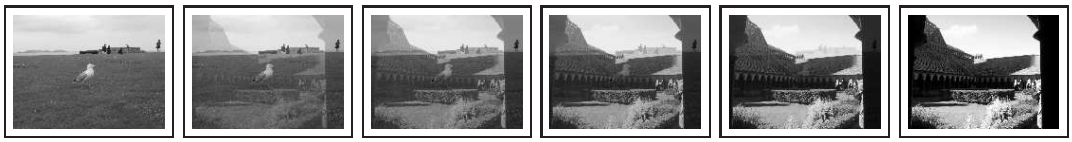


(a) (b)

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

(a) Cut

(b) Fade-out



(c) Dissolve

Figure 2: Example of transitions

Another approach to the video segmentation problem is to transform the video images into a 2D image representation, $\mathcal{R}$, and to apply image processing methods on $\mathcal{R}$ to extract the different patterns related to each transition. Informally, each frame is transformed into a vertical line of $\mathcal{R}$, as illustrated in Fig. 1(a). This approach can be found in [16, 4, 13, 9, 8, 7]. A definition of visual rhythm is given in [4] and in [13], both definitions are related to the same video transformation and sub-samplings of each frame, like the principal diagonal sub-sampling, as illustrated in Fig. 1(b). In [4], Chung et al. applied statistical measures to detect some patterns, but the number of false detections is very high. In [13], Ngo et al. applied Markov models for shot transition detection which fails when there is low contrast between textures of consecutive shots. In [9], we proposed a method to identify cuts based on visual rhythm analysis considering morphological and topological tools. In [8], we proposed a method to identify fades based on visual rhythm by histogram analysis that considers discrete line identification. Here, we propose a new approach to detect both cut and gradual transitions. The idea of this approach is to identify fuzzy and abrupt transitions vertically aligned, in the visual rhythm representation, through multi-scale gradient operators. In this work, we consider three different morphological multi-scale gradient operators: the Soille's multi-scale gradient and two other proposed variants of this method based on ultimate erosion and thinning concepts. And also, we use the Canny's filter [3] to compute the gradient values.

This paper is organized as follows. In Sec. 2, we introduce the visual rhythm representation. In Sec. 3, we present the basic concepts on mathematical morphology considered here. In Sec. 4, we define the multi-scale morphological gradient operators. In Sec. 5, we propose a methodology based on multi-scale operators to identify both the cut and the gradual transitions, and discuss some experiments. Finally, in Sec. 6, we give some conclusions.

2 Video transformation

Let $D \subset Z^2$, $D = \{0, ..., M-1\} \times \{0, ..., N-1\}$, where M and N are the width and the height of each frame, respectively.

Definition 2.1 (Frame) A frame f_t is a function from D to Z, where for each spatial position (x, y) in D, $f_t(x, y)$ represents the grayscale value at the pixel position (x, y).

Definition 2.2 (Video) A video V, in domain $2D + t$, can be seen as a sequence of frames f_t. It can be described by

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

where *time* is the number of frames contained in the video. Next, we give some definitions about the video transitions, considered in this work, namely, cut, fade and dissolve.

2.1 Cut

Definition 2.3 (Cut) A cut is a kind of transition in which two consecutive shots are concatenated.

In Fig. 2(a), we illustrate an example of cut.

2.2 Fade

The fade transition is characterized by a progressive darkening of a shot until the last frame becomes completely black, or inversely [2]. A more general definition of fade is given in [12] where the black frame is replaced by a monochrome frame. The fades can be subdivided into fade-ins and the fade-outs.

Definition 2.4 (Fade) A fade is a progressive transition from a monochrome frame M to a visual intensity frame P, or inversely.

$$T_{f_i}(t) = \alpha(t) \times M + (1 - \alpha(t)) \times P(t) \quad (fade - in)$$
$$T_{f_o}(t) = \alpha(t) \times P(t) + (1 - \alpha(t)) \times M \quad (fade - out)$$

where $\alpha(t)$ is a transformation function that is usually a linear function and $t \in [0, t_f - 1]$, t_f being the duration time of the fade. M is a monochrome frame (e.g., white or black). Fig. 2(b) illustrates an example of fade-out.

2.3 Dissolve

Unlike the cut, the dissolve transition is characterized by a progressive transformation of a shot P into another shot Q.

Definition 2.5 (Dissolve) Dissolve is a progressive transition from a shot (P) to a consecutive shot (Q) with a non-null duration. Each frame of the transition can be given by:

$$T_d(t) = (1 - \alpha(t)) \times P(t) + \alpha(t) \times Q(t) \tag{2}$$

Usually, the dissolve is considered as a generalization of the fade, where the monochrome frame is replaced by the first (last) frames of the shot. Fig. 2(c) illustrates an example of dissolve.

2.4 Visual rhythm

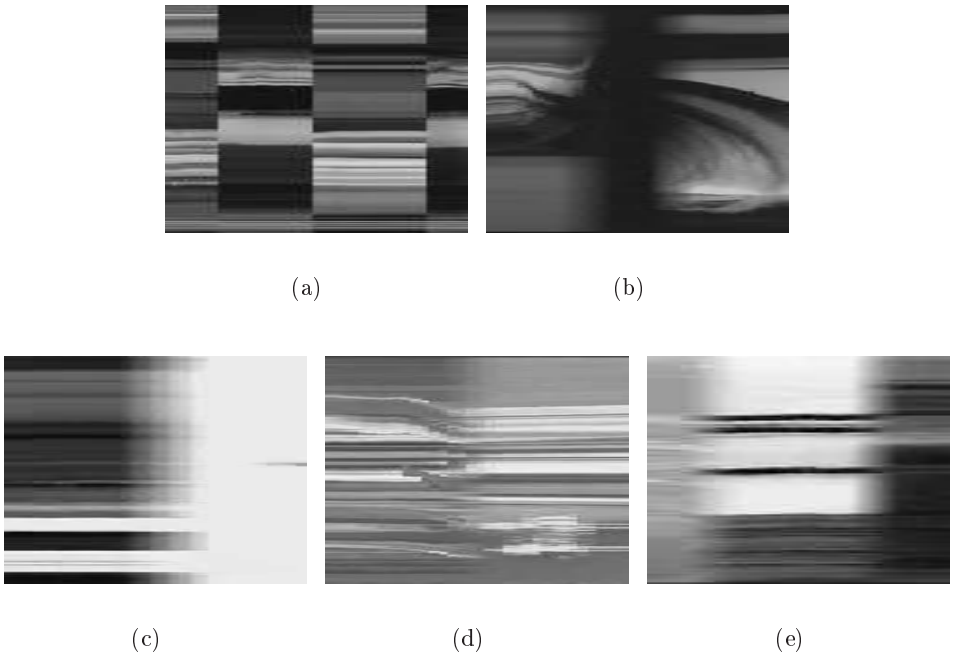
When we work directly on the video, we have to cope with two main problems: the processing time and the choice of a dissimilarity measure. Looking for reducing the processing time and using tools for 2D image segmentation instead of a dissimilarity measure, we transform the video into a 2D image, called visual rhythm [4, 13].

Definition 2.6 (Visual rhythm) Let $\mathrm{V} = (f_t)_{t \in [0, time-1]}$ be an arbitrary video, in domain $2D + t$. The visual rhythm ϑ, in domain $1D + t$, is a simplification of the video where each frame f_t is transformed into a vertical line on the visual rhythm

$$\vartheta(t, z) = f_t(r_x * z + a, r_y * z + b) \tag{3}$$

where $z \in \{0, ..., M_\vartheta - 1\}$ and $t \in \{0, ..., N_\vartheta - 1\}$, M_ϑ and N_ϑ are the height and the width of the visual rhythm, respectively, r_x and r_y are ratios of pixel sampling, a and b are shifts on each frame. Thus, according to these parameters, different pixel samplings could be

considered, for example, if $r_x = r_y = 1$ and $a = b = 0$ and $M = N$, then we obtain all pixels of the principal diagonal. If $r_x = -1$ and $r_y = 1$ and $a = M$ and $b = 0$ and $M = N$, then we obtain all pixels of the secondary diagonal. If $r_x = 1$ and $r_y = 0$ and $a = 0$ and $b = N/2$, then we obtain all pixels of a central horizontal line. If $r_x = 0$ and $r_y = 1$ and $a = M/2$ and $b = 0$, then we obtain all pixels of a central vertical line.



(a) (b)

(c) (d) (e)

Figure 3: Example of patterns on the visual rhythm associated with cut and gradual transitions.

The choice of the pixel sampling is an interesting problem because different samplings can produce different visual rhythms with different patterns. In [4] we can see some pixel samplings together with their corresponding visual rhythm patterns. In this work, it is said that the best results are obtained by considering diagonal sampling of the images because they can encompass horizontal and vertical features. In Fig. 3 we give some examples of patterns that can be found on the visual rhythm. According to the features of the transitions, we note that all cuts are represented by "vertical sharp lines". Also, all gradual transitions are represented by vertical aligned gradual regions, independently on the pixel sampling. Fig. 3(a) illustrates the cut transition. Fig. 3(b) and (c) give examples of fade and, finally, Fig. 3(d) and (e) correspond to dissolves.

3 Basic concepts

In this section, we describe some basic morphological operators considered in this work (see [15, 14] for more details). Let B be a flat structuring element (SE) and its homotetic

representation ($\lambda B = \{\lambda b \mid b \in B\}, \lambda \geq 0$), such that the size λ represents the radius information of the SE.

Definition 3.1 (Morphological gradient) Let δ_n and ε_n be the dilation and erosion with a structuring element B of size n, respectively. The morphological gradient ρ_n of size n is defined by

$$\rho_n = \delta_n - \varepsilon_n \tag{4}$$

The morphological gradient, also called thick gradient, gives the maximum variation of the function in a neighbourhood of size n. If the size n equals the width of the transition between regions of homogeneous grayscale, the morphological gradient will output the contrast value between these regions. However, the output of this gradient is represented by thick edges.

Another interesting morphological operator that extracts light regions of the image, smaller than a certain SE, is the white top-hat defined as follows.

Definition 3.2 (White top-hat) Let γ_n be an opening by a structuring element B of size n ($\gamma_n = \delta_n\varepsilon_n$). Let f be an original image. The white top-hat WTH_n of size n corresponds to the residue of the opening, and is defined by

$$\mathrm{WTH}_n = Id - \gamma_n \tag{5}$$

where Id is the identity operator, "$-$" may represent the binary or grayscale difference for binary and grayscale images, respectively.

The white top-hat represents the residues related to a specific size of the SE. Another operator can be considered if we replace the opening by the erosion in white top-hat definition. This new operator is called inf top-hat [5].

Definition 3.3 (Inf top-hat) Let ε_n be an erosion by a structuring element B of size n (ε_n). Let f be an original image. The inf top-hat ITH_n of size n corresponds to the residue of the erosion, and is defined by

$$\mathrm{ITH}_n = Id - \varepsilon_n \tag{6}$$

Two other important operators, the ultimate erosion and the thinning, that will be explored in this work, are associated with the simplification of the images. The ultimate erosion indicates the moment in which a connected component disappears. Its definition is given next.

Definition 3.4 (Ultimate erosion) The ultimate erosion represents the set of all components of an image that disappears from one erosion step to the other, when we consider increasing SE. It is defined for binary and grayscale images as follows.

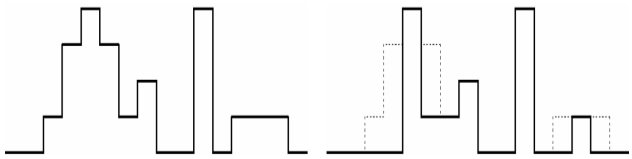
$$ULT(X) = \bigcup_n \{{\varepsilon_1}^{(n)}(X) \backslash G_{{\varepsilon_1}^{(n)}(X)}[{\varepsilon_1}^{(n+1)}(X)]\} \tag{7}$$

$$ULT(f) = \bigvee_n \{{\varepsilon_1}^{(n)}(f) - G_{{\varepsilon_1}^{(n)}(f)}[{\varepsilon_1}^{(n+1)}(f)]\} \tag{8}$$

where G is the morphological reconstruction by dilation operator [15, 14].

Definition 3.5 (Thinning) Let us consider a point x in a 1D image (or signal) g. We say that a point x is destructible for g if one neighbor of x has a value greater than or equal to $g(x)$ and the other neighbor has a value strictly smaller than $g(x)$. The thinning procedure consists in repeating the following steps until stability: i) select a destructible point x; ii) lower the value of x downto the value of its lowest neighbor.

Selection of destructible points must be done in increasing order of value, so that each point is modified at most once. Points having the same value are scheduled with a fifo policy which guarantees that, in case of large flat maxima, the thinned signal is "well centered" with respect to the original one. This procedure is in fact a particular case, in the 1D domain, of a topological operator introduced by [1]. Topological operators have as aim to simplify the image while maintaining the topology. [1] presented operators for image segmentation based upon topology which generalizes to 2D grayscale images the notions of binary digital topology ([11]). In Fig. 4, we illustrate the thinning of a 1D image.



(a) Original (b) Result

Figure 4: Example of 1D thinning. Dotted line in (b) represents transformed pixels of the original image.

4 Multi-scale gradient based on WTH

The possibility to detect smooth variations between regions by considering thick gradient operators provides an important tool to identify gradual video transitions. Nevertheless, the result of these operators yields thick edges as result and, also, false edges (merge of the gradual boundaries) that can be produced if the distance between two transitions is smaller than the width of the used structuring element. These problems can be avoided if we consider the multi-scale gradient proposed by Soille in [15]. In this section, we describe this gradient and propose two variants of this operator based on the morphological ultimate erosion and thinning transformations.

4.1 Soille's gradient

The multi-scale gradient model proposed by Soille can be represented by the block diagram in Fig. 5. According to this model [15], the problem of thickness introduced by the morphological gradient is avoided by the application of erosions on the thick image. To avoid

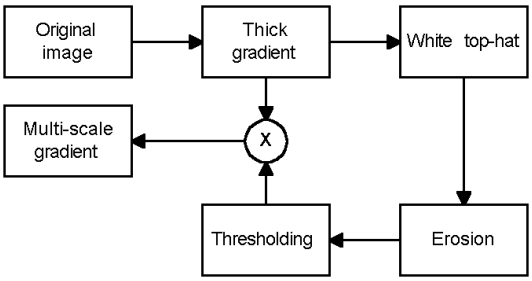


Figure 5: Soille's multi-scale gradient.

the merge of boundaries, a white top-hat is applied to the thick gradient before the erosion operation.

Definition 4.1 (Soille's morphological multi-scale gradient) The morphological multi-scale gradient at scale n is given by:

$$\rho_n^S = \rho_n * L_{[1]}\varepsilon_{(n-1)}\mathrm{WTH}_n\rho_n \tag{9}$$

where ρ_n, $\varepsilon_{(n-1)}$ and WTH_n represent the thick gradient of size n, the erosion of size $n-1$ and the white top-hat of size n, respectively. $L_{[1]}$ represents a thresholding operation.

For the application considered here, the problems of this approach are directly associated with the choice of the SE family. The first problem we can identify is related to the quality of the detection. For example, let us consider a 1D signal (Fig. 6(a)) and a homotetic family of SE. If we apply the Soille's gradient on such configuration the transition is not well identified for the corresponding scale (Fig. 6(b) and (c)). The second problem concerns the elimination of regions due to the introduction of the erosion operation. Finally, from this approach, it is not possible to classify the detected transitions according to a parameter of size, i.e., we can not identify the transitions of a specific size k because these transitions, identified at a certain scale i, correspond to the transition size smaller than i. Intuitively, this problem can be avoided by the difference between two consecutive levels, but the gradient values depend on the thick gradient that may vary at consecutive scales. In Fig. 7(b) we show the result of the Soille's gradient applied to the image in Fig. 7(a). Here, the parameter of size is in the range $[1, 7]$, and the final result corresponds to the supremum of the gradient values at all levels.

To cope with the above problems, we propose to replace the erosion by the ultimate erosion operation. In this case, we preserve the same properties of the Soille's gradient, that is, the computation of thin edges and non merge the gradual boundaries. We also preserve all regions smaller than the corresponding SE.

4.2 Multi-scale gradient based on ultimate erosion

Here, we replace the erosion in Fig. 5 by the ultimate erosion.

Definition 4.2 (Multi-scale gradient based on ultimate erosion) Our proposed morphological multi-scale gradient at scale n is defined by:

$$\rho_n^U = \rho_n * L_{[1]} ULT(\mathrm{WTH}_n \rho_n) \tag{10}$$

In Fig. 6(d) and (e) we illustrate the application of the multi-scale gradient based on ultimate erosion, considering a certain scale and a range of scales, respectively.

The problem identified here is associated with noise sensitivity, and with the fact that the transitions are not thin. Also, it is not possible to classify the transitions regions according to a size criterion. In Fig. 7(c), we show the result of the multi-scale gradient based on ultimate erosion of the image in Fig. 7(a). The parameter of size is in the range $[1, 7]$, and the final result corresponds to the supremum of the gradient values at all levels.

Next, we propose a variant that replaces the ultimate erosion by a thinning operator.

4.3 Multi-scale gradient based on thinning

While the erosion operation may produce thick edges, the thinning transformation defines one-pixel-thin edges. Another feature of this operator is that it preserves topology in the sense of [1].

Definition 4.3 (Multi-scale gradient based on thinning) Our proposed morphological multi-scale gradient at scale n is defined by:

$$\rho_n^T = \rho_n * L_{[1]} \mathcal{T}(\mathrm{WTH}_n \rho_n) \tag{11}$$

where $\mathcal{T}$ represents the thinning operator described previously.

In Fig. 6(f) and (g) we illustrate the application of the multi-scale gradient based on thinning, considering a certain scale and a range of scales, respectively.

The problem of this approach is the high noise sensitivity. In Fig. 7(d), we illustrate the gradient based on thinning applied to the image in Fig. 7(a). Here, the parameter of size is in the range $[1, 7]$ and the final result corresponds to the supremum of the gradient values at all levels.

5 Transition detection

The possibility to identify different kinds of transitions constitutes an important and interesting aspect of our approach. This is possible mainly due to the use of the morphological multi-scale analysis. In this section, we present our algorithm for video transition identification, followed by a description and analysis of the realized experiments.

5.1 Algorithm

In [9], we proposed an approach for cut transition identification illustrated by the block diagram in Fig. 8(a). The method can be subdivided in two main modules: the gradient module and the spatio-temporal module. The former computes the presence of a gradient and the second one verifies if the gradient values are vertically aligned. Here, we propose

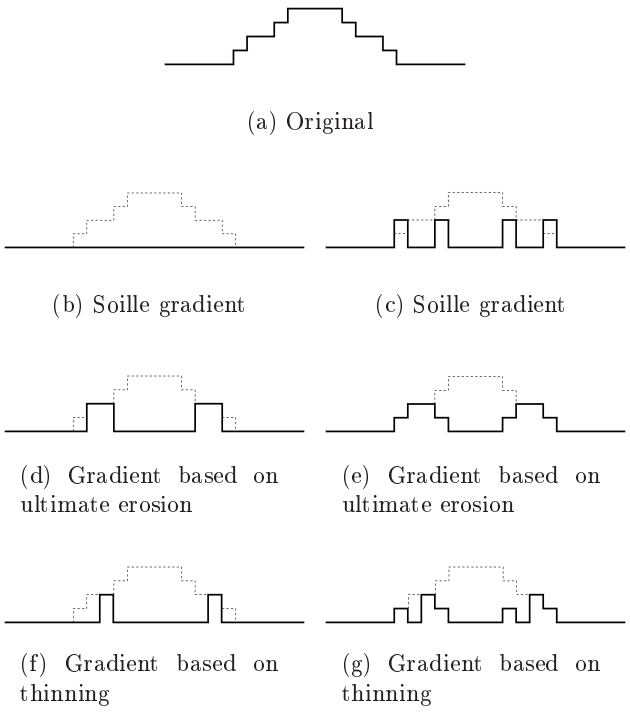
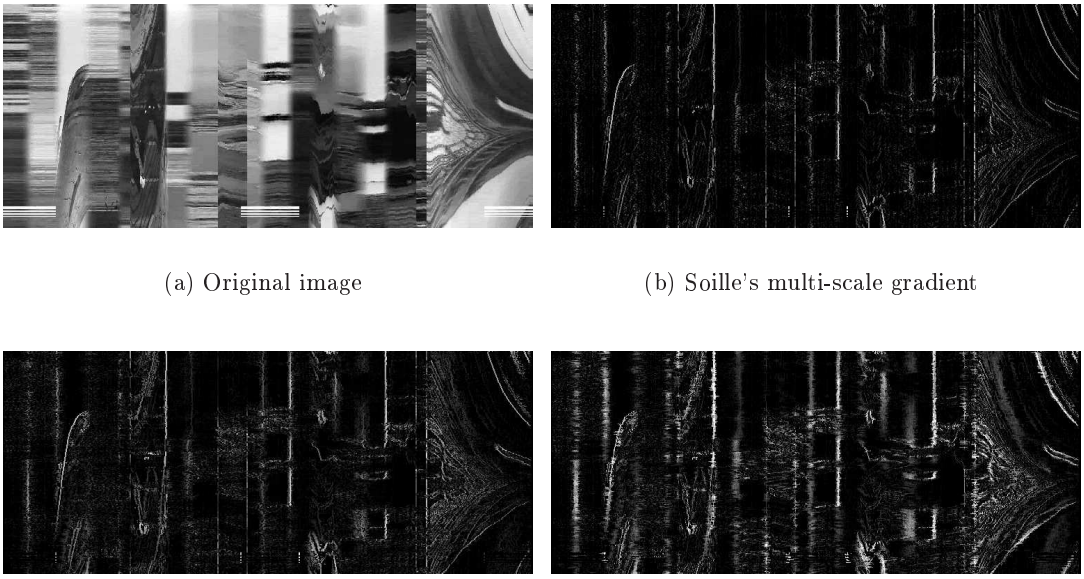


Figure 6: Morphological multi-scale gradient: (b,d,f) correspond to the gradient values at a specific level $n = 4$ and (c,e,g) correspond to the supremum of the gradient values at different levels ($n = [1, 5]$) .

to use a very similar approach, where the gradient module is replaced by a multi-scale gradient module and the spatio-temporal module is internally modified to adequate the possibility of identifying thick transitions. In Fig. 8(b) we illustrate a block diagram of our proposed method to identify both cut and gradual transitions. Next, we describe each module separately.

Let K be the maximum size of the transition to be detected. This value is related to the size n of the SE, and according to the SE family used in this work, $K = 2n + 1$. The multi-scale gradient module allows the identification of abrupt and fuzzy regions. As stated before, we consider three different types of gradient operators:

- Soille's gradient: this operator produces one-pixel-thin edges, but due to the fixed size of the SE used in the erosion operation, some gradient regions are not detected.
- Ultimate erosion: this operator preserves all gradient regions according to the properties of this erosion, and also one-pixel-thin or two-pixel-thin edges are produced in each level.
- Thinning: this operator preserves all gradient regions, as well as the topology (in the sense of [1]) of the result obtained from the thick gradient.



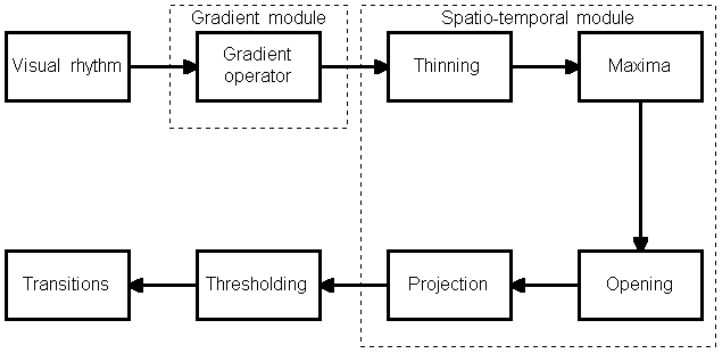
(a) Original image

(b) Soille's multi-scale gradient

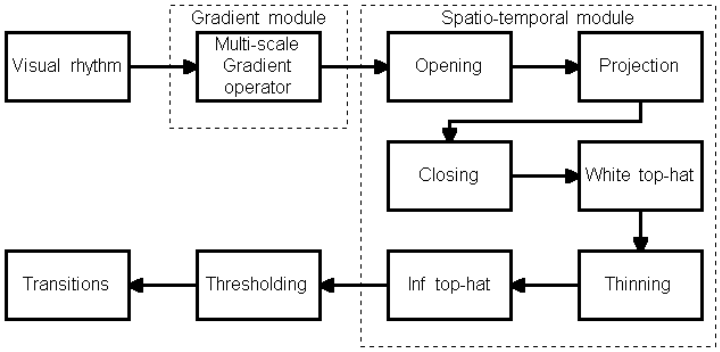
(c) Multi-scale gradient based on ultimate erosion

(d) Multi-scale gradient based on thinning

Figure 7: Examples of multi-scale gradients where the SE size is in range $[1, 7]$. These results correspond to the supremum of gradient values of all levels.



(a) Simple gradient



(b) Multi-scale gradient

Figure 8: Block diagram for detection of transitions.

These operators are computed with respect to a certain scale n, and to combine the results of various scales, a supremum operation is realized. This operation may produce thick edges for all gradient operators.

Here, we notice that we are interested in the presence of the gradient, without taking into account its values. Thus a thresholding operator is applied to the result of the gradient operator. In Fig. 9 we illustrate examples of thresholded multi-scale gradient obtained from the image of Fig. 7(a) in which we compute the Soille's gradient at scale $n = 7$ (Fig. 9(a)), the multi-scale gradient based on ultimate erosion at scale $n = 7$ (Fig. 9(b)), and finally, the supremum of the multi-scale gradient based on thinning at scales in the range $n = [1, 7]$ (Fig. 9(c)).

The spatio-temporal module executes the identification of vertical aligned transitions. Next, we describe each step of this module according to the block diagram described in Fig. 8(b). Here, the input of the spatio-temporal module is a binary image corresponding to the output of the gradient module:

- Opening: this operation is used to eliminate small spatial components, i.e., we ignore the time information. So, we consider a vertical SE with size (radius) equals to 2.
- Projection: this operation is used to compute the number of non-zero gradient values in each column. A 1D signal is produced by this step.
- Closing: the aim of this operation is to eliminate small "holes" on the projected image. Empirically, we verified that a good size of SE for the closing is 3, which is associated with the size of the smallest shot ($k = 2 * 3 + 1 = 7$).
- White top-hat: the aim of this operation is to consider only the "dome" that are related to a transition size n . Thus, we compute WTH_n.
- Thinning: the aim of this operation is to find the center of transitions represented by the regional peaks on the projection image.
- Inf top-hat: this operation computes the peak values with respect to its neighborhood. As the results of the thinning is an image with one-pixel-thin edges, an inf top-hat of level 1 is sufficient to calculate the peak value. This operation constitutes a type of filtering in which the plateau values are not considered. This filtering can eliminate regions that are not vertically aligned but vary with respect to time.

In Fig. 9(b), Fig. 9(d) and Fig. 9(f) we illustrates the 1D signals produced by the projection operation, computed by Fig. 9(a), Fig. 9(c) and Fig. 9(e), respectively.

5.2 Experiments

Nowadays, our video database contains approximately 500 videos, but we consider only 29 videos that are chosen according to the presence of zoom, tilt, flash, dissolve, fade, cut, etc. In Table 1, we outline some features of our video corpus.

To compare the experiments, we use the quality measures defined in [9]. The experiments realized in this work consider the multi-scale gradient by Soille, and its variants based on the ultimate erosion and on the thinning transformation. Also, we realize an

	Number of events	Average size
Cut	250	0
Gradual ($\leq$15 frames)	140	9.00
Gradual (>15 frames)	56	23.2

Table 1: Features of the video corpus.

experiment considering the Canny's filter [3] to compute the multi-scale gradient instead of morphological gradient.

The main feature of the morphological gradient operators presented here is that all transitions smaller than the parameter K are detected considering a SE of size n. In our experiments, the parameter of SE size is equal to 7, and consequently we search to detect the transitions of size smaller than $K = 2n + 1 = 15$, this value corresponds to the average size of the gradual transitions. According to the features of the multi-scale gradient based on thinning, we realize its experiments with the parameter of SE size in the range $n = [1, 7]$.

With respect to the Canny's filter, the parameter used is $\alpha = 0.5$, that is associated with the size of the smallest shot. Afterwards the computation of this gradient for each row on the visual rhythm, we realize the following algorithm: the opening operation, to reduce the noise; the thinning operation, to find the center of each dome; to detect the horizontal maxima; and finally, to apply the spatio-temporal analysis as stated above.

In Fig. 10, we show the graphics that relate some quality measures to the threshold value. In Table 2, we outline the quality measure values obtained from these experiments.

5.3 Analysis of results and parameters

As illustrated in Fig. 10, the results of our experiments are very similar. In general, the method based on thinning is more sensitive to low and high values of threshold due to the higher number of regions. On the other hand, the size of regions on the 1D projected image is directly associated with the size of the transitions (as we can observe in Fig. 9). As we can note in Fig. 9, the method based on thinning produces more information and, consequently, a high number of correct detections is identified as well as false detections.

According to all quality measures, the multi-scale gradient proposed by Soille presents slightly better results, with respect to the studied variants, due to the fact that it produces less irrelevant information than the other gradients and, consequently, the reward function between the false detection and correct detection is higher. The problem of Canny's filter is its sensitivity to threshold values, that is, the number of correct detections decreases very fast when the values of threshold increase, with respect to the other gradient used here.

The misses in our experiments are mainly due to the quality of the videos and the presence of motion in the gradual transitions and, also, the presence of chained dissolves.

Finally, we can identify two types of parameters used in these experiments: the fixed parameters, which are set only once, associated with the size of the SE and used to filter the intermediate images; and the variable parameters, related to the size of transitions to be

	μ	E_m	R_f	γ
Soille	**0.21**	0.40	**0.47**	0.70
Ultimate	**0.21**	0.46	**0.47**	0.67
Thinning	0.17	**0.36**	0.35	0.72
Canny	0.12	**0.36**	0.30	**0.74**

Table 2: Quality measures: robustness $\mu(0.3, 0.3)$, gamma measure γ, $E_m(0.08)$, $R_f(0.02)$

detected and the considered threshold. The tuning of these parameters is desirable because it allows the user to find a compromise between over-segmentation and under-segmentation.

6 Conclusion

In this work, we study the problem of video segmentation, where we transform the video into a 2D image applying tools of image processing. Here, we propose a new approach to identify cut and gradual transitions. Our approach is based on morphological multi-scale gradient operators, more specifically, the Soille's gradient and two other variants that are based on ultimate erosion and thinning. An interesting aspect we remark here is the inclusion relation between the multi-scale ultimate erosion gradient and Soille's gradient, where the former contains the second one.

In a general way, the Soille's gradient present the best results, but the gradient based on thinning contains more information concerning, for example, the transition size. In this work, we are interesting mainly in identifying the presence of cut and gradual transitions without taking account the size of these transitions. This information will be considered in a future work.

From the above experiments, we observe that the use of the multi-scale gradient can represent an interesting approach to cope with the problems of gradual transitions. According to the features of the gradient based on thinning, a study about the relationship between the transition size and this thinning can be envisaged. Also, a study of the behavior of this method could be done considering its application directly to the video data.

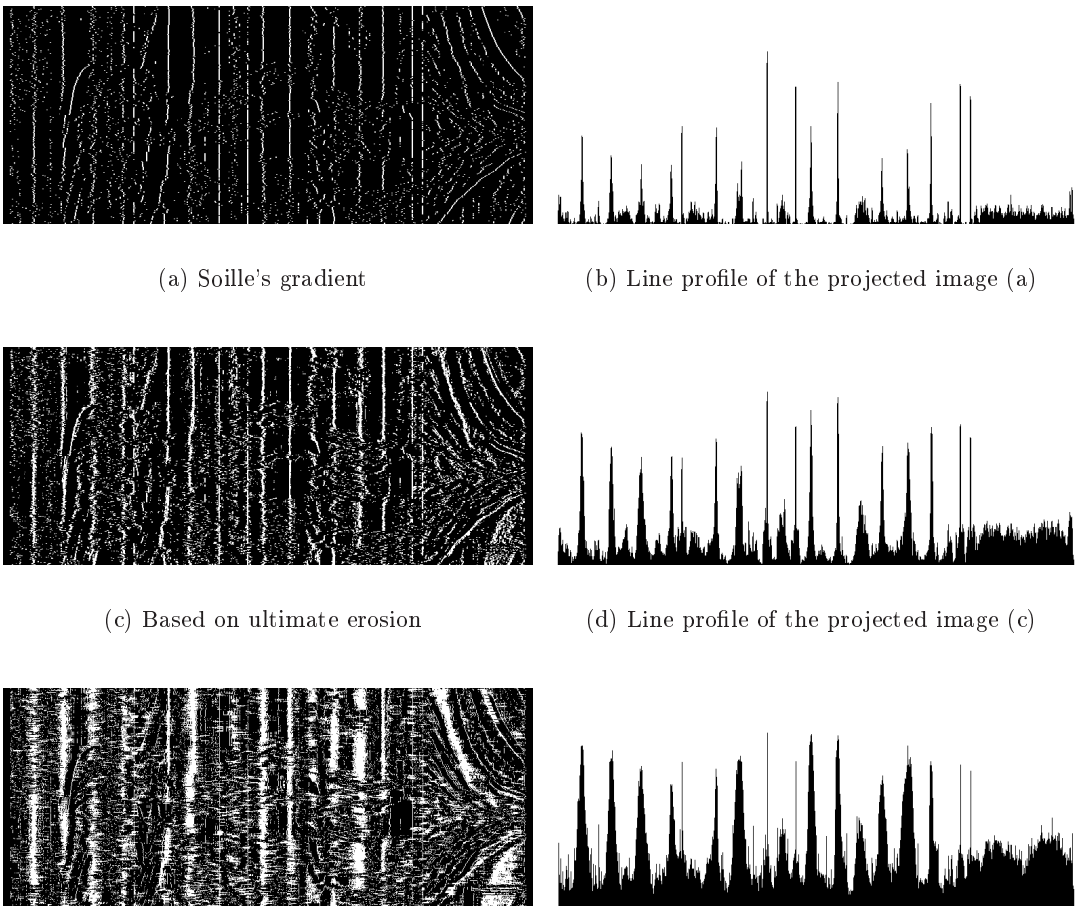
Acknowledgements

The authors are grateful to CNPq, CAPES/COFECUB, FAPEMIG and the SIAM DCC/PRONEX Project for the financial support of this work.

References

[1] G. Bertrand, J.-C. Everat, and M. Couprie. Image segmentation through operators based upon topology. *Journal of Electronic Imaging*, 6:395–405, 1997.

[2] A. D. Bimbo. *Visual Information Retrieval*. Morgan Kaufmann, 1999.



(a) Soille's gradient

(b) Line profile of the projected image (a)

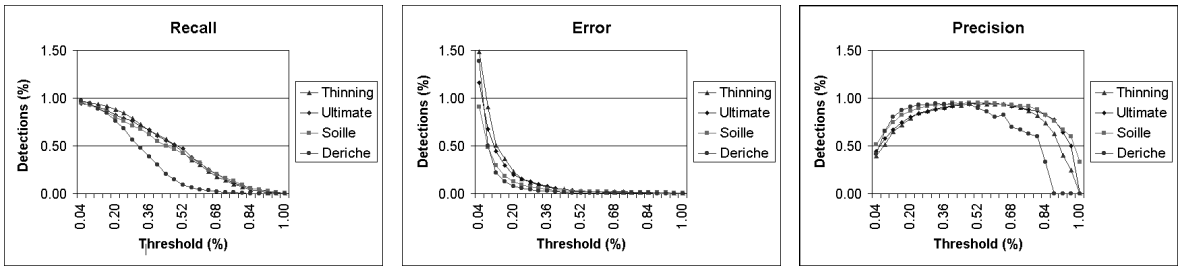
(c) Based on ultimate erosion

(d) Line profile of the projected image (c)

(e) Based on thinning

(f) Line profile of the projected image (e)

Figure 9: Example of multi-scale gradient analysis for transition detection, where $n = 7$ for the Soille's multi-scale gradient and gradient based on ultimate erosion, and $n = [1, 7]$ for the gradient based on thinning: the gradient computation (a,c,e) and the line profile of the projected image (b,d,f).



(a) (b) (c)

Figure 10: Graphics of the quality measures.

[3] J. Canny. A computational approach to edge detection. *IEEE Trans. on PAMI*, 8(6):679–698, 1986.

[4] 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*, 4(1):4–14, 1999.

[5] C.-H. Demarty. *Segmentation et Structuration d'un Document Vidéo pour la Caractérisation et l'Indexation de son Contenu Sémantique*. PhD thesis, École Nationale Supérieure des Mines de Paris, Janvier 2000.

[6] W. A. C. Fernando, C. N. Canagarajah, and D. R. Bull. Fade and dissolve detection in uncompressed and compressed video sequences. In *Proc. of the IEEE ICIP*, pages 299–303, 1999.

[7] S. J. F. Guimarães, M. Couprie, N. J. Leite, and A. A. Araújo. A method for cut detection based on visual rhythm. In *Proc. of the IEEE SIBGRAPI*, pages 297–304, Brazil, 2001. ISBN 0769513301.

[8] S. J. F. Guimarães, M. Couprie, N. J. Leite, and A. A. Araújo. Video fade detection by discrete line identification. In *Proc. of the 16th ICPR*, volume 2, pages 1013–1016, Quebec, Canada, Aug 2002. ISBN 0 7695 1699 8.

[9] S. J. F. Guimarães, M. Couprie, A. A. Araújo, and N. J. Leite. Video segmentation based on 2D image analysis. *Pattern Recognition Letters*, 24:947–957, April 2003.

[10] A. Hampapur, R. Jain, and T. E. Weymoth. Production model based digital video. *Multimedia Tool and Aplications*, 1:9–46, 1995.

[11] T. Y. Kong and A. Rosenfeld. Digital topology: Introduction and survey. *CVGIP*, 48:357–393, 1989.

[12] R. Lienhart. Comparison of automatic shot boundary detection algorithms. In *SPIE Image and Video Processing VII*, volume 3656, pages 290–301, Jan 1999.

[13] C. W. Ngo, T. C. Pong, and R. T. Chin. Detection of gradual transitions through temporal slice analysis. In *Proc. of the IEEE CVPR*, pages 36–41, 1999.

[14] J. Serra. *Image Analysis and Mathematical Morphology*, volume 1. Academic Press, 1982.

[15] P. Soille. *Morphological Image Analysis*. Springer-Verlag, 1999.

[16] Y. Tonomura, A. Akutsu, K. Otsuji, and T. Sadakata. Videomap and videospaceicon: Tools for anatomizing video content. In *ACM Interchi*, pages 131–136, 1993.

[17] Y. Wang, Z. Liu, and J.-C. Huang. Multimedia content analysis. *IEEE Signal Processing Magazine*, pages 12–36, 2000.

[18] R. Zabih, J. Miller, and K. Mai. A feature-based algorithms for detecting and classifying scene breaks. *Multimedia Systems*, 7:119–128, 1999.

[19] H. Zhang, A. Kanknhalli, and S. W. Smoliar. Automatic partitioning of full-motion video. *Multimedia Systems*, 1:10–28, 1993.