Identification of flashes in video based on $2\mathbb{D}$ image analysis

Silvio Jamil Ferzoli GUIMARÃES[1,2], Michel COUPRIE[2],
Neucimar Jerônimo LEITE[3], Arnaldo de Albuquerque ARAÚJO[4]

[1]NPDI/DCC - Universidade Federal de Minas Gerais, Caixa Postal 702, 30161-970, Belo Horizonte, MG, Brasil

[2]A[2]SI/ESIEE - Cité Descartes, BP 99, 93162, Noisy le Grand, France

[3]Instituto de Computação - UNICAMP Caixa Postal 6176, 13083-970, Campinas, SP, Brasil

{sjamil,arnaldo}@dcc.ufmg.br, {guimaras,coupriem}@esiee.fr, neucimar@dcc.unicamp.br

Abstract – The flash presence is very common in digital videos, mainly in television journal videos. When a flash occurs, an increase of the luminosity in a few frames is produced. When we calculate a dissimilarity measure, we can see that this measure is very high in the frames affected by a flash. In fact, the presence of flashes often disturbs the identification of video transitions. In this work, we transform the video into a $2\mathbb{D}$ image in which the flash is represented by a vertical light line. Considering this image, we propose two methods for flash detection without taking into account dissimilarity measures. The first one considers a filtering using white top-hat by reconstruction and the second one considers a filtering of the max tree calculated from statistical measures of each frame (or frame sub-samplings).

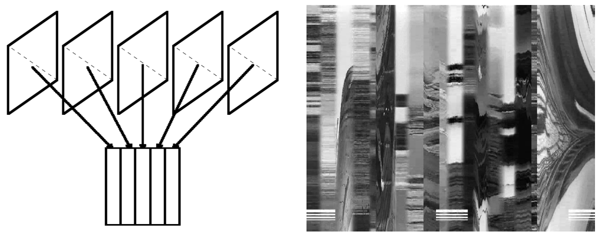
1 Introduction

The identification of flashes can be considered as an essential step to reduce the number of false detections when we try to identify shot boundaries, mainly cuts. Another application concerning the identification of flashes is associated with an important event, like when we have the President of Brazil making a speech, many photos are taken, and consequently, many flashes are dispatched. An interesting approach can be found in [7], in which a flash and a cut are distinguished according to some defined models. In this method, it is computed the average intensity of each frame. A flash is identified if the relation between different frames (sliding window) is verified. This method fails when a flash occurs at the same time of a cut. In Fig. 1, we illustrate two different models for a flash: Fig. 1(a) represents a flash that occurs in the middle of a shot; and Fig. 1(b) represents a flash that occurs in the boundary of a shot.

(a) (b)

FIG. 1: Flash model: (a) flash occurrence in the middle of the shot and (b) flash occurrence in the boundary of the shot.

Another approach to the flash identification problem is to transform the video into a $2\mathbb{D}$ image representation, VR, and to apply image processing methods on VR to extract the different patterns related to each transition [2]. For example, a flash is represented by a vertical light line. Informally, each frame is transformed into a vertical line of VR, as illustrated in Fig. 2(a). In this work, we propose two different methods to identify flashes, both from visual rhythm by sub-sampling: top-hat based and max-tree based. The first one can be related to the method proposed by [7] in the sense of using a neighborhood, but the principles are different, and also here we consider the values of the pixels without computing the statistical values. The second method is more sophisticated in the sense of identifying all types of flashes, even if it uses the average value of each frame.



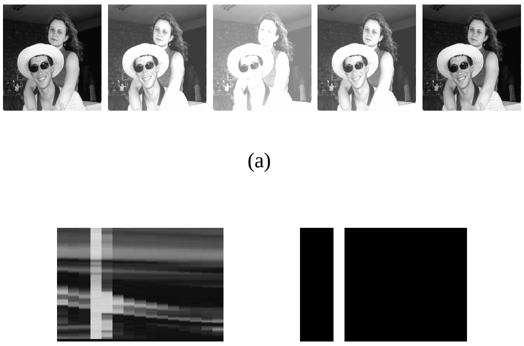
(a) (b)

FIG. 2: 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.

This work is organized as follows: in Sec. 2, we propose a method based on top-hat filtering. In Sec. 3, another method is proposed, in this case, a filtering of the statistical values is considered. In Sec. 4, some experiments are showed. Finally, some conclusions are given in Sec. 5.

2 Based on top-hat filtering

On the visual rhythm, we can observe that the video flashes are transformed into thin light vertical lines, as shown in Fig. 3. So, we can easily extract these lines from a white top-hat by reconstruction. The white top-hat by reconstruction is a mathematical morphology operator and represents the difference between the original image g and the opening by reconstruction of g [5, 6]. Informally, this operator detects light regions according to the shape and the size specifications of the structuring element.



(a) (b) (c)

FIG. 3: Flash video detection: (a) some frames of a sequence with the flash presence; (b) visual rhythm by sub-sampling; (c) detected flash.

Visual rhythm computation. The visual rhythm by sub-sampling is computed from the principal diagonal of each frame. In Fig. 4, we illustrate the visual rhythm by principal diagonal sub-sampling computed between the frames 740 and 800 from the video “example.mpg”.

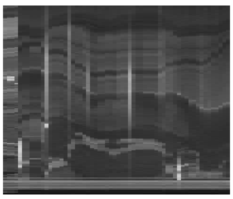
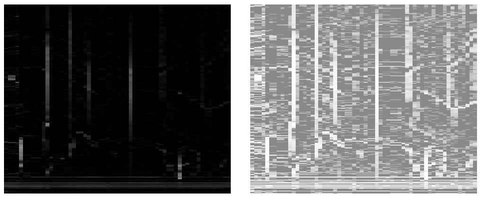


FIG. 4: Visual rhythm by principal diagonal sub-sampling computed for video “example.mpg”.

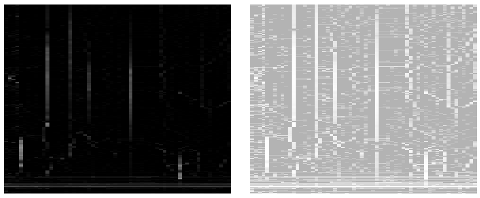
White top-hat filtering. Applying the white top-hat by reconstruction with a linear horizontal structuring element of size $\lambda = 5$. This size is associated with the potential duration of a flash. In Fig. 5, we illustrate the white top-hat and the result of the histogram equalization of the white top-hat to facilitate its visualization.

Thinning. Apply a $1\mathbb{D}$ thinning ([1]) to each horizontal line to identify the center of each event. In Fig. 6, we illustrate the result of the thinning and the result of the histogram equalization of the thinning to facilitate its visualization.



(a) (b)

FIG. 5: White top-hat filtering: (a) result of the white top-hat and (b) result of the histogram equalization of (a).



(a) (b)

FIG. 6: Thinning: (a) result of the thinning and (b) result of the histogram equalization of (a).

Detection of local maxima points. Considering that we have a thinned image, the computation of the maxima points is very trivial. We need to detect these points to verify if they are vertically aligned. In Fig. 7, we illustrate the result of this step in which the maxima points are detected.

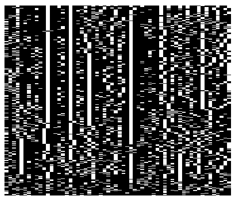


FIG. 7: Detection of maxima points

Maxima image filtering. Apply an opening by reconstruction with vertical structuring element of size $\lambda = 7$, defined empirically. If the maxima points are not aligned, then they do not represent a flash, and consequently they are eliminated. In Fig. 8, we illustrate the result of the maxima points filtering.

Detection of flashes. Depending on the number of maxima points that are present in the maxima image filtering, a flash must be identified. For that, we must calculate the number of maxima points in each vertical line, followed by application of thresholding. In Fig. 9, we illustrate the flashes detected, that are represented by white vertical bars.

This methodology detects the center of the regions in matter, in this case, regions with peak luminosity. Thus, we can have false detection in regions of high luminosity change that do not represent a flash. Usually, this method produces good results when the flash appears in the middle of the shot. The problem of this method is its noise sensitivity, and due to features of the

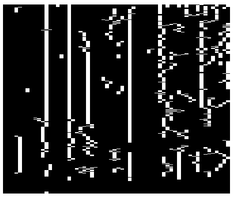
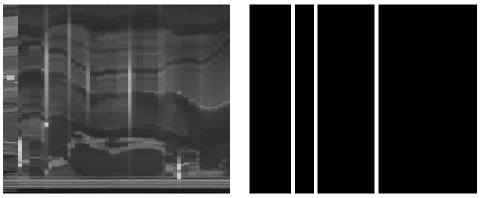


FIG. 8: Maxima image filtering.



(a) (b)

FIG. 9: Detection of flashes in which the flashes are represented by white vertical column bars: (a) original image with 4 flashes and (b) result of the flash detection.

used operators, we can not apply any filtering.

3 Based on max tree filtering

Usually, the frames affected by a flash are visually similar to their neighbors but with a higher luminosity. Here, the analysis of flash presence can be given by the computation of some statistical measures like mean or median, because the frames affected by a flash present higher mean and median values in regarding to their neighbors. From the computation of these statistical measures for all frames of the video, we can create a $1\mathbb{D}$ image from the visual rhythm, or preferably from the original video data. From this $1\mathbb{D}$ image, we need to find the "peaks" which the "height" is greater than a certain value h, and a "basis area" less than or equal to a value S that corresponds to the duration of the flash. Fig. 1 illustrates the peaks which must be identified jointly with their features. Next, we describe the main steps of this method.

Visual rhythm computation. The visual rhythm by sub-sampling is computed from the principal diagonal of each frame. In Fig. 10, we illustrate the visual rhythm by principal diagonal sub-sampling computed between the frames 740 and 800 from the video "example.mpg".

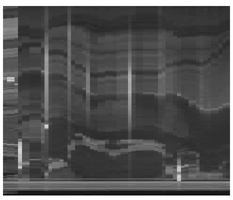


FIG. 10: Visual rhythm by principal diagonal sub-sampling computed from video "example.mpg".

Average. For each column of the visual rhythm, we compute its average value producing a $1\mathbb{D}$ signal. This computation is very important because a frame that is affected by a flash has a higher average value than its neighbors. In Fig. 11, we illustrate the average computation.

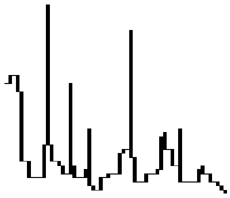
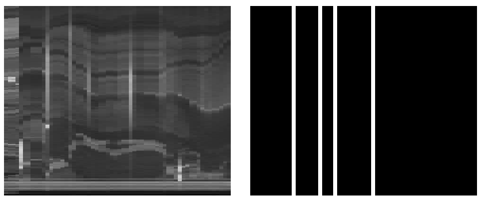


FIG. 11: Average computation: result of the average of the elements for each column.

Max-tree filtering The notions of peak, height and basis area can be precisely defined thanks to a data structure called *max-tree* [4] (refer to this paper for more details). The parameter h influences the sensitivity of the method and has a role similar to the threshold in Sec. 2. In Fig. 12, we illustrate the result of the max-tree filtering.

FIG. 12: Result of the max-tree filtering.

Detection of flashes In this work, we consider that the maximum flash duration is 5 frames, i.e. $S = 5$. Thus, for each flash candidate we need to verify if the size is smaller than S. In Fig. 13, we illustrate the flashes detected, that are represented by white vertical bars.



(a) (b)

FIG. 13: Detection of flashes in which the flashes are represented by white vertical column bars: (a) original image with 4 flashes and (b) result of the flash detection.

Differently from the first method, this one does not identify the center, but the problem is quite the same, mainly when there is high luminosity change that does not represent a true flash. Regarding other methods, this one identifies very well all types of flashes thanks to the data structure used to model the peak values.

	Videos	Flashes	Frames	Frames/event (mean)
Corpus	10	23	8392	3

TAB. 1: Features of the videos which were selected for the experiments.

	μ	E_m	R_f	γ
Top-hat	0.05	**0.61**	0.26	0.56
Max tree	**0.11**	0.67	**0.43**	**0.69**

TAB. 2: Quality measures for flash detection: robustness $\mu(0.40, 0.30)$, missless $E_m(0.05)$, falseless $R_f(0.01)$ and gamma measure γ.

4 Experiments

In these experiments, we apply the methods described in Sec. 2 and in Sec. 3 to the Corpus illustrated in Table 1. To compare the methods, we consider the basic quality measures, like, recall, precision, error, and some other quality measures that are proposed by Guimarães like, missless error, falseless recall, gamma measure and robustness (refer to [3, 2] for more details).

In Fig. 14, we illustrate some experimental results. The quality measures of the filtering of the max tree and the top-hat filtering are outlined in Table 2. In Table 3, we illustrate the basic quality measures when we consider the values associated with the gamma measure.

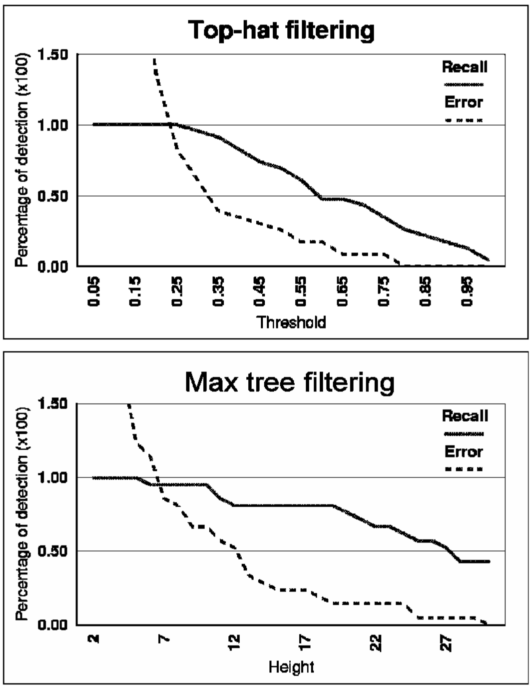


FIG. 14: Experimental results for flash detection.

5 Conclusions and discussions

In this work, we proposed two methods for flash detection, in which the extraction of peaks by max tree analysis allows a detection of flashes in all positions of the shot. We computed the quality measures for flash detection, and according to these measures, the method based on max tree analysis generally presents the best results.

Exp	Flash	Recall	Precision	Error	Threshold
Top-hat	23	91%	68%	43%	35%
Max-tree	23	87%	87%	13%	18

TAB. 3: Basic quality measures related to the gamma measure.

Acknowledgements

The authors are grateful to FAPEMIG, CAPES/COFECUB, CNPq 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] S. J. F. Guimares. *Video transition identification based on 2D image analysis*. PhD thesis, UFMG - Brasil, March 2003.

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

[4] P. Salembier, A. Oliveras, and L. Garrido. Anti-extensive connected operators for image and sequence processing. *IEEE Trans. on Image Processing*, 7(4):555–570, 1998.

[5] J. Serra. *Image Analysis and Mathematical Morphology: Theoretical Advances*, volume 2. Academic Press, 1988.

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

[7] D. Zhang, W. Qi, and H. J. Zhang. A new shot boundary detection algorithm. *Lectures notes in Computer Science*, 2195:63–70, 2001.