

# VIDEO SHOT META-SEGMENTATION BASED ON MULTIPLE CRITERIA FOR GRADUAL TRANSITION DETECTION

*Efthymia Tsamoura, Vasileios Mezaris, Ioannis Kompatsiaris*

Informatics and Telematics Institute / Centre for Research and Technology Hellas,  
1st Km Thermi-Panorama Rd, Thessaloniki 57001, Greece  
{tsamoura, bmezaris, ikom}@iti.gr

## ABSTRACT

Shot boundaries provide the basis for almost all high-level video content analysis approaches, validating it as one of the major prerequisites for efficient video indexing and retrieval in large video databases. The successful detection of both gradual and abrupt transitions is necessary to this end. In this paper a new gradual transition detection algorithm is proposed, based on novel features exhibiting less sensitivity to local or global motion than previously proposed ones. These features, each of which could serve as a stand-alone transition detection approach, are then combined using a machine learning technique, to result in a meta-segmentation scheme. Besides significantly improved performance, advantage of the proposed scheme is that there is no need for threshold selection, as opposed to what would be the case if any of the proposed features were used by themselves and as is typically the case in the relevant literature. Comparison of the proposed approach with four popular algorithms of the literature reveals the significantly improved performance of it.

## 1. INTRODUCTION

Video manipulation and most notably retrieval has become in recent years an integral part of our everyday lives. This necessitates the development of techniques that cover the entire range of video analysis and understanding, to support efficient and effective indexing and retrieval of vast amount of video content. A basic analysis step contributing towards this goal is video shot segmentation.

Video shot is a collection of frames with high degree of affinity that constitute a self-contained visual entity. In most cases shots contain fundamental information for events that occur in videos. Shot boundaries provide the basis for almost all high-level video content analysis approaches, validating it as one of the major prerequisites for successful indexing and retrieval in large video databases.

Transitions between video shots can be clustered into two main categories; abrupt transitions and gradual transitions. In

abrupt transitions, the differences between two consecutive frames are extremely high. On the other hand, there are several kinds of gradual transitions including fades, dissolves and wipes. The common characteristic between them is the existence of frames that clearly mark the transition period. The above bounding frames constitute the ending of the current shot and the beginning of the next shot. Each frame in the transition period is typically a combination of the starting and ending frames.

Many algorithms have been proposed for abrupt as well as gradual transition detection. In [1], a machine learning approach was presented for abrupt transitions detection. A number of features, such as color histograms and shape descriptors, were exploited for that purpose. This approach was shown to provide very accurate detection of abrupt transitions; however, its generalization for detecting gradual transitions as well is not straightforward.

A twin threshold algorithm for detecting fades and dissolves, based on a histogram difference metric, is proposed in [2]. One global threshold bounds the total differences of all consecutive frames that participate in the transition, while two extra local thresholds bound the difference of each pair of consecutive frames. If the difference between the current frame with the last frame belonging to the transition period is bounded by the local thresholds and the sum of differences between transition frames is bounded by the global threshold, then the current frame is appended to the transition period.

In [3] transition detection is based on the analysis of intensity edges. An edge pixel that appears far from an existing edge pixel is defined as an entering pixel, while a previously existing edge pixel that disappears is defined as an exiting pixel. Transitions can be detected by counting the number of exiting and entering pixels and in particular the Edge Change Ratio (*ECR*). In [4] dissolves and fades were detected employing different distance metrics between consecutive frames. J.S.Boreczky [5] made an extensive comparison of various distance metrics for abrupt and gradual transition detection.

In [6] dissolve detection is implemented via analysis of a characteristic curve estimated from the input video. Ana-

---

This work was supported by the European Commission under contracts FP6-045547 VID-Video, FP6-027685 MESH and FP6-027026 K-Space.

lytically, a dissolve modelling error is being compared with the actual dissolve modelling error estimated from the input video. Let  $p(x, y)$  and  $q(x, y)$  be the starting and ending frames of a dissolve transition respectively. It is hypothesized that a dissolve is being constructed by a linear combination of the above frames. If  $f(x, y, t)$  represents a frame during dissolve at time  $t$  then  $f(x, y, t) = \frac{t_2-t}{t_2-t_1} p(x, y) + \frac{t-t_1}{t_2-t_1} q(x, y)$  where  $t_1, t_2$  are the start and end time of the dissolve and  $t_1 < t < t_2$ . Under this assumption it was shown in [6] that the variance of pixel luminance curve exhibits parabolic shape during an actual dissolve. Candidate dissolve areas are identified using the characteristics of the first and second derivatives of the luminance variance curve and are verified using an adaptive threshold. The same philosophy is exploited in [7]. Although the above modelling is reasonable, it is inefficient to capture transition areas where pixel intensities change non-linearly. C.W. Su et al. [8] proposed a model to overcome the above problem. A sliding window that covers a number of frames is used, in order to observe the change of pixel intensities. If the number of pixels that undergo monotonous intensity changes (either increasing or decreasing) within the window is above a threshold, it is concluded that the frames included within the window belong to a dissolve area.

Although most of the proposed algorithms attempted to overcome limitations such as sensitivity to global or local motion, their performance is moderate. From this it becomes clear that a single feature alone (e.g. *ECR*) is difficult to accommodate for all possible effects that influence shot detection, such as local and camera motion, measurement noise, etc. In this paper a novel approach for dissolve and fade transitions is presented. Two main directions are followed. As a first direction, novel shot change detection features are proposed, exhibiting reduced sensitivity to motion activity. As a second direction, a method of combining them in a meta-segmentation scheme, in order to achieve more accurate detection results, is proposed. The segmentation process under the proposed approach can be summarized as follows: selected features are initially computed for every video frame. Given a couple of consecutive frames, the distances between the above features are then computed forming distance vectors. These vectors are subsequently supplied to a trained binary classifier. The output of this classifier denotes the class membership of each of the examined frames, i.e. whether it is part of a transition area or not.

The paper is organized as follows. In section 2, a set of new image features showing less sensitivity to global or local motion effects are presented. In section 3, a machine learning approach based on a Support Vector Machine (*SVM*) classifier is employed for combining the aforementioned features for gradual transition detection. In section 4 the proposed approach is compared with 4 very popular gradual transition detection algorithms. The experimental results show that the proposed algorithm has significant performance gains. Finally, conclusions are drawn in section 5.

## 2. INDIVIDUAL FEATURES FOR GRADUAL TRANSITION DETECTION

The fundamental element of successful gradual transition detection is the selection of the appropriate image features. In this work we focused on finding image features that are affected from motion events as little as possible but are at the same time sensitive to gradual transitions.

### 2.1. Macbeth Color Histogram

A feature that is widely employed for shot detection is color. In [9], several simple statistics based on color histograms are compared for gradual transition detection. Best results were obtained by breaking images into 16 regions, using a  $\chi_2$  test on color histograms and discarding the 8 largest differences to reduce object motion and noise influence. In general, even in the presence of object or camera motion, the color distribution is not expected to show abrupt changes. Especially in cases where the moving object cover a small percentage of the image pixels, or the camera is not moving very fast, color changes should be relatively small since only a small portion of pixels are affected each time. On the other hand, when a gradual transition occurs, values of color histogram bins may vary enough to detect transitions, since changes may take place in a much bigger part of image.

A feature based on the Macbeth color pallet [10] is introduced in this work. The Macbeth pallet consists of twenty four colors which were selected according to human color perception, i.e. the clusters that humans categorize colors with the ones included in Macbeth pallet are approximately the same. Color pallets have been used successfully for retrieval in large image databases. Use of a predefined color pallet has low computational complexity. Pixel colors are mapped to one of the 24 colors of Macbeth pallet, constructing a 24 bins histogram.

Let  $I_t$  be a video frame at time  $t$ .  $Z_i$   $i = 1 \dots 24$  are the Macbeth pallet color clusters. Pixel  $I_t(x, y) = [R_{xy} \ G_{xy} \ B_{xy}]$  belongs to the color cluster  $Z_i$  for which the manhattan distance  $d = \text{manhattan}([R_{xy} \ G_{xy} \ B_{xy}], Z_i)$  is minimized. We choose the *manhattan* distance due its better suitability for the *RGB* color space, as compared to other distance functions, e.g. the euclidian. The distance between frames  $I_t$  and  $I_{t-1}$  using their estimated Macbeth color histograms  $M_t$  and  $M_{t-1}$ , can then be defined as:

$$D_t^M = \sum |M_t - M_{t-1}|, \quad (1)$$

where  $|\cdot|$  is the  $L_1$  distance. Computing the distances between each pair of consecutive frames based on the Macbeth color histogram feature for an input video, a curve  $D_t^M$ ,  $t = 0, \dots, T$  is produced, where  $T$  corresponds to the video duration. Values higher than a threshold  $D^H$  indicate that corresponding frames are identified as belonging to a transition area, while the rest ones as belonging to non-transition areas.

A simple threshold estimation procedure, similarly to [8], can be followed to determine threshold value  $D^H$ .

## 2.2. Luminance Center of Gravity

When a dissolve occurs, the spatial distribution of pixel intensities changes. A simple yet neglected in the shot detection literature characteristic is the luminance center of gravity. It's definition is analogous to an object's center of mass: it is the point where luminance is concentrated on. Let  $L_t(x, y)$  be the luminance value calculated for pixel  $\mathbf{I}_t(x, y)$ . Then the luminance center of gravity  $\mathbf{R}$  is computed as:

$$\mathbf{R} = [R_x \ R_y] \quad (2)$$

$$R_x = \frac{\sum_x x l_t(x, y)}{\sum_x l_t(x, y)} \quad (3)$$

$$R_y = \frac{\sum_y y l_t(x, y)}{\sum_y l_t(x, y)} \quad (4)$$

Eq.(2) treats the image as a system of particles with masses equal to their luminance. A great advantage of the above feature is its persistence to slight pixel luminance variations as well as to measurement noise effects. If a small percentage of pixel values changes due to local or global motion, then the location of the center of luminance is not likely to be significantly affected. On the other hand it is sensitive to dissolves, since in that case intensities would change for a significant amount of image pixels. One disadvantage of the above feature is that it is sensitive to fast camera motion or local motion of large objects. Nevertheless, it is a good characteristic that can contribute towards detection of gradual transitions.

The distance of the luminance centers of gravity between frames  $\mathbf{I}_t$  and  $\mathbf{I}_{t-1}$ , having  $\mathbf{R}_t$  and  $\mathbf{R}_{t-1}$  respectively is

$$D_t^R = \|\mathbf{R}_t - \mathbf{R}_{t-1}\|, \quad (5)$$

where  $\|\cdot\|$  denotes the euclidian distance. Computing distances between frames based on the luminance center of gravity feature for an input video, a  $D_t^R$  curve,  $t = 0, \dots, T$  is produced. As the value of curve  $D_t^R$  increases, the distance between a couple of consecutive frames becomes bigger. A threshold can be set as in the case of color Macbeth histogram curve, for distinguishing between transition and non-transition frame areas.

## 3. META SEGMENTATION APPROACH

As previously mentioned, a single feature alone is difficult to accommodate for all possible effects that influence gradual shot detection. To alleviate this drawback, a combination of multiple individual features can be employed. To this end, a machine-learning based classification approach is adopted in this work. Commonly employed machine learning techniques

for binary or multi-class classification include Neural Networks (NN) and Support Vector Machines (SVM). The latter can effect non linear classification with the use of kernel functions. Kernel functions, such as gaussian and sigmoid, map input data into a higher dimensional space, where data are linearly separable. Two main properties of SVM lead us to use them instead of NN. The first is that only 2 classes exist in our problem; transition occurrence, non-transition. The second is that SVM have the potential to estimate the optimal separating hyperplane between the two examined classes. Gradual transition detection, under the proposed SVM machine learning meta-segmentation technique, has two phases: training and evaluation. For training, the classifier must be supplied with training vectors and each training vector must be manually assigned to the appropriate class. We denote with “-1” the transition and with “1” the non-transition class. When the training step is completed, new input vectors can be classified by the SVM to one of the two aforementioned classes. In our case input vectors are made of distances of various features, estimated from consecutive video frames.

We construct input vectors to the SVM by combining distances of the two features proposed in section 2, with an existing feature: monotonous intensity change. In [8], it is hypothesized that in a gradual transition, pixel intensities vary monotonously. The percentage of pixels with monotonously varying intensities is calculated and if it exceeds a certain threshold, a dissolve/fade transition is detected. Let  $f(x, y, t) = L_{t+1}(x, y) - L_t(x, y)$ . Then, the monotonous change of intensity is evaluated using the following equation:

$$g(x, y, t) = \begin{cases} 1, & f(x, y, t)f(x, y, t-1) \geq 0 \\ 0, & f(x, y, t)f(x, y, t-1) < 0 \end{cases} \quad (6)$$

It is clear that  $g$  at time  $t$  and pixel  $(x, y)$  is equal to 1, if and only if the intensity in  $(x, y)$  varies monotonously at time  $t$ . Then the percentage of pixels with monotonously varying intensities at a given time  $t$  can be defined as  $h(t) = \sum g(x, y, t)/XY$ , where  $X$  and  $Y$  are the width and height of input frames.

Consequently, under the proposed approach each input vector to the SVM classifier, corresponding to a pair of consecutive frames, is made of the distance of the Macbeth color histograms, the distance of the luminance center of gravities, and the monotonous intensity change feature  $h(t)$ . The 3 dimensional distance vector  $D_t$  between frames  $I_t$  and  $I_{t-1}$  is:

$$\mathbf{D}_t = [D_t^M \ D_t^R \ h(t)] \quad (7)$$

For classification we used a C-SVM with a radial basis function kernel of  $3^{rd}$  degree. In the SVM training stage training vectors are computed (Eq.7) and manually assigned to one of the two classes (transition, non-transition). These are then used by the SVM for estimating the corresponding optimal separating hyperplane. After training, for detecting gradual

transitions in a non-segmented video set, distance vectors between all consecutive frames are again estimated. The output of the classifier for every input vector is “-1” or “1”, indicating its class membership.

A significant advantage of the proposed scheme is that input vectors are automatically classified, without any threshold. The SVM classifier has bypassed the need for threshold selection by learning from the training set the optimal separating hyperplane instead. This characteristic is very important since it alleviates the need for threshold selection and fine-tuning.

#### 4. EXPERIMENTAL RESULTS

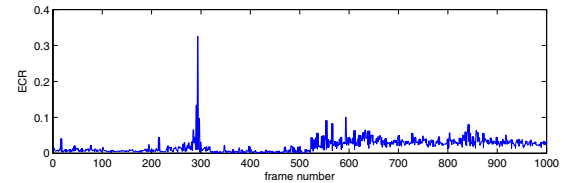
The performance of our algorithm was evaluated on 2 videos of the TRECVID-2007 video set. Both video sequences had a frame size of  $352 \times 288$  pixels at 25 fps. Gradual transitions were evaluated against manually generated ground truth results. These videos exhibited a large number of gradual transitions and intense motion events. It must be emphasized that about 30% of the included gradual transitions were relatively easy to identify, since they did not involve significant motion or color changes. The remaining gradual transitions involved global or local motion and/or luminance changes that made their automatic detection a challenging task. The use of such a challenging test set has made possible the comparative evaluation of the proposed approach under realistic conditions. A 10 minutes segment of the above video set was used for training of the proposed system, whereas the remaining 50 minutes were used in the evaluation experiments presented below.

Two measures were employed to compare the proposed algorithm with the existing ones; precision and recall. Precision is defined as the fraction of true gradual transitions returned by the algorithm over all transitions returned by the algorithm. On the other hand recall is defined as the fraction of true gradual transitions returned by the algorithm over all true gradual transitions present in the test corpus. Table 1 shows the results of the proposed meta-segmentation approach and compares them with results of four other approaches. Results for shot boundary detection based on each single feature of the two proposed in section 2, are also shown in the same table. To produce the latter, the distance curves  $D_t^R$  and  $D_t^M$  were computed as described in section 2. It should be noted that the results shown in Table 1 have been estimated considering gradual transitions only. These results are therefore, not directly comparable with those of any evaluation activity considering both gradual and abrupt transitions at the same time, particularly since in the various test corpora gradual transitions typically account for only a small portion of the overall transitions.

We compared our algorithm with 4 popular approaches for gradual transition detection: Edge Change Ratio (ECR) [3], standard deviation of pixel intensities [6], Effective Average Gradient (EAG) [7], and monotonous intensity change [8].

Method	Recall	Precision
Standard Deviation of Pixel Intensities [6]	0.765	0.484
ECR [3]	0.654	0.125
EAG [7]	0.28	0.146
monotonous intensity change [8]	<b>0.78</b>	0.409
Luminance center of gravity	0.75	0.15
Macbeth color histogram	<b>0.78</b>	0.45
Proposed meta-segmentation approach	<b>0.78</b>	<b>0.67</b>

**Table 1.** Recall and precision evaluation of gradual transition detection algorithms. The proposed approach significantly outperforms previous approaches of the literature.



**Fig. 1.** *ECR* curve for the frame range [1 1000] of a test video.

During implementation of [3], input videos were smoothed by convolution with a gaussian kernel of standard deviation  $\sigma = 0.5$ . Edge detection was performed using the Canny edge detector. The expansion distance  $r$ , (parameter which controls the minimum distance between and entering and exiting pixel) is set to 6. Zabih et al. claimed that transitions can be detected by analyzing the shape of *ECR* signal. Analytically, during the first half of a dissolve number of input edges will be greater than output. The opposite stands for the second half. However, as also mentioned in [11], many dissolves did not show the characteristic behavior described by Zabih et al, especially in cases where the dissolves had long duration. This criterion is also sensitive to motion, despite motion compensation. The threshold defined in [3] is set to 0.05 and is manually chosen so as to maximize the precision rate. In Fig. 1 the *ECR* signal is plotted for 1000 frames of a test video. Using the selected threshold (0.05) the method of [3] cannot detect all gradual transition areas, such as those appearing in frame ranges [1 30] and [80 100] of the same video (Fig. 2(a)-(b)). Since the duration of the above dissolve areas is long, the number of entering and exiting edge pixels does not change rapidly. On the other hand, due to camera and object motion in frame range [8504 8525] new entering edges appear at a high rate and this frame range is falsely identified as representing a gradual transition (Fig. 2(c)).

The second examined approach [6] is based on the analysis of a signal computed from the standard deviation of pixel intensities. Dissolve areas correspond to *U* shaped regions of



(a)



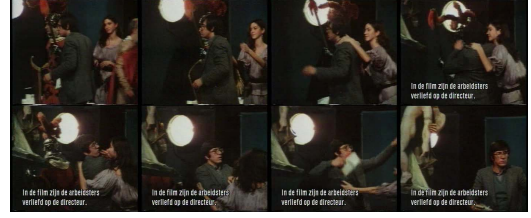
(b)



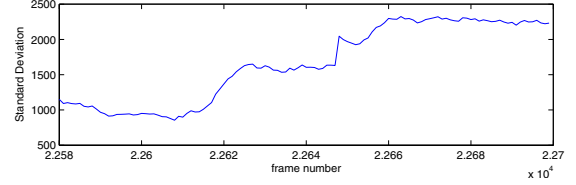
(c)

**Fig. 2.** (a) & (b) Examples of true gradual transitions whose duration is long, resulting in the number of entering and exiting edge pixels not changing rapidly over time. As a result these areas are not detected by [3]. (c) Example of non-transition area where, due to camera and object motion, new entering edges appear at a high rate. This area is wrongly identified as a transition area by [3].

this signal. One disadvantage of this approach is the need for manual selection of correlation parameter  $c$ . A high value of  $c$  would make this criterion sensitive to motion events. On the other hand a small value of  $c$  would allow the detection of dissolves only in the case where two composing frames of the dissolve have very low correlation, i.e. are almost completely different. Our experiments were conducted with  $c = 0.5$ , as suggested in [3]. With this parameter, only dissolves with correlation between starting and ending transition frames up to 0.5 are correctly detected. Some dissolve areas, such as the one in frame range [1 30] with correlation between composing frames more than 0.5, cannot be detected. On the other hand, motion in area [22608 22626] caused changes in intensity variance analogous with those caused from a dissolve



(a)



(b)

**Fig. 3.** (a) Example of non-transition area exhibiting local object motion (frame range [22608 22626]), (b) corresponding standard deviation curve of [6]. Based on the latter the frames of (a) are falsely identified as belonging to a transition area.

transition and was falsely identified as such (Fig. 3(a)-(b)).

The Effective Average Gradient criterion [7] counts the average magnitude of intensity gradient per pixel. Let

$$G^2[I(x, y, t)] = G_x^2[I(x, y, t)] + G_y^2[I(x, y, t)]$$

be the gradient magnitude at pixel  $(x, y)$  of frame  $I_t$ . Then  $EAG = \frac{TG}{TP}$ , where

$$TG = \sum_{x,y} G(x, y)$$

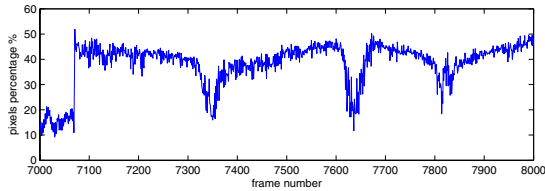
is the total magnitude value of the gradient image and  $TP = \sum I(x, y)$ . Dissolves are areas which correspond to  $U$  shaped regions, as in the standard deviation of pixel intensities criterion. In our experiments, this algorithm was shown to be very sensitive to motion events (about 80% of detections where areas with motion activity).

Monotonous Intensity Change [8] has very high recall rate but an average precision rate caused from local or global motion events. Threshold was manually set to 0.5, so as to maximize precision rate. Monotonous Intensity Change values for frames in the range [7025 8000] are plotted in Fig. 4(a). Although there is not dissolve transition but only object and camera motion, this area is misclassified as a dissolve Fig. 4(b).

In contrast to existing methods, using the proposed meta-segmentation approach dissolve areas [1 30] and [80 100] are correctly identified. The Macbeth color histogram descriptor is shown to be more robust to object and camera motion events than the luminance center of gravity. For example, frame ranges [639 652] and [716 728] correspond to object



(a)



(b)

**Fig. 4.** (a) Example of non-transition area exhibiting global (camera) motion (frame range [7025 8000]), (b) corresponding Monotonous Intensity Change curve of [8]. Due to the camera motion the Monotonous Intensity Change curve exhibits large values that would indicate the presence of a gradual transition.

motion events; in these ranges, the Luminance center of gravity descriptor shows large picks, while the Macbeth color histogram is not influenced (Fig. 5(a)-(c)).

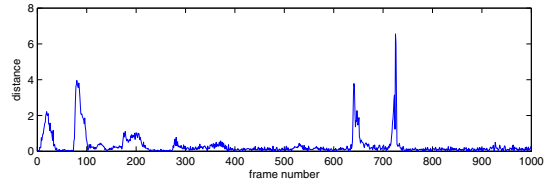
The combination of the proposed features leads to a very powerful gradual transition detection scheme as can be observed from the presented experimental results. The Macbeth color histogram descriptor is very useful in order to achieve more robustness to motion activity, while the rest can easily detect motion and transition events. The precision rate was increased by over 25%, while the recall rate achieved the best proposed performance over all approaches examined in this work.

## 5. CONCLUSIONS

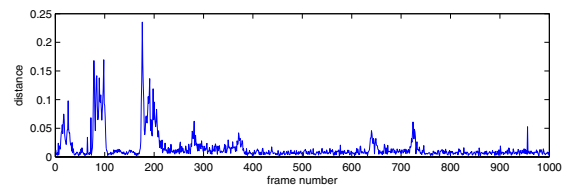
A new gradual transition detection scheme was proposed. New features were introduced showing less sensitivity to global or local motion effects. These features, each of which could serve as a stand-alone transition detection approach, were then combined using a machine learning technique, to result in a meta-segmentation scheme. We compared the performance of our algorithm with 4 other popular algorithms. The experimental results show that our algorithm has significantly high performance. Precision rate had an increase of over 25%, without decreasing the recall rate. Future work involves the integration in the meta-segmentation scheme of additional features that can contribute to further improvements in accu-



(a)



(b)



(c)

**Fig. 5.** (a) Frame range [1 1000], (b) curve of the distances between couples of consecutive frames based on the Luminance Center of Gravity and (c) on the Macbeth color histogram. The Macbeth color histogram descriptor is more robust to object and camera motion events than the luminance center of gravity, as indicated by the high peaks in frame ranges [639 652] and [716 728] of curve (b) in contrast to curve (c); both these frame ranges do not correspond to true gradual transitions.

rate gradual transition detection.

## 6. REFERENCES

- [1] G. C. Chvez, M. Cord, S. Philip-Foliguet, F. Precioso, and A. de A. Arajo, "Robust scene cut detection by supervised learning," *EUSIPCO*, 2006.
- [2] A.M. Alattar, "Detecting and compressing dissolve regions in video sequences with a dvi multimedia image compression algorithm," *IEEE International Symposium on Circuits and Systems (ISCAS)*, vol. 1, pp. 13–16, 1993.
- [3] R. Zabih, J. Miller, and K. Mai, "A feature-based algorithm for detecting and classifying production effects," *Multimedia Systems*, vol. 7, no. 2, pp. 119–128, 1999.

- [4] A. Miene, A. Dammeyer, T. Hermes, and O. Herzog, "Advanced and adaptive shot boundary detection," *ECDL WS generalized Documents*, 2001.
- [5] J.S. Boreczky and L. A. Rowe, "Comparison of video shot boundary detection techniques," *IST SPIE Symposium Proceedings, Storage and Retrieval for Image and Video Databases IV*, vol. 2670, January 1996.
- [6] J.W., Y.S. Chung, I.S. Kim, J.G. Choi, and K.H. Park, "Correlation based video-dissolve detection," *Information Technology: Research and Education, 2003. Proceedings. ITRE2003. International Conference on*, pp. 104 – 107, August 2003.
- [7] H.B. Lu, Y.G. Zhang, and Y.R. Yao, "Robust gradual scene change detection," *Image Processing, 1999. ICIP 99. Proceedings. 1999 International Conference on*, vol. 3, pp. 304 – 308, 1999.
- [8] C.W. Su, H.Y.M. Liao, H.R. Tyan, K.C. Fan, and L.H. Chen, "A motion-tolerant dissolve detection algorithm," *IEEE transactions on multimedia*, vol. 7, no. 6, pp. 1106–112, December 2005.
- [9] A. Nagasaka and Y. Tanaka, "Automatic video indexing and full-video search for object appearances," *Visual Database Systems II Elsevier Science Publishers*, pp. 113–127, 1992.
- [10] C. McCamy, H. Marcus, and J. Davidson, "A color rendition chart," *Journal of Applied Photographic Engineering*, vol. 2, no. 3, pp. 95–99, 1976.
- [11] R. Lienhart, "Comparison of automatic shot boundary detection algorithms," *IST SPIE Symposium Proceedings, Storage and Retrieval for Image and Video Databases VII*, vol. 3656, pp. 290–301, 1999.