METHOD FOR VIDEO SHOT DETECTION AND SEPARATION

David Asatryan, Manuk Zakaryan

Abstract: *Shot boundary detection is main step in video management systems like browsing and indexing. In this paper, we shortly describe an earlier proposed shot detection algorithm based on the structural properties of video frames. Two mathematical models for decision making, i.e. for similarity threshold determination are proposed and compared. The first model allows determination of threshold in case of using mean-square deviation for frames similarity and when the image pixel is assumed to be normal distributed random variable. The second model is based on using the structural similarity measure based on Weibull model for image gradient magnitude distribution. Global and adaptive approaches are considered for frames similarity threshold determination. Results of experiments to detect real video shots are given. It is shown that adaptive threshold determination method generally gives more acceptable results than global threshold determination method. At the same time the* W^2 *based threshold determination approach gives more accurate results than PSNR based approach, therewith corresponds to HVS perception.*

Keywords: *video segmentation, shot detection, similarity measure, threshold determination*

ACM Classification Keywords: *Image Processing and Computer Vision*

Introduction

In recent years content-based indexing and retrieval of digital video becomes an active research area. Why it is so important to have reliable transition detection algorithm and techniques? Firstly, shots are generally considered as the elementary units constituting a video. Detecting shot boundaries means recovering those elementary video units, which can be a ground for existing video abstraction and high-level video segmentation algorithms. Secondly, during video production each transition type is chosen carefully in order to support the content and context of video sequences. So in future it may be easy to automatically detect transitions according to their type and content. Also shot detection can be used in security purposes, so that if a usual scene is changed abruptly, it can be easily detected by controlling tools.

There is several video segmentation algorithms proposed in the literature currently which are based on various statistical properties of the shots. 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. They can be classified as: intensity/color based, edge/contour based and motion based [Lienhart, 2001]. In the class of intensity/color– based methods, the histogram–based methods achieve better results comparing with other methods [Gargi, 2000]. Various approaches to histogram–based cut detection were proposed using color based histogram metrics or histogram differences in several color spaces.

In this paper, we will make short overview of our proposed method and algorithm, which already showed the effectiveness in cut detection problems [Zakaryan, 2012; Asatryan, 2014]. The main idea of this method is based to well-known thesis that the human visual system makes an imaginary segmentation mainly using the notional and structural features of the frame.

After having good shot boundary detection algorithms which is mainly based on similarity or dissimilarity measures between consequent frames the next important step of segmentation is to have a threshold which describes a comparison value. We can divide threshold setting methods 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 iterations and must find different optimal threshold for various 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 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, 2005]. There are several ways for adaptive threshold detection in existing literature. Average calculation in which supposed that when there is a shot boundary, the frame difference is more than the average frame difference before the shot boundary. And if there is a hard cut or shot boundary, the frame difference is also more than the average frame difference after the shot boundary. Another widely used method is sliding window method, when a window of specific length is chosen and similarity comparison between frames is done in rang of this window, and after each step the window is moving forward. Our model of adaptive threshold detection is based on statistical model of image similarity measure described in [Asatryan, 2009].

Proposed Algorithm for Shot Detection

Shot detection algorithm is usually based on consecutive determination of similarity of neighboring frames and detecting abrupt similarities between them. Let consider the frames $f_1, f_2, ..., f_k, f_{k+1}$ and the sequence of corresponding values $\mu_{1,2}, \mu_{2,3}, ..., \mu_{k,k+1}$ of predefined similarity measure μ . 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 [Asatryan, 2014]. The sequence of frames between neighboring cut points is considered as a shot. The quality of shot detection algorithm is generally depends on properties of chosen similarity measure.

Measure for structural similarity estimation of images. We consider a model of image structure based on the set of edges which are determined by the gradient field of the image. For simplicity we consider the Gray Scale (8 bit) format image $I = \{I(m,n)\}$ with $M \times N$ sizes, $m = 0,1,...,M$; $n = 0,1,...,N$. 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

$$M(m,n) = \sqrt{G_H^2(m,n) + G_V^2(m,n)} \quad (1)$$

We follow [Asatryan, 2009] and suppose that the gradient magnitude (1) is a random variable with Weibull distribution density

$$f(x;\eta,\sigma) = \frac{\eta}{\sigma}\left(\frac{x}{\sigma}\right)^{\eta-1} \exp\left[-\left(\frac{x}{\sigma}\right)^{\eta}\right], x \ge 0, \eta > 0, \sigma > 0 \quad (2)$$

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

$$W^2 = \frac{\min(\eta_1,\eta_2)\min(\sigma_1,\sigma_2)}{\max(\eta_1,\eta_2)\max(\sigma_1,\sigma_2)}, \ 0 \le W^2 \le 1, \quad (3)$$

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

Two Mathematical Models for Threshold Determination

As already has been mentioned the fixed threshold which is also called *global threshold* is a technique based on time series of feature values of a similarity measure μ, which in the ideal case is supposed to show a single large peak at hard cut locations. A hard cut is declared each time the feature value $f(t)$ surpasses a globally fixed threshold.

In general, similarity threshold determination is a challenging problem due to large variety of types of video sequences. Therefore this problem is mostly solved in experimental way. Any theoretical approach to this problem requires some general limitations on formal features of video frames. So it is necessary to take into account that the results obtained under these limitations can also have limited reliability [Yi, 2012].

We propose two mathematical models for determination of acceptable threshold when the similarity measures PSNR (peak-to-noise-ratio) and W^2 defined by formula (3) are used.

Global threshold by PSNR. It is well known that when PSNR>30 dB, the human visual system (HVS) practically does not notice the difference between two compared images. But there arises a question: when else HVS is not notice the similarity of images? It can be accepted the following answer: when the structures of images are quite different or images don't have any interpretable structure. In the capacity of such images we can consider random images, i.e. when the intensities of pixels of images are independent random variables. Thus we can set a statistical model which based on assumption that the adjacent frames consists of pixels with independently distributed random intensities. To concretize the model we assume that the intensities have normal distribution $X \sim N(m_j, \sigma_j)$, $j = 1,2$. Then, of course,

$$P\left\{-3\sigma_j \le \left|X - m_j\right| \le 3\sigma_j\right\} \approx 0.997 \qquad (4)$$

Let PSNR is calculated by formula as follows

$$PSNR = 10 \log_{10} \frac{\Delta^2}{MSE^2}, \quad MSE^2 = \frac{1}{MN} \sum_m \sum_n \left[I_1(m,n) - I_2(m,n)\right]^2, \qquad (5)$$

where $\Delta = \max_{m,n} \left|I_1(m,n) - I_2(m,n)\right|$ is dynamic range of difference between image pixels intensities.

Let, for simplicity, $\sigma_1 = \sigma_2 = \sigma$, $m_1 = m_2$, $M,N >> 1$. Then under (4) we can calculate $\Delta \approx 6\sigma$, $MSE^2 \approx 2\sigma^2$, and

$$PSNR = 10 \log \frac{\Delta^2}{2\sigma^2} = 10 \log \frac{36\sigma^2}{2\sigma^2} \approx 12.5\,dB \qquad (6)$$

so we can put $t_c = 12.5\,dB$.

We have shown in [Asatryan, 2014] that the decisions on cut presence using threshold of $t_c = 12.5dB$ conform with visual analysis of PSNR chart for video sequence.

Adaptive threshold by PSNR. In general, determination of adaptive threshold for cut detection can be based on using the samples of similarity measure μ between consecutive frames of the ordinary shot. Decision on cut presence is made as it is described in previous section.

In case of PSNR it is reasonable to use the ratio $\frac{\Delta^2}{MSE^2}$ instead of PSNR. Threshold t_c is calculated by formula

$$t_c = \overline{X} - lS \text{ for } l = 2 \text{ or } l = 3 .$$

Unfortunately, nowadays there is no convenient model for distribution of measure W^2 for random images like considered above. Therefore determination of relevant threshold was performed experimentally. The series of experiments show that acceptable threshold t_c for W^2 varies between 0.6 and 0.8.

Adaptive threshold by measure W^2. For adaptive threshold detection we use a statistical model based on similarity measure of consecutive frames. Unfortunately, the exact distribution of measure W^2 is not yet known, but because of large number of frames in each shot we can use the normal model for asymptotic distribution of W^2 and write

$$P\{\overline{x} - 2s < W^2 \le \overline{x} + 2s\} \approx 0.95 \tag{7}$$

where $\overline{x}$ is the average value and s is the mean square deviation of corresponding samples $W_1^2, W_2^2, ..., W_K^2$, within the current shot, and K is the number of current frame.

If W^2 between K-th and $K+1$-th frames satisfies equation $W^2 < \overline{x} - 2s$, then we consider that there is a cut in $K+1$-th frame.

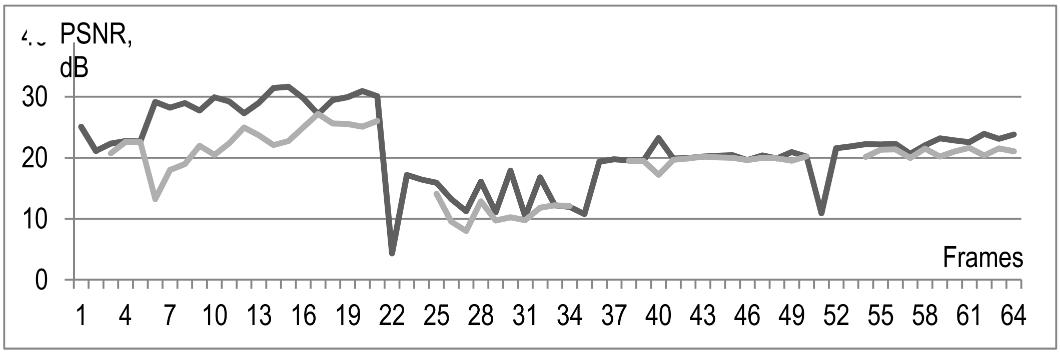

Figure 1. Dependence of PSNR (black curve) and PSNR adaptive threshold (Grey curve) on frame number

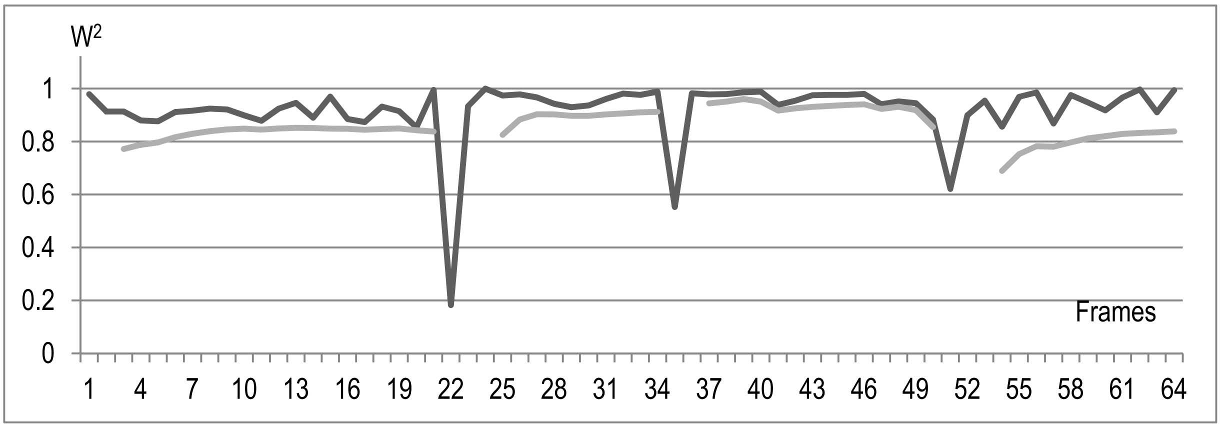

Figure 2. Dependence of W^2 (black curve) and W^2 adaptive threshold (Gray curve) on frame number

Results of Experiments

Described method of shot boundary detection was tested for various video sequences and some results have been given in previous articles [Zakaryan, 2012; Asatryan, 2014]. Here we illustrate the effectiveness of method when adaptive threshold is used.

The results of experiments for both PSNR and W^2 methods are shown in Figure 1 and Figure 2 with black it is shown the similarity graph and the gray one is the adaptive threshold determined above.

More clear result have been reached when we use $l = 3$. As it can be seen from these two figures, for both methods our proposed adaptive algorithm works acceptable, but better and accurate result we gone for W^2 method.

Conclusion

In this paper, we consider the problem of video shots detection and propose two mathematical models for similarity threshold determination. Methods for global and adaptive threshold determination for each model are proposed. Results of experiments to detect real video shots are given. It is shown that adaptive threshold determination method generally gives more acceptable results than global threshold determination method. At the same time the W^2 based threshold determination approach gives more accurate results than PSNR based approach, therewith corresponds to HVS perception.

Bibliography

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

[Asatryan, 2014] D.G. Asatryan, M.K. Zakaryan. Improved Algorithm for Video Shot Detection. International Journal "Information Content and Processing", vol. 1, pp. 66-72, Number 1, 2014

[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.

[Gargi, 2000] U. Gargi, R. Kasturi, S. H. Strayer, Performance Characterization and Comparison of Video - Shot – Change Detection Methods, IEEE Transactions on Circuits and Systems for Video Technology, vol. 10, pp. 1-13, 2000.

[Kim, 2005] 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.

[Lienhart, 2001] R. Lienhart, Reliable transition detection in videos, A survey and practitioners guide. International Journal of Image and Graphics, vol. 1, pp. 469-486, 2001

[Yi, 2012] Huo Yi, Zhang Pengzhou , Wang Yanfeng, Adaptive threshold based video shot boundary detection framework, Image Analysis and Signal Processing (IASP), 2012 International Conference on Digital Object Identifier, pp. 1- 5, 2012.

[Zakaryan, 2012] M.K. Zakaryan. Novel Approach to Video Cut Detection and Segmentation. В сборнике научных трудов 7-й годичной научной конференции Российско-Армянского (Славянского) университета, vol. 1, pp. 1-5, 2012.

[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.