Real Time Image Processing: Shot detection

Guillaume Lemaître - Mojdeh Rastgoo - Warakorn Gulyanon
Heriot-Watt University, Universitat de Girona, Université de Bourgogne
g.lemaitre58@gmail.com - mojdeh.rastgoo@gmail.com - tonghoho@gmail.com

I. Introduction

Automatic shot transition detection is an important field of research. Shot transition detection is the first step to simplify any processing on video. Shots are considering as units of a video. They give information about time. For instance, in a scene interpretation algorithm, it will be difficult to give any interpretation with a unique sequence constituted of two shots decorrelate. However, if the sequence is split in two different shots, it will be easier to interpret independently both shots.

In this paper, we will present an overview of the shot transition detection problem. In the section II, we will define the basic technical terms. In the following section III, we will present the different types of transitions and illustrated them. In the section IV, we will detail methods which permit to score and detect transitions. Finally, we will conclude in the section V by a presentation of three different applications.

II. Definition of vocabulary

In this part, we will define several basic aspect used in shot detection:

- Video structure definition.
- Shot definition.
- Scene definition.

A. Video structure definition

Figure 1 represents the structure of a video. A video is composed of an image serie taken at constant interval short (between 25 and 30 frames per seconds). The video can be cut in series of shot. These different shots constitute some scenes and these scenes constitute some sequences. We will define in the next part what is a shot and a section and sequence.

B. Shot definition

A shot is by definition an unbroken sequence of frames from only one camera. Figure 2 presents an example of shot.

On this sequence of image, only one camera takes the sequence.

C. Scene definition

A scene is collection of one or more shots *focusing* on objects of interest. Figure 3 presents an example of a scene.

On this sequence of image, we can see that you have two different cameras that are focus on the same person. Basically, this scene is composed of two different shots. The first shot is composed of the figures 3(a) and 3(b) while the second shot is composed of the figures 3(c), 3(d), 3(e) and 3(f).

D. Sequence definition

A sequence is defined as a collection of several scenes. Figure 4 presents an example of a sequence.

On this sequence of image, we can see that the sequence is composed of three different scenes. The first scene is composed of the figure 4(a) and 4(b) while the

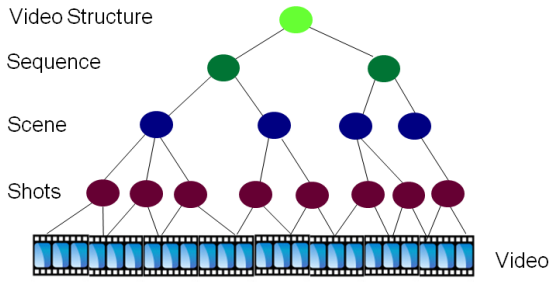
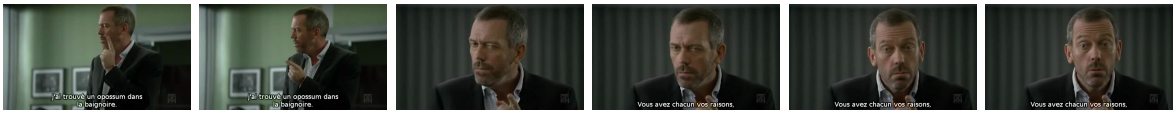


Figure 1. Video structure

(a) shot 1 (b) shot 2 (c) shot 3 (d) shot 4 (e) shot 5 (f) shot 6

Figure 2. Example of a sequence constituting a shot



(a) scene 1 (b) scene 2 (c) scene 3 (d) scene 4 (e) scene 5 (f) scene 6

Figure 3. Example of a sequence constituting a scene

(a) sequence 1 (b) sequence 2 (c) sequence 3 (d) sequence 4 (e) sequence 5 (f) sequence 6

Figure 4. Example of a sequence constituting a sequence

second scene is composed of the figures 4(c) and 4(d) and the third scene is composed of the figures 4(e) and 4(f).

III. Shot transition classification

Shots transition can be split in two groups:

- Hard cut.
- Gradual transition.

A. Hard cut

A hard cut is an instantaneous transition from one shot to the next one. Figure 5 presents an example of hard cut.

In this figure, the hard cut transition is between the frames 5(b) and 5(c). The transition between both frames is abrupt.

B. Gradual transition

Instead of hard cut which is an abrupt transition, gradual transitions are smooth transition between two shots. It exists four different types of gradual transitions:

- Fade-in.
- Fade-out.
- Dissolve.
- Wipe.

1) Fade - in: A fade-in is a gradual transition from a constant image generally black to the frame wanted. Figure 6 presents an example of fade-in.

On this sequence of images, the brightness increases. At the beginning, frames 6(a), 6(b) and 6(c), the luminosity is weak and the image is black. Then, the intensity of the light increase and the interest scene start to be perceived, frames 6(d) and 6(e). Finally, the scene is explicit on the frame 6(f).

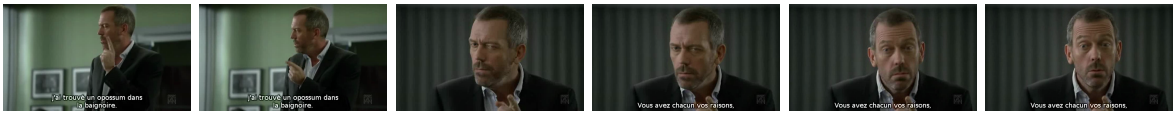
2) Fade - out: A fade-out is the opposite process to a fade-in. It is a gradual transition from a frame wanted to a constant image generally black. Figure 7 presents an example of fade-out.

On this sequence of images, the brightness decreases. The scene is explicit on the frame 7(a). Then, the intensity of the light decrease gradually from the frame 7(b), 7(c), 7(d) and 7(e). Finally, on the frame 7(f), a black constant image appears.

3) Dissolve: A dissolve is defined by a succession of a fade-out followed by a fade-in. Figure 8 presents an example of a dissolve.

On the sequence of images, the figure 8(d) shows a decreasing intensity of the previous scene presenting on the frames 8(a), 8(b) and 8(c) and in the same time an increasing intensity of the next scene presenting frames 8(e) and 8(f).

4) Wipe: A wipe is a transition where the previous shot is replace by the new shot using a regular pattern



(a) hard cut 1 (b) hard cut 2 (c) hard cut 3 (d) hard cut 4 (e) hard cut 5 (f) hard cut 6

Figure 5. Example of hard cut

(a) fade-in 1 (b) fade-in 2 (c) fade-in 3 (d) fade-in 4 (e) fade-in 5 (f) fade-in 6

Figure 6. Example of fade-in

(a) fade-out 1 (b) fade-out 2 (c) fade-out 3 (d) fade-out 4 (e) fade-out 5 (f) fade-out 6

Figure 7. Example of fade-out

(a) dissolve 1 (b) dissolve 2 (c) dissolve 3 (d) dissolve 4 (e) dissolve 5 (f) dissolve 6

Figure 8. Example of dissolve

as a line. Figure 9 presents an example of wipe

The frame 9(a) presents the previous shot while frame 9(f). The image presented in figure 9(f) appears gradually, substituting pixel of the frame 9(a) following an horizontal line.

IV. Methods

In this section, we will present methods which allow to evaluate the dissimilarities between two images and methods which allow to take a decision if the transition is a shot detection or not.

A. Scoring

The scoring methods are methods which allow to compute the dissimilarities between two images. Several methods exist:

- Uncompressed features
  - Pixel differences
  - Statistical differences
  - Histogram differences
  - Edge tracking
- Compressed features
  - Compression differences
  - Motion vectors

In this section, we will present only histogram and edge tracking methods.

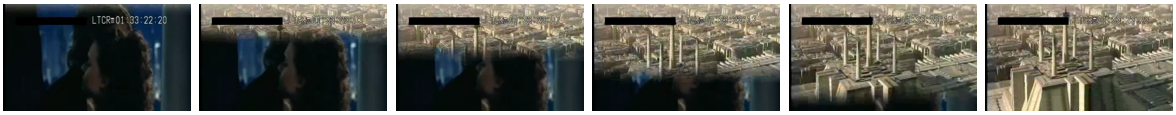
1) Histogram differences: Histograms methods is one of the more used to find a score between two images. Figure 10 presents two images belong to the same shot so with a lot of similarities while figure 11 presents two images belong to two different shot with a lot of dissimilarities.

The score is computing using the following formula:

$$d(H_f, H_g) = \sqrt{\sum_{i=0}^{255}(H_f(i) - H_g(i))^2} \quad (1)$$

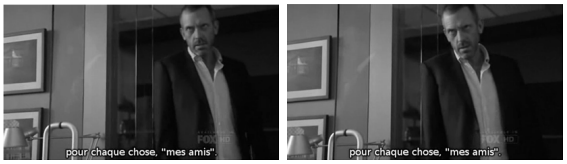
where H_f and H_g is the histograms of two consecutive frames from a video.

The basic idea of this method is the following:

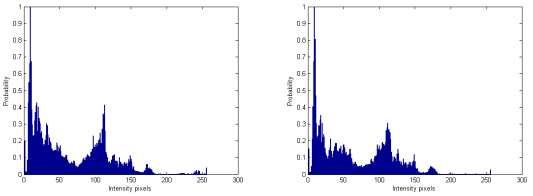


(a) wipe 1 (b) wipe 2 (c) wipe 3 (d) wipe 4 (e) wipe 5 (f) wipe 6

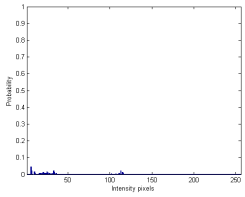
Figure 9. Example of wipe



(a) Image at $t-1$ (b) Image at t



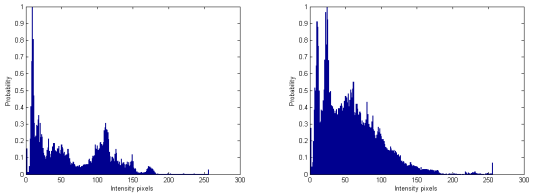
(c) Histogram of image at $t-1$ (d) Histogram of image at t



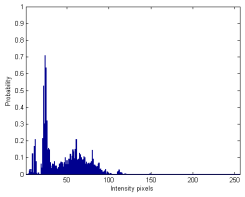
(e) Square difference of histigrams

Figure 10. Example of histograms methods with similarities

(a) Image at $t-1$ (b) Image at t



(c) Histogram of image at $t-1$ (d) Histogram of image at t



(e) Square difference of histigrams

Figure 11. Example of histograms methods with dissimilarities

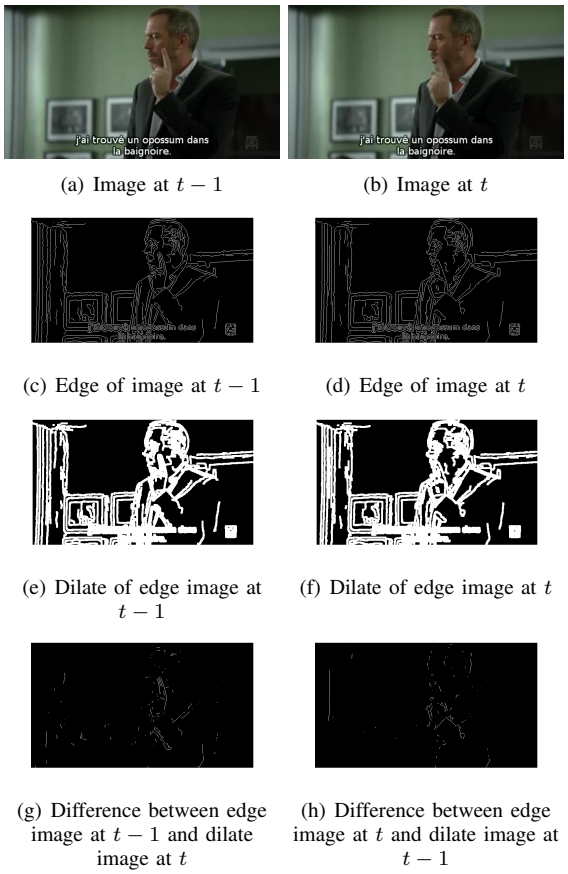
- Compute histograms of both images as shown in figures 10(c) and 10(d) for the example presented on the figure 10 and figures 11(b) and 11(c) for the example presented on the figure 11.
- Compute the bin-wise histogram which is the square of the difference of the two previous histograms computed before. Figure 10(e) represents the bin-wise histogram for the example shown in the figure 10. Figure 11(d) represents the bin-wise histogram for the example shown in the figure 11.
- Then, the square root of the sum of the bin-wise histogram is computed and give the score. For the example figure 10, the score computed is equal to 0.6155. This score is weak and represents a big similitude between both images. For the example figure 11, the score computed is equal to 3.0167. This score is important and represents a big dissimilitude between both images.

2) Edge tracking: The second method to compute the score representing the dissimilarities is based on edge tracking. The aim is to find edges two consecutives frames and compute how the images are different. We present to example where figure 12 represents an example where both images belong to the same shop and figure 13 represents an example where both images belong to two different shots.

The score can be computed using the formula as follow:

$$ECR_n = \max(\frac{X_n^{in}}{\delta_n}, \frac{X_n^{out}}{\delta_{n-1}}) \quad (2)$$

where X_n^{in} and X_n^{out} are the number of edge pixels



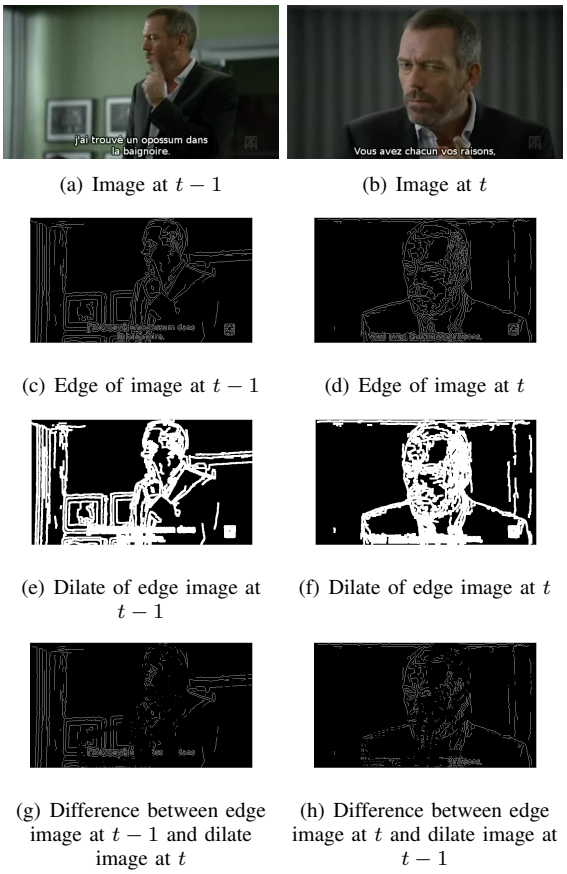
(a) Image at $t-1$ (b) Image at t

(c) Edge of image at $t-1$ (d) Edge of image at t

(e) Dilate of edge image at $t-1$ (f) Dilate of edge image at t

(g) Difference between edge image at $t-1$ and dilate image at t (h) Difference between edge image at t and dilate image at $t-1$

Figure 12. Example of edge similarities



(a) Image at $t-1$ (b) Image at t

(c) Edge of image at $t-1$ (d) Edge of image at t

(e) Dilate of edge image at $t-1$ (f) Dilate of edge image at t

(g) Difference between edge image at $t-1$ and dilate image at t (h) Difference between edge image at t and dilate image at $t-1$

Figure 13. Example of edge dissimilarities

entering and exiting and δ_n and δ_{n-1} are the number of edge pixels in the frame at t and $t-1$.

The implementation of this equation can be as follow:

- Detect edges using Canny detector for instance as shown on figures 12(d), 12(e), 13(c) and 13(d).
- Apply a dilatation on Canny image as shown on figures 12(f), 12(g), 13(e) and 13(f).
- Compute the difference of the edge image at time t and the dilate image at time $t-1$ and the difference of the edge image at time $t-1$ and the dilate image at time t as shown on figures 12(h), 13(a), 13(g) and 13(h)
- Compute the ratio of the number of edge pixels of the difference image over the number of edge pixels. For the example shown on the figure 12, the score computed is 0.0858 and for the example on the figure 13, the score computed is 0.6456

B. Decision

Two methods can be used to decide if the transition can be considered as a shot detection:

- Simple threshold.
- Adpative threshold.

1) Simple threshold: The simple threshold method is just to compare the score with a threshold. If the score is bigger than a threshold, we consider that we detect a transition of shot. Otherwise, the sequence is considered in the same shot.

2) Adaptive threshold: An adaptive threshold is a threshold which changes depending of position in the sequence of image. A method to detect shot detection was to satisfy the two following conditions:

- The middle sample is the maximum in the window.
- The middle sample is greater than $\max(\mu_{left} + T_d\sqrt{\sigma_{left}}, \mu_{right} + T_d\sqrt{\sigma_{right}})$ where T_d is given a value of 5.

V. Real time applications

A. A method for fast shot boundary detection based on SVM

Xue Ling, Ouyang Yuanxin, Li huan , and Xiong Zhang[1] purposed a new shot detection algorithm using SVM. The main flow of their algorithm is extracting features and feeding them to SVM. The features that they use is based on uncompress data which some of them is expensive computation; however, they purposed a way to reduce the number of frame of the video before extracting features. Therefore, the normal method that is not supposed to be real time can reach 36 frames per second which passes the real-time timing. Here is the list of main step of this method.

- Cut the smooth interval
- Extract video features
- Detect hard cut using SVM
- Detect gradual change using temporal multi-resolution analysis

1) Cut the smooth interval: The purpose of cutting the smooth interval is to reduce the computation time. The basic element that used to define the smooth interval is the intensity variance. The result video sequence from cutting is call Reordered Frame Sequence (RFS). The procedure of extract RFS is shown below.

- Extract the intensity variance from each frame
- Calculate the difference between the two consecutive variances according to the following formula

$$D(t) = |\vartheta^2(t) - \vartheta^2(t+1)| \tag{3}$$

where $\vartheta^2(t)$ is the intensity variance sequence, $D(t)$ represents the sequence of the difference between the two consecutive variances. According to the figure 14, we can find that $D(t)$ has a peak value where the abrupt cut happens and a series of local maximum value where gradual change is occurring. In order to detect the smooth interval, the threshold is created which the different of intensity variance is less than threshold; it is considered as the smooth interval.

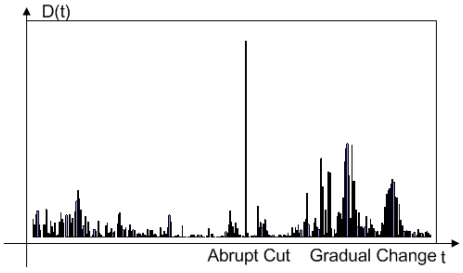


Figure 14. The intensity variance difference

- Detect short local minimum sequences in the gradual transitions. The gradual change such as dissolve can be mistaken eliminate as the smooth interval. The variance curve of dissolve as in figure 15 shows a clear parabolic shape which this unique character is used to detect local minimum sequences. The basic idea is to compare the maximum variance in the local minimum sequence with the two peak points at the start and end of the gradual change.

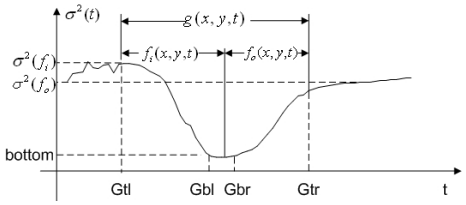


Figure 15. Intensity variance of dissolve change

2) Feature extraction: They used three features in their research which are

- Intensity Pixel-wise Difference
- Color Histogram Differences in HSV Space
- Edge histogram differences.

These features are calculated from RFS. The special configuration of these three features is the interval of two frames that used to find the difference.

3) Hard cut detection: The support vector machine (SVM) is a supervised learning technique from the field of machine learning Application. The SVM performs classification by non-linear mapping the input space into a very high dimensional feature space and constructing an optimal separating hyperplane in this space
In their research, the radial basis function (RBF) is used as the kernel as shown in equation 4.

$$K(x_i, x_j) = exp(-\gamma||x_i, x_j||^2), \gamma > 0 \tag{4}$$

where x_i and x_j are training vectors, γ is the kernel parameter. For detecting a hard cut, a five-dimensional normalized feature vector ($D_P^1(t)$, $D_H^1(t)$, $D_S^1(t)$, $D_X^1(t)$, $D_Y^1(t)$) is extracted from two consecutive frames in the RFS and used as the input vector to the SVM.

4) Gradual Change Detection: The gradual change often occurs over several frames; therefore, a high resolution cannot detect the gradual change. In order to solve this problem, a lower resolution is implemented.

This method is so called the temporal multi-resolution analysis.

- Eliminate the detected cuts from RFS.
- Reduce the sampling rate in the RFS and follow the feature extraction procedure to extract the features of $D_P^2(t)$, $D_H^2(t)$,$D_S^2(t)$, $D_X^2(t)$, $D_Y^2(t)$, $D_P^3(t)$, $D_H^3(t)$, $D_S^3(t)$, $D_X^3(t)$, $D_Y^3(t)$.
- The feature vectors at different resolutions are concatenated and form a single feature vector to serve as the input vector to the SVM. Then the SVM detect whether the candidate is a gradual change or not.

B. An Adaptive Shot Change Detection Algorithm and Its Implementation on Portable Multimedia Player

Won-Hee Kim and Jong-Nam Kim[2] proposed algorithm consists of the following elements: sub-sampling, weighting variance, and adaptive thresholds. The experiments obtained the detection rate of about 94.4
In order to match the performance of real-time in PMP, computational complexity can be reduced with sub-sampling in spatial domain. In actual PMP player software, coded video stream cannot be manipulated with detail because of encapsulation property of the software. Therefore, the shot detection algorithm is implemented in spatial domain with decoded video stream. The basic idea is that mean and variance have a higher detection rate than pixel based methods, a detection rate for sub-sampled video higher than histogram based ones, and a computational reduction greater than block based methods. For adaptive thresholds, the mean values of weighting variance is used as a feature reference frames

1) Weighting variance as a comparison feature: The difference of mean and variance of two consecutive frame can be calculated as in the equation 5 and 6 below.

$$m = \frac{\sum_{x=0}^{W-1}\sum_{y=0}^{H-1}(f_i(x,y) - f_{i+1}(x,y))}{W \times H} \tag{5}$$

$$v = \frac{\sum_{x=0}^{W-1}\sum_{y=0}^{H-1}(f_{i+1}(x,y) - m)^2}{W \times H} \tag{6}$$

In the equations, m is the mean of frame difference and v means variance of the next frame. f_i is a current frame, and f_{i+1} the subsequent frame of f_i. W and H are the horizontal and vertical size of a frame. The weighting variance can be calculated by dividing the scaled W and H as shown in equation 7, 8.

$$m' = \frac{\sum_{x=0}^{W-1}\sum_{y=0}^{H-1}(f_i(x,y) - f_{i+1}(x,y))}{(W/w_d) \times (H/w_d)} \tag{7}$$

$$v' = \frac{\sum_{x=0}^{W-1}\sum_{y=0}^{H-1}(f_{i+1}(x,y) - m')^2}{(W/w_d) \times (H/w_d)} \tag{8}$$

As a result, m' is the same as $w_d^2 \times m$. Also v' is much larger than v. As shown in equation 8. v' is not increased linearly, however it is increased abruptly more than ten times on shot change frames as shown in figure 16. Figure 16 represents variance of frames, in which

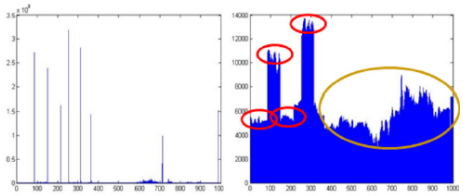


Figure 16. Variance v, v'

(a) shows variance of frame with weighting factors and (b) variance of frame without weighting factors. With this weight variance, the shot boundary can be detected easily. The proper w_d also has to be calibrated properly.

2) Setting adaptive thresholds: The purposed method of adaptive thresholds is based on these conditions. Firstly, The method uses the statistical information of a video sequence to apply the adaptive thresholds obtained to various video sequences. Secondly, only previous frames are used from a current frame to get adaptive thresholds. Third, the proposed method uses the results already calculated to reduce computational time. The adaptive threshold setting scheme satisfies the above conditions by using the mean of weighting variance to set thresholds as shown in equation 9

$$TH = \frac{\sum_{i=s_p+1}^{s_c-1} v'(i)}{s_c - s_p + 1} \times w_t \tag{9}$$

Here, s_p means the number of the most recent shot change frame, s_c the number of a current frame, and w_t a weighting factor. From equation 9, the adaptive thresholds is calculated from the video frames of some temporal range, which are previous frames. The current frame is regarded as a shot change frame when variance, $v'(s_c)$ of a current frame is larger than TH. By

multiplying weighting wt additionally, the threshold is prevented from being too low which the suitable w_i is around 10-15.

C. Real Time Efficient Detection of Shot Changes in MPEG soccer Video Using Macro - Block Type information

This paper was published by L.Yin, H.Morita, I.Baskara and L.Zhang in 2009. Due to the increasing demand of video services in the field of efficient detection approaches ,they introduced the efficient method for real time analysis of the soccer videos.The introduced method is based on macro - blocks information. The algorithm uses the statistics and the distribution of the macro blocks types in every B and P frame in order to detect the shot changes in real time. The result of the algorithm shows that the introduced method has high level of recalls and precisions. This part of the report will present a summary on this paper, the algorithm, result, brief introduction of video compression and Defining the macro blocks.

1) Video Compression: Video is a three dimensional array. Two dimensions represent the spatial dimensions of the moving pictures and one dimension shows the time domain. All the pixels that corresponds to one time moment are known as one frame data. Video compression refers to reducing the amount of the data which is used in the digital video. Its basically the combination of the spatial image compression and temporal motion compression. The video frames can be compressed using different algorithm. These algorithms have their advantage and disadvantages, they will be different mostly in terms of the amount of data they can compress.These algorithms are named as picture type and frame type. Three main algorithm are :

- I frame type : They are the least compressible frames , but they do not need other video frames for decodin.Since they are depending on the current frame to compress the data
- P frame type: This type can use the previous data as a reference to decompress the data. Due to this reason it has the ability to compress the data more
- B frame type: This type can used both, the previous video frames and the feature video frames as a reference in order to decompress the data , and has the highest ability of compression.

There are different standards in order to compress the data , In this paper they used MPEG2 encoder to compress the information. In General the MPEG group of frames (compressed frames) will have the following form,$IBBPBBPBBPBBPBB$. As it mentioned before in this algorithm they are interested in the amount of information of macro-blocks (MB) in P and B frame , for that reason they consider they consider the $PBBP$ consecutive frames as P_f as front P , B_f as front B, B_r as rear B and P_r as the rear P respectively.

2) Macro -Blocks (MB): Using the MPEG-2 standard video compression each frame will be divided in to 16 by 16 blocks. These blocks are called macro blocks. Based on the MPEG standard macro blocks can be divided in to 4 mode (4 types)

- I MB type: Intra prediction , I mode.
- F MB type: Forward prediction, F mode.
- B MB type: Backward prediction, B mode
- BI MB type: Bidirectional prediction, BI mode

The frame types are different in terms of the number of MB mode they have , and this feature will make a difference between them. It means they can be differentiated based on the number and the mode of MB they have. In I frame type there are only I mood MB. P frame type has both the I mode MB and F mode MB, and the B frame type has all four types of MB mode. In this research they detect the shot changes based on the difference on these numbers in each B and P frame.

3) Patterns of Macro-blocks type in Abrupt Transition: Abrupt transition or cut transition as explained in the previous section is defined when the content of the image suddenly changes between the two consecutive frames. In this paper they divided these changes in to three categories, scene changes before the *P*frame or *I* frame, scene changes before B_f *frame(front B)* and scene changes before B_r *frame (rear B)*. These classification is based on the fact that when shot occurs the number of the I MB mode in B and P frames changes. In the introduced algorithm , in order to analyse the number of MB mode in P and B frames in related to the shot changes , they consider different cases of having the shot in the consecutive frames and they divide the cut shots in to five categories due to the reason that each type of cut shots has individual static and distribution of MB modes in P and B frames. first,scene change occurs before P frame or I frame(SCBPI). Second,scene change occurs before front B BB_fframe(SCBFB). Third,scene change occurs before rear B B_r frame(SCBRB),fourth,scene change

occurs at front B B_f frame (SCAFB) and fifth,scene change occurs at rear B B_r frame(SCARB).

In the first case of the abrupt transition the largest number of the MB mode in the P_r frame are I mode since these frame represent the first frame of the new shot, and most of the MB modes in the two consecutive B frames are F MB mode since they have more similarity to I modes in the P_f frame rather than P_r frame.

In the second case of the cut (SCBFB) most of the MB modes in the P_r frame are I MB mode again since its the first reference frame of the new shot. and then the two consecutive B frames contains mostly B MB modes. This is due to the image changes between the P_f and the next two B frame.

In the third case (SCBRB), the scene change is happening before the rear B(B_r). most of the MB mode in P_f are I mode , and B_f contains mostly F MB modes and most of the MB frames in B_r frame are B modes.

In the fourth case(SCAFB), the cut transition is happening at the B_f frame. Due to that reason, a lot of I MB mode exist in the B_f frame and B_r contains a lot of B MB Modes . Again since a B_f frame contains the information of the new shot and previous shot it also contains the BI MB modes.

In the fifth case (SCARB), the shot change occurs at the B_r frame, so these frame will have some BI MB modes since again its contains the previous and new shot information. and because of the new shot it contains a lof of I MB Modes. Because the change occurs in B_r the B_f frame contains a lot of F MB Modes.

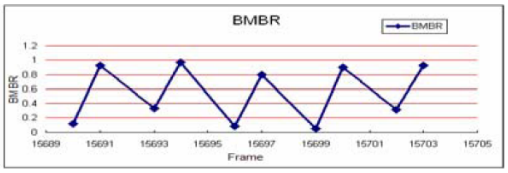
Based on the mentioned analysis on the number of the MBs in B and P frame they classify again the frames in to 7 types. Figure 17 shows this classification. With reference to the pattern of the frame they conclude the existence of the abrupt shot. If the pattern belongs to one of the following forms: 220 , 240, 270, 440, 740.

4) Patterns of Macro-blocks type in Dissolve Transition: As it is mentioned before the dissolve transition is a change from one shot to another one with linear or non linear decreasing the intensity of the first shot and increasing of the second shot. Through the done research by the authors of the paper, they faced the problems that some of the dissolves transitions occurs only on one frame in the soccer video and in the other cases the dissolve transition may lasts several frames. In this case they define to kind of dissolve, short dissolve for detecting the dissolve over one frame and long dissolve in the other case. In the case of the short dissolve the

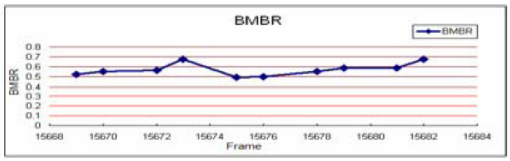
Frame type	Conditions	
0	Number of I mode MB occupies more than 10% in P frame	
1	Number of I mode MB is the largest in B frame	
2	Number of F mode MB is the largest in B frame	number of I mode MB occupies less than 27%
3		number of I mode MB occupies more than 27%
4	Number of B mode MB is the largest in B frame	number of I mode MB occupies less than 27%
5		number of I mode MB occupies more than 27%
6	Number of BI mode MB is the largest in B frame	number of I mode MB occupies less than 27%
7		number of I mode MB occupies more than 27%

Figure 17. Classification of the frames in to 7 type in order to obtain the abrupt transition

dominant number of the I MB modes in B or P frame was used as a prove. In the case of the long dissolve they used the definition of the Backward Macro Block Ratio in order to define the characteristics of the MB modes in the B and P frame. These is due to the fact that this factor changes a lot during the dissolve transition in compare to the other times, and can be used as a good parameter to describe the dissolve transition. Figure 18 shows the behaviour of this factor during the dissolve shot and Non dissolve regions.



(a)Dissolve region



(b)Non-dissolve region

Figure 18. behaviour of the Backward Macro Block Ratio (BMBR)during the dissolve transition and Non Dissolve transition

The Backward Macro Block Ratio (BMBR) is computed using the following expression. The num_f is the number of the F MB mode in one frame and num_b is the number of B MB mode in one frame.

$$BMBR = \begin{cases} \frac{num_b}{num_b + num_f} & \text{if } num_b + num_f \neq 0 \\ 0 & \text{otherwise} \end{cases}$$

By using the above expression they defined the dissolve transition if the BI MB mode in the B frame are dominant and the value of the BMBR is smaller than the upper threshold in the front B (B_f) frame and larger than the lower threshold in the rear B (B_r). If more than three consecutive pairs of B frame meets the mentioned conditions It means that they have long dissolve.

5) result and evaluation: The authors of this algorithm tested their algorithm over two video , each 100 minutes. In order to evaluate the performance of the algorithm they introduce two variable recall and precision. Recall is the ratio of the correct detection of the shots over the total number of the correct detection and missed shots. Precision is the ration of the correct detected shots to the total number of the correct detection and false detection. Figure 19 shows the evaluation of their algorithm. As it could be seen the performance is acceptable and since their processing time was obtained as $0.3ms$ it could be conclude that their system is operating in the real time.

		Recall	Precision	Time
First Video	Hard Cut	92.66%	95.91%	100 min
	Dissolve	87.69%	82.60%	
Second Video	Hard Cut	93.15%	97.61%	100 min
	Dissolve	87.80%	81.82%	

Figure 19. Evaluation of the algorithm performance

VI. Conclusion

In this paper, we presented an overview of the shot transition detection problem. We defined in the section II the basic technical terms. Then, we presented the different types of transitions and illustrated them in the section III. In the section IV, we detailed methods which allow to score and detect transitions. We concluded in the section V by a presentation of three different applications.

References

[1] X. L. O. Y. L. Huan and X. Zhang, "A method for fast shot boundary detection based on svm," *Proceedings of the 1st International Congress on Image and Signal Processing (CISP '08)*, vol. 1, pp. 445–449, 2008.

[2] W.-H. Kim, Y.-J. Jeong, K.-S. Moon, and J.-N. Kim, "Adaptive shot change detection technique for real-time operation on pmp," *Convergence Information Technology, International Conference on*, vol. 1, pp. 295–298, 2008.

[3] L. Yin, H. Morita, I. B. Nugraha, and L. Zhang, "Real-time efficient detection of shot changes in mpeg soccer video using macro-block type information," *Proceedings of IC-NIDC, IEEE*, vol. 2, pp. 707–711, 2009.