IMPROVED ALGORITHM FOR VIDEO SHOT DETECTION

David Asatryan, Manuk Zakaryan

***Abstract**: Currently digital video are widely used in various fields of science and technology and in human daily activities. Intensively growing and already existing huge amount of digital video data need to be managed, so the shot boundary detection is the first and important step for content-based video retrieval and indexing. Each algorithm aimed for this approach should accurately detect boundaries between camera shots and do a segmentation of a video. In this paper a new method of abrupt transitions detection is proposed, based on Weibull distribution of each frame's gradient magnitude. Experimental results successfully show that it can effectively detect hard cuts and has certain advantages against widely used other methods, which are using image pixels point-by-point comparison method.*

***Keywords**: segmentation, cut detection, similarity measure, Weibull distribution.*

***ACM Classification Keywords**: Image Processing and Computer Vision*

Introduction

Internet, telecommunications and digital television made explosive rate of video recordings. With such enormous video data resources there are rising the issues for the efficient browsing, searching and retrieval of digital videos. However, the traditional video indexing method, which uses human beings to manually annotate or tag videos with text keywords, is time-consuming and not efficient.

For video data to be useful, its content must be represented so that it can be stored, queried, and displayed in response to user information needs. The process of shot detection is a fundamental component in automatic video indexing, editing, archiving and browsing [Hanjalic, 2002]. Any approach to indexing and archiving video for retrieval requires parsing the video and extracting key frames to generate an indexed database. A typical video indexing technique is to segment a video sequence into shots and then select representative key-frames [Fernando, 2000]. A shot is defined as an unbroken sequence of frames from a single camera. In a video sequence there can be a number of different types of transitions between shots, such as a cut (abrupt shot change between two frames) or gradual transitions such as fades, dissolves and wipes [Borecsky, 1996]. Here, we focus only on cut detection problem, using a new approach to determine the similarity or dissimilarity of the sequential frames.

The major techniques that have been used for shot boundary detection are pixel differences, statistical differences, histogram comparisons, edge differences, compression differences, and motion vectors [Borecsky, 1996].

Pixel Differences: The easiest way to detect if two frames are significantly different is to count the number of pixels that change in value more than some threshold. This total is compared against a second threshold to determine if a shot boundary has been found. This method is sensitive to camera motion and somewhat slow [Borecsky, 1996].

Statistical Differences: Statistical methods expand on the idea of pixel differences by breaking the images into regions and comparing statistical measures of the pixels in those regions. This method is reasonably tolerant of noise, but is slow due the complexity of the statistical formulas.

Histograms: Histograms are the most common method used to detect shot boundaries. The simplest histogram method computes gray level or color histograms of the two images. If the bin-wise difference between the two histograms exceeds a threshold, a shot boundary is assumed.

Compression Differences: It used differences in the discrete cosine transform (DCT) coefficients of JPEG compressed frames as their measure of frame similarity, thus avoiding the need to decompress the frames.

Edge Tracking: The method aligned consecutive frames to reduce the effects of camera motion and compared the number and position of edges in the edge detected images. The percentage of edges that enter and exit between the two frames is computed. Shot boundaries were detected by looking for large edge change percentages. This is more accurate at detecting cuts than histograms and much less sensitive to motion than chromatic scaling.

Motion Vectors: Method uses motion vectors determined from block matching to detect whether or not a shot was due to zoom or a pan. Because shots with camera motion can be incorrectly classified as gradual transitions, detecting zooms and pans increases the accuracy of a shot boundary detection algorithm.

The most of existing methods for video cut detection use some inter-frame difference metric. In frame pair where this difference is greater than a predefined threshold is deemed to be a shot boundary or cut location. Probably the simplest of these methods is based on pair-wise pixel comparison, and the widely used modification of them is the mean-squared error (MSE). Despite of many advantages of MSE-based methods there are certain problems of image processing technique, where application of MSE measure for two images dissimilarity assessment comes to conflict with the Human Visual System (HVS) perception [Smoliar, 1994] . HVS, in contrast to MSE, is less sensitive to camera or object motion because it is extracting basically the structural and intentional but not pixel-by-pixel information from an image. Last two decades bring many new metrics for images similarity assessment, beginning from [Wang, 2002 - Wang, 2009], which give the results more coherent with HVS perception.

As it is mentioned above the shot boundary detection algorithms mainly uses similarity or dissimilarity measures between consequent frames. The comparison value is called a threshold, and it is one of important elements in the shot change detection algorithm. We can divide threshold setting into two groups: the fixed threshold method and the adaptive threshold method [Zhi, 2005].

The fixed threshold method determines optimal thresholds from repeated experiments. However, they require much experimental iteration and must find other optimal threshold for other video sequences. Most of them iterate adjustment of thresholds until they get the best results. These methods may have long processing time. In general, variation of thresholds is relatively large to use a fixed threshold for all video sequences. Thus, some algorithms for shot detection were improved by analyzing the whole video sequences for setting multiple thresholds instead of a fixed threshold [Cheng, 2002]. These methods may also have long processing time. Thus, it is difficult to apply them to actual applications requiring real-time operations. Meanwhile, the adaptive threshold based segmentation algorithms get sub-optimal threshold according to [Kim, 2009].

This paper introduces a novel video cut detection technique using the similarity measure [Asatryan, 2009] based on comparing the structural properties of images. The approach proposed in [Asatryan, 2010] successfully applied to some problems of image registration even the images are rotated or scaled.

Shot Detection Algorithm

Let's consider a sequence of frames $f_1, f_2, ..., f_i, ...$, where f_1 is the first frame of video sequence or the first frame which follows the previous cut.

The simplest decision rule for cut detection is based on consecutive comparison of contiguous frames using certain similarity measure. When the level of similarity measure exceeds some predefined threshold t_c, then the corresponding frame is considered as a cut frame. Of course, there are other decision rules which use the information of previous frames [Cheng, 2002] or use a few subsequent frames before decision making [Kim, 2009]. However, in any case the quality of a decision rule depends on used similarity measure. In this paper, for simplicity, we choose the first type of cut detection algorithm and compare it with the new measure based on structural properties of an image.

For simplicity we consider the Gray Scale (8 bit) format image $\mathrm{I} = \{\mathrm{I}(m,n)\}$ with $\mathrm{M} \times \mathrm{N}$ sizes, $m = 0,1,...,\mathrm{M}$; $n = 0,1,...,\mathrm{N}$.

The standard algorithm of inter-frame comparison by using a similarity measure μ can be presented as follows

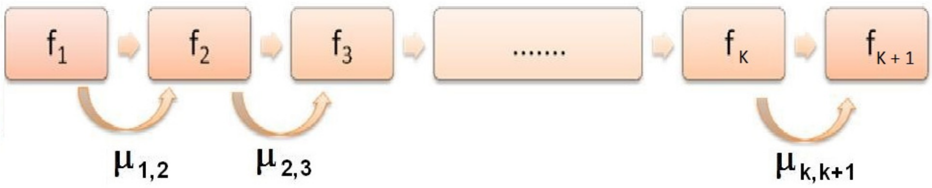


Fig.1 Algorithm of inter-frame comparison

Let consider the frames $f_1, f_2, ..., f_k, f_{k+1}$ and the sequence of corresponding values of the similarity measure $\mu_{1,2}, \mu_{2,3}, ..., \mu_{k,k+1}$. When $\mu_{i,i+1} \le t_c$ for $i = 1,2,...,k-1$ and $\mu_{k,k+1} > t_c$ then point k is assumed as a cut point.

To demonstrate the advantages of shot detection algorithm with using new similarity measure we have to choice some popular, simple and interpretable measure for inter-frame similarity estimation. As it is noted in the Section 1 the simplest measure is MSE-based similarity measure PSNR, which is done according the formula (1) as follows [Fernando, 2000]

$$\mathrm{PSNR} = 10\log_{10}\frac{\max\limits_{m,n}\left|\mathrm{I}_1(m,n)-\mathrm{I}_2(m,n)\right|^2}{\mathrm{MSE}^2}, \quad MSE^2 = \frac{1}{MN}\sum_m\sum_n\left[I_1(m,n)-I_2(m,n)\right]^2 \qquad (1)$$

New measure is based on structural properties of an image, therefore it can be expected that the new measure will provide the detection of content changes more adequately than any point-by-point measure.

The mentioned measure is described below. We consider a model of image structure based on the set of edges which are determined by the gradient field of the image [Asatryan, 2009]. Here we rest upon the fact that the HVS uses the edge information for understanding and analyzing the structure of an image [Wang, 2002], [Wang, 2009]. It is also very important that the edges are invariant to image rotation, scaling and other transformations, so they provide more adequate extraction of structural information from any image [Wang, 2004]. In [Asatryan, 2009] a measure is proposed for images similarity assessment based on using the gradient magnitude distribution.

Let $\|G_H(m,n)\|$ and $\|G_V(m,n)\|$ at a point (m, n) of an image be the horizontal and vertical gradients, determined by one of known gradient methods, and the matrix of gradient magnitude $\|M(m,n)\|$, where

$$M(m,n) = \sqrt{G_H^2(m,n) + G_V^2(m,n)} \tag{2}$$

Following to [Asatryan, 2009] we suppose that the gradient magnitude (2) is a random variable with Weibull distribution density

$$f(x;\eta,\sigma) = \frac{\eta}{\sigma}\left(\frac{x}{\sigma}\right)^{\eta-1} \exp\left[-\left(\frac{x}{\sigma}\right)^{\eta}\right], x \geq 0,\ \eta > 0,\ \sigma > 0 \tag{3}$$

As a measure of structural similarity of two images with probability distribution functions of gradient magnitude $f_1(x;\eta_1,\sigma_1)$ and $f_2(x;\eta_2,\sigma_2)$ accordingly, we take

$$W^2 = \frac{\min(\eta_1,\eta_2)\min(\sigma_1,\sigma_2)}{\max(\eta_1,\eta_2)\max(\sigma_1,\sigma_2)},\ 0 < \mathrm{W}^2 \leq 1, \tag{4}$$

where the corresponding parameters are represented as statistical estimations gotten from the corresponding samples of gradient magnitude.

As it has been shown in [Asatryan, 2009 - Asatryan, 2012] the measure (4) has certain advantages against widely used other methods, which use image pixels point-by-point comparison method.

Results of Experiments

We tested our method with several video clips of different themes and varying nature. The videos fluctuate widely in content and length. But in this section we include analyzes and results on one exact video. The purpose of this section of our work is to illustrate how our algorithm works and compare the results with widely known PSNR method mentioned above.

We take manually detected positions of the shot boundaries as the ground truth, defining in that way the number of missed detections and false detections.

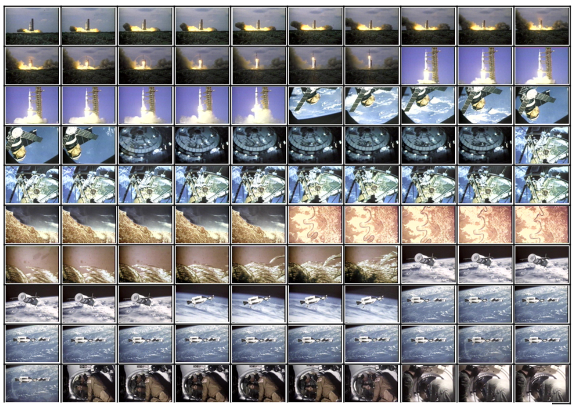


Fig.2 Consecutive frames from video sequences presenting abrupt cuts

The frame sequence fragment of video under test is demonstrated in Fig.2.

The video was divided into 130 frames and contains 9 cuts. The values of PSNR and W^2 have been calculated for all adjacent frames. The experimental results are shown in Fig.3 and Fig.4. From graphical representation it is clearly visible that the threshold value for decision making regarding presence of cuts is of 0.8. But this value can vary for other type and content videos. Our experiments show that acceptable threshold t_c for W^2 vary between 0.6 and 0.8.

It can be noted that a reasonable algorithm for cut detection using PSNR curve can be based on choosing some extremal values of PSNR instead of using a threshold. However, we can propose a statistical model for determination of the threshold for cut detection by using PSNR. The model is based on assumption that the adjacent frames consists of pixels with gradient magnitudes, which are samples from independently and normally distributed random variable. Let Δ be the dynamic range and σ^2 be the variance of the magnitudes related to a shot. Then $\Delta = 6\sigma$, $MSE^2 \approx 2\sigma^2$, $PSNR = 10\log\frac{\Delta^2}{2MSE^2} \approx 12.5\,dB$, so we can put $t_c = 12.5$.

The values of specified thresholds are shown in Fig. 3 and Fig.4. Solid circles are real cuts; gray circles are false cuts and white circles - cuts that have not been detected. One can see that our proposed method correctly detect all the cuts, while PSNR gives missed hits and false hits.

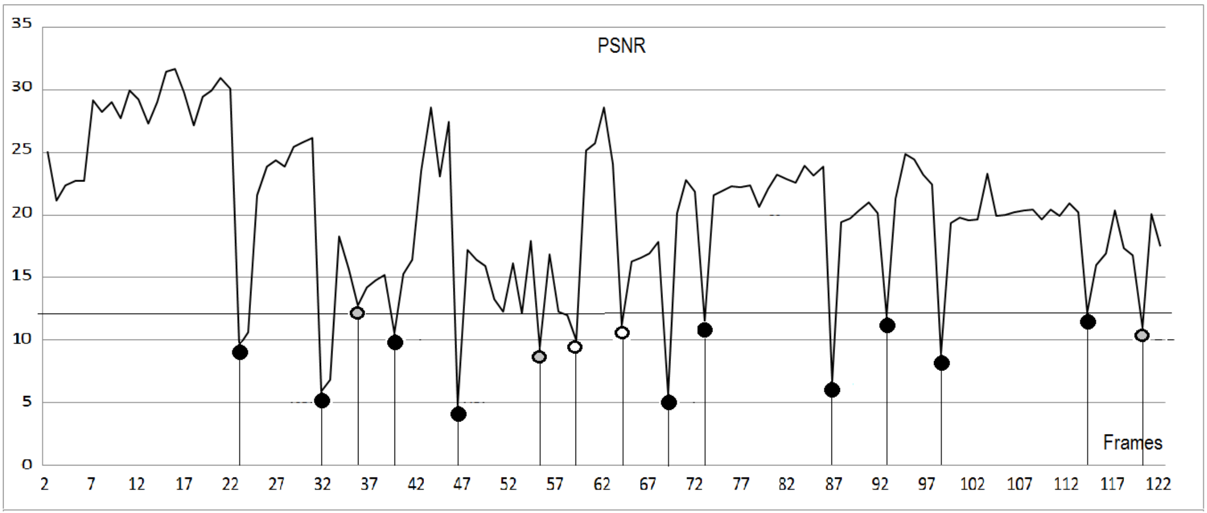


Fig.3 Results of cut detection by using PSNR

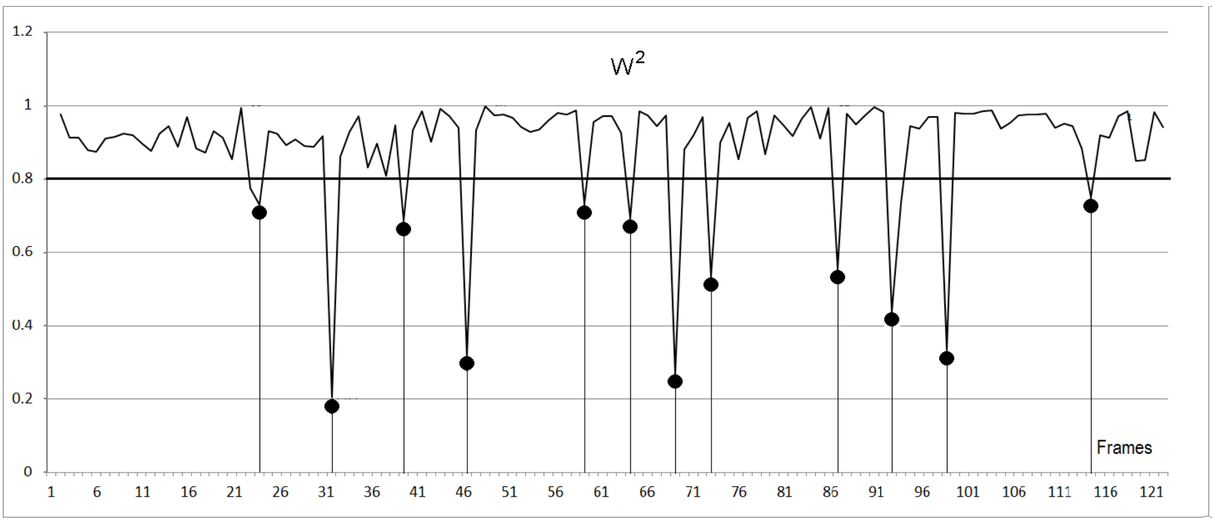


Fig.4 Results of cut detection by using W^2

As it was already mentioned to compare similarity of two frames with our method, we need to estimate two parameters of corresponding Weibull distributions. It is interesting to investigate the behavior of content of the frames within each frame of a shot by graphical analyzing the (η, σ) - scatter.

Fig.5 shows that corresponding shot did not change actively, while Fig. 6 shows that the sense was dynamically changed from the beginning to end. This kind of analyzes may help to analyze some physical process, which is controlled by video camera.

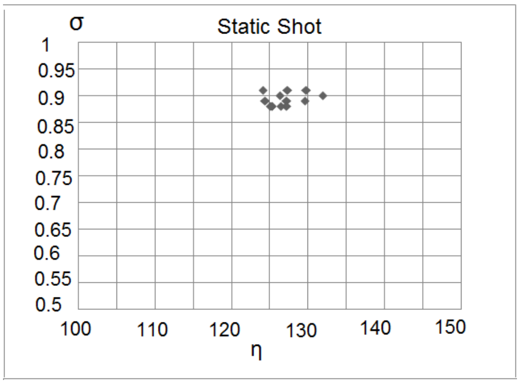


Fig.5 Distribution of σ and η parameters for static shot

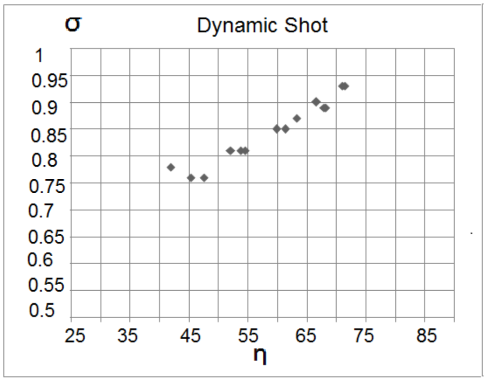


Fig.6 Distribution of σ and η parameters for dynamic shot

Conclusion

Shot boundary (cut) detection is the main and important step of searching and browsing the digital video. In this paper we propose a novel method for shot detection based on Weibull distribution model. This is a new technique for shot detection not based on existing methods or their combinations. It is using structural properties of the image and gives results which are closer to human visual system perception. Earlier the method was successfully compared to some of widely using methods reported in literature. The experimental results show the effectiveness of proposed method of cut detection, in comparison with method based on widely used mean-square measure. Convenient algorithm used for images similarity measurement, which is using only two parameters of an image, may be easily and effectively used to do more deeper analyzes inside and also between different shots, which can give general information about type of whole video or separate segments on it. We believe that this method can be successfully applied to detection of gradual transitions, such as fade in/ fade out, dissolves and etc., and our farther works will be devoted to them.

Acknowledgment

This work was supported by the State Committee of Science of Armenia (13A-1b25). http://scs.am/

Bibliography

[Asatryan, 2009] D. Asatryan and K. Egiazarian. Quality Assessment Measure Based on Image Structural Properties. International Workshop on Local and Non-Local Approximation in Image Processing, Finland, Helsinki, pp. 70-73, 2009.

[Asatryan, 2010] D. Asatryan, K. Egiazarian and V. Kurkchiyan. Orientation Estimation with Applications to Image Analysis and Registration. International Journal "Information Theories and Applications", vol. 17, no. 4, pp. 303-311, 2010.

[Asatryan, 2012] D. Asatryan, V. Kurkchiyan and M. Bagramyan. A Method for Quality Assessment of Image Resizing Algorithms. Mathematical Problems of Computer Science, vol. 36, pp. 128-132, 2012.

[Borecsky, 1996] J. S. Borecsky and L. A. Rowe. Comparison of video shot boundary detection techniques. Journal of Electronic Imaging, vol.5, no.2, pp.122–128, 1996.

[Cheng, 2002] Y. Cheng, X. Yang, and D. Xu. A method for shot boundary detection with automatic threshold. Proceedings of IEEE TENCON, vol. 1, pp. 582-585, 2002.

[Fernando, 2000] W.A.C. Fernando, C.N. Canagarajah, and D.R. Bull. A unified approach to scene change detection in uncompressed and compressed video. IEEE Transactions on Consumer Electronics, vol. 46, no. 3, pp. 769–779, 2000.

[Hanjalic, 2002] A. Hanjalic. Shot-Boundary Detection: Unraveled and Resolved? IEEE Transactions on Circuits and Systems for Video Technology, vol. 12, no. 2, pp. 90-105, 2002.

[Kim, 2009] W.Kim, K. Moon and J.Kim. An automatic shot change detection algorithm using weighting variance and histogram variation. 11th International Conference on Advanced Communication Technology, vol. 2, pp.1282-1285, 2009.

[Liu, 2010] W. Liu. An Effective Method For Abrupt Scene Change Detection. 2010 Sixth International Conference on Natural Computation , vol. 2, pp. 850-852, 2010

[Smoliar, 1994] S.W. Smoliar and H.J. Zhang. Content-based video indexing and retrieval. IEEE Multimedia, vol. 1, no. 2, pp. 62-72, 1994.

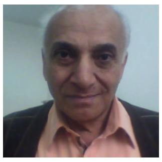
[Wang, 2002] Z. Wang, A.C. Bovik. A universal image quality index. IEEE Signal Processing Letters, vol. 9, no. 3, pp. 81-84, 2002.

[Wang, 2004] Z. Wang, A.C. Bovik, H.R. Seikh, and E.P. Simoncelli. Image quality assessment: From error visibility to structural similarity. IEEE Transactions on Image Processing, vol. 13, no. 4, pp. 600-612, 2004.

[Wang, 2009] Z. Wang and A.C. Bovik. Mean Squared Error: Love It or Leave It?. IEEE Signal Processing Magazine, vol. 26, no.1, pp. 98-117, 2009.

[Zhi, 2005] M. Zhi and A. N. Cai. Shot change detection with adaptive thresholds. IEEE International Workshop on VLSI Design and Video Technology, pp. 147-149, 2005.

Authors' Information



David Asatryan *– Professor, Head of group of the Institute for Informatics and Automation Problems of NAS Armenia, 1, P.Sevaki Str., 0014, Yerevan, Armenia; e-mail: dasat@ipia.sci.am*

Major Fields of Scientific Research: Digital signal and image processing.

Manuk Zakaryan *– Ph.D. student at Russian – Armenian (Slavonic) University, Software Developer at EGS Armenia; e-mail: zakaryanmanuk@yahoo.com*

Major Fields of Scientific Research: Digital signal and image processing, Software developing.