Robust Gradual Scene Change Detection

H.B. Lu, Y.J. Zhang, Y.R. Yao
Department of Electronic Engineering,
Tsinghua University, Beijing 100084, China
zhangyj@ee.tsinghua.edu.cn

Abstract

Content-based temporal sampling of video is an efficient method for representing the visual information contained in the video sequence by using only a small subset of the video frames, which can be detected by a scene change detection method. Scene changes include not only the abrupt transitions, but also gradual transitions such as fade, dissolve. Robust detection of gradual transitions has been considered to be difficult. A high-performance gradual scene change detection method is presented in this paper. The proposed method is based on the video edit model. It can be derived that, for dissolve transition without violent motions, the variance, gradient magnitude and double chromatic difference (DCD) of image sequence during dissolve will show parabolic shapes. These features can be used for robust detection of dissolve transition. We approve that in term of recall rate the variance and gradient magnitude have the same efficiency. So, we simply use variance sequence to select possible positions of dissolve transitions and then DCD is used for further confirmation.

1. Introduction

Automatic video partitioning is a prerequisite for organizing and indexing video sources. Many methods have been proposed on the abrupt scene change detection by using pixel-wise difference, histogram comparison, edge comparison, and motion information. These methods deal with abrupt scene changes fairly well. A comparison of these algorithms can be found in [1]. However, many video programs, especially movies, contain a large number of gradual transitions such as fade-in, fade-out, dissolve. In order to properly segment the program into individual shots, the detection of such gradual transitions is required. Relatively fewer works on the detection of gradual transitions have been reported [2-4]. Robust detection of gradual transitions is still an open issue.

Zhang [2] proposed a twin-comparison algorithm for solving the problem of searching the positions of gradual transition. But camera or object motions may result in a sustained increase in the image disparity values same as

gradual changes and cause false detection. Yeo [3] suggested to look for the "plateau" in the curve of spatial difference with every K frames. A "plateau" indicates that a gradual scene change takes place. This method cannot distinguish dissolve from camera motion. Zabin [4] used edge tracking to detect gradual transitions. It requires global motion compensation before computing dissimilarity. Low precision rate and time-consuming computation are the drawbacks of this method.

In this paper, we present our method of fade and dissolve detection, which performs on the DC sequence extracted from MPEG stream [3]. The video edit model is first presented and then some features derived from this model are discussed. An efficient method is proposed and tested with real video sequences.

2. Gradual Transition Detection

A block diagram for shot detection is shown in Figure 1. In this diagram, the abrupt scene changes are first detected using the method proposed in [5]. Then, we calculate variance of image sequence for detection of fade-in and fade-out transitions. Possible dissolve positions are picked out by only searching the parabolic shapes in variance sequence that are not corresponding to the fade transition positions. DCD feature is applied here to confirm the real dissolve positions from those possible positions as it can identify dissolve from camera motion such as wipe, pan and zoom.
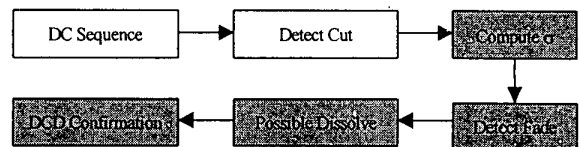

Figure 1. Block diagram for shot detection

2.1 Video edit model

We model gradual transition: fade-in, fade-out, and dissolve as follows:

$$f(x,y,t) = \alpha(t)p(x,y,t) + \beta(t)q(x,y,t) , \quad 0 \le t \le T \qquad (1)$$

where $\alpha(t)$ is a decreasing function during the gradual scene change with $\alpha(0)=1$ and $\alpha(T)=0$; and $\beta(t)$ is a increasing function with $\beta(0)=0$ and $\beta(T)=1$.

For fade-in, we have $\alpha(t)=0$ and (1) becomes

$$f(x,y,t) = \beta(t)q(x,y,t), \quad 0 \le t \le T \qquad (2)$$

For fade-out, we have $\beta(t)=0$ and (1) becomes

$$f(x,y,t) = \alpha(t)p(x,y,t), \quad 0 \le t \le T \qquad (3)$$

We notice that an image may fade-out to black, white, or other pure color and can also fade-in from any pure color. So some fade transitions will be missed using the method under the assumption of fading to and from black.

One can easily see that the variance of image at the end of fade-out and at the beginning of fade-in will be zero. It is a very important feature that is used in our method to achieve robust detection of fade. We design the fade detector to find the position where the variances are near zero. If the value of variance at the right side of this position increases for several successive frames, a fade-in is declared. If the value of variance at the left side of this position decreases for several successive frames, a fade-out is declared.

In the following discussion, two assumptions are made.

$$\alpha(t) + \beta(t) = 1 \qquad (4)$$

$$p(x,y,t) = p(x,y) \text{ and } q(x,y,t) = q(x,y) \qquad (5)$$

The first assumption (4) does not consider $\alpha(t)$, $\beta(t)$ as a piece-wise linear function, while most existing methods use the following linear functions to detect gradual transitions.

$$\alpha(t) = 1 - t/T \text{ and } \beta(t) = t/T \qquad (6)$$

But it is not a good choice since $\alpha(t)$ and $\beta(t)$ during most dissolve and fade transitions do not satisfy linear requirement. Equation (6) is just the special case of (4). In our method, it is only required that $\alpha(t)$ is a decreasing function.

The second assumption (5) is that during those transitions no violent object and camera motions happen. In fact, most gradual transitions satisfy this assumption and as we perform detection on DC sequence, image smoothing can compensate for minor camera or object motions.

2.2 Variance sequence

Considering equations (1)(4)(5), the mean of image sequence during dissolve can be expressed as

$$E(f(t)) = \alpha(t)E(p) + \beta(t)E(q)$$
$$= \alpha(t)[E(p) - E(q)] + E(q) \qquad (7)$$

and variance as

$$\sigma^2(f) = E(f - \bar{f})^2$$
$$= E[\alpha(t)p + \beta(t)q - \alpha(t)\bar{p} - \beta(t)\bar{q}]^2$$
$$= \alpha^2(t)\sigma^2(p) + \beta^2(t)\sigma^2(q) + 2\alpha(t)\beta(t)E[(p - \bar{p})(q - \bar{q})] \quad (8)$$
$$\approx \alpha^2(t)\sigma^2(p) + \beta^2(t)\sigma^2(q)$$
$$= [\sigma^2(p) + \sigma^2(q)]\alpha^2(t) - 2\sigma^2(q)\alpha(t) + \sigma^2(q)$$

Equation (8) shows that the variance sequence curve is a parabolic one. We can obtain its local minima at

$$\alpha(t) = \frac{\sigma^2(q)}{\sigma^2(p) + \sigma^2(q)} \qquad (9)$$

2.3 Gradient magnitude sequence

The gradient magnitude $G(x,y)$ at pixel (x,y) can be computed with:

$$G^2[f(x,y)] = G_x^2[f(x,y)] + G_y^2[f(x,y)] \qquad (10)$$

Now, using video edit model (1), we can wirte

$$G_x[f(x,y)] = \alpha(t)G_x[p(x,y)] + \beta(t)G_x[q(x,y)]$$
$$G_y[f(x,y)] = \alpha(t)G_y[p(x,y)] + \beta(t)G_y[q(x,y)] \qquad (11)$$

Let $TG^2(f) = \sum_{x,y} G^2[f(x,y)]$, then

$$TG^2(f) = \sum_{x,y} \{[\alpha(t)G_x(p) + \beta(t)G_x(q)]^2$$
$$+ [\alpha(t)G_y(p) + \beta(t)G_y(q)]^2\} \qquad (12)$$

Considering $\sum_{x} G_x(p)G_x(q) \approx 0, \sum_{y} G_y(p)G_y(q) \approx 0$, we can get

$$TG^2(f) \approx [TG^2(p) + TG^2(q)]\alpha^2(t)$$
$$- 2TG^2(q)\alpha(t) + TG^2(q) \qquad (13)$$

Clearly, the gradient magnitude of image sequence during dissolve also show parabolic shape, the local minima is at

$$\alpha(t) = \frac{TG^2(q)}{TG^2(p) + TG^2(q)} \qquad (14)$$

Comparing equation (8) with equation (13), we can see that during dissolve transition both variance and gradient magnitude always show parabolic shapes, as depicted in Figure 2, which is an important feature that can be used to select possible dissolve transition positions with high recall rate. In our experiment, we also find that the curve shapes of variance and gradient magnitude sequence are similar.

We choose variance of image sequence in our method instead of edge magnitude because we detect gradual transition on DC sequence, which is a low-frequency

version of original image and is not suitable for edge detection. Another reason is that edge calculation is time-consuming.
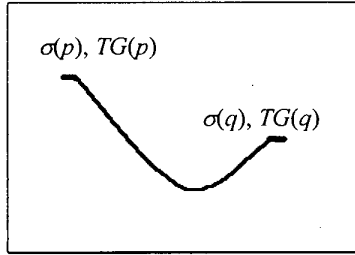

Figure 2. Variance or gradient magnitude of image sequence shows parabolic shape during dissolve

Clearly, when variance or gradient magnitude of image p or q is low, the valley of parabola becomes less distinct. One example is shown in Figure 3. There are two dissovle transitions existing in the video clip. Both parabolic valleys of this variance sequence are not distinct, Figure 4, while those of EAG (discussed in the following paragraph) sequence distinct. But, because parabolic shape always exits when a dissolve takes place and the local minima always locates at somewhere on this piece of parabola, such positions will also be selected using variance sequence. That is, the recall rate will not be affected.
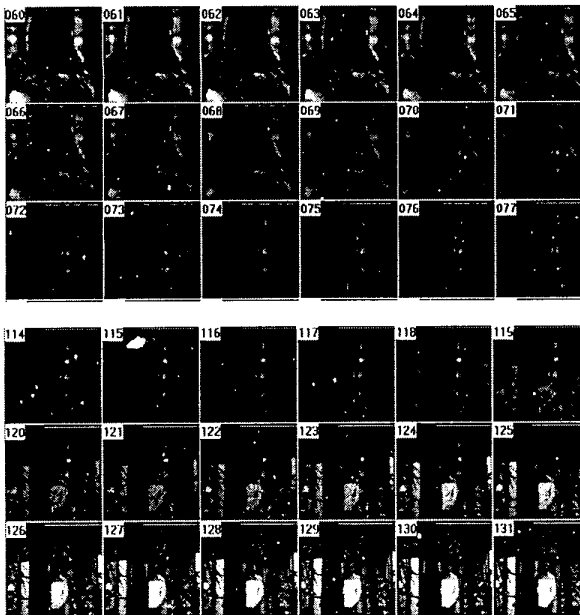

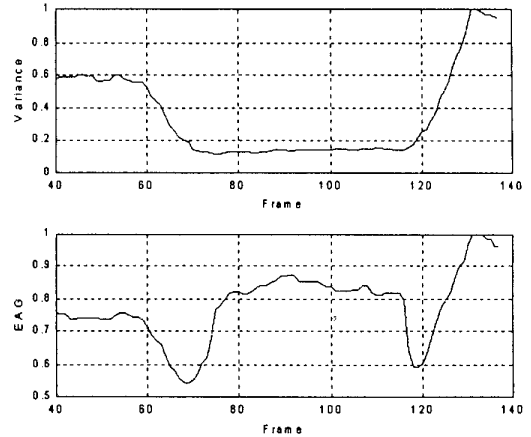Figure 3. Two dissolve transitions in "Casablanca"

Figure 4. Variance and EAG sequence of Figure 3

It is notable that effective average gradient (EAG) [6] can be used to the same purpose. EAG is defined by the following equation:

$$EAG = TG / TP \tag{15}$$

where

$$TG = \sum_{x,y} G(x,y) \tag{16}$$

is total magnitude value of the gradient image, and

$$TP = \sum_{x,y} p(x,y) \tag{17}$$

is total number of pixels with non-zero gradient values, as $p(x,y)$ is defined by

$$p(x,y) = \begin{cases} 1 & \text{if } G(x,y) > 0 \\ 0 & \text{if } G(x,y) = 0 \end{cases} \tag{18}$$

Clip transformation can be used to reduce the influence of noise. Given an original image $f(x,y)$ and a gray-level L, a transformed image $f_L(x,y)$ can be obtained by using the following clip transformation:

$$f_L(x,y) = \begin{cases} f(x,y) & \text{if } f(x,y) > L \\ L & \text{if } f(x,y) \le L \end{cases} \tag{19}$$

From the transformed image $f_L(x,y)$, a gradient image $g_L(x,y)$ can be obtained. Then the EAG can be calculated by using $g_L(x,y)$.

In most cases, the valley of parabola becomes more distinct in EAG sequence than in variance and gradient magnitude sequences, as shown in Figure 4. Since the variance sequence is efficient enough to select possible dissolve position and DC sequence is not suitable for edge detection, EAG has not been used in the following.

306

2.4 Double chromatic difference

Another problem in dissolve detection is that the large object motion and camera motion will introduce similar parabola shape and cause false alarm. We need confirmation by means of DCD test proposed by Yu [7]. The feature can identify dissolve from zoom, pan and wipe.

$$DCD(t) = \sum_{x,y} F(|\frac{f(x,y,t_0) + f(x,y,t_N)}{2} - f(x,y,t)|) \quad (20)$$

where $t_0 \le t \le t_N$, t_0 and t_N are two points during dissolve period, $0 \le t_0 < t_N \le T$. $F(\bullet)$ is a thresholding function. For convenience, let threshold to be zero. We have the following results using equation (1) and assumptions (4) and (5).

$$DCD(t) = \sum_{x,y} |\frac{f(x,y,t_0) + f(x,y,t_N)}{2} - f(x,y,t)|$$

$$= \sum_{x,y} |\frac{\alpha(t_0) + \alpha(t_N)}{2} - \alpha(t)| \times |p(x,y) - q(x,y)| \quad (21)$$

$$= |\frac{\alpha(t_0) + \alpha(t_N)}{2} - \alpha(t)| \times \sum_{x,y} |p(x,y) - q(x,y)|$$

Because $\alpha(t)$ is a decreasing function, $DCD(t)$ is approximately a parabola. Under the assumption of $\alpha(t) + \beta(t) = 1$, there always exits a frame t_m, $t_0 < t_m < t_N$, where

$$\frac{f(x,y,t_0) + f(x,y,t_N)}{2} = f(x,y,t_m) \quad (22)$$

i.e., $DCD(t_m) = 0$.

From equation (21), we can see that for any t_0, t_N satisfying $0 \le t_0 < t_N \le T$, $DCD(t)$ will always show approximate parabolic shape. That is, the positions of starting point and ending point of dissolve are not essential in DCD calculation. Actually, it is difficult to find out starting point and ending point of dissolve accurately.

On the other hand, since local object or camera motion is inevitable in movie production, there will be relatively large motion cumulation between the actual starting and end points of dissolve for those dissolves with long duration.

So we choose t_0, t_N close to each other in the interval [0,T], as shown in Figure 5. This means can compensate for motion cumulcation to a certain extent. In Figure 5, there are three possible dissolves in the curve of variance, indicated by number 1, 2, 3. The first two are the same as those shown in Figure 4. Black dots are the position of local maximum or minimum; Solid arraws point to the position of t_0 selected and dashed t_N.
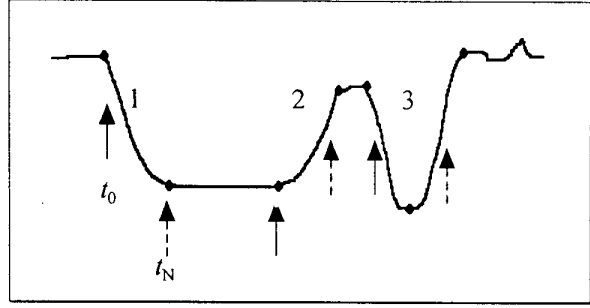

Figure 5. Select positions of t_0 and t_N

3. Experiment Result

Several video sequences containing 12 fade-in, 11 fade-out are used to test the efficiency of fade detector. Perfect recall and precision rates are achieved in our experiment. One fade-in clip is shown is Figure 6 and its variance and EAG sequence Figure 7. We can see that the variances of frame 0-3 are zero but the EAG have relatively big values. Because EAG decreases from the frame 4 on, this clip will not be marked as fade-in. So, EAG is not suitable for fade detection.
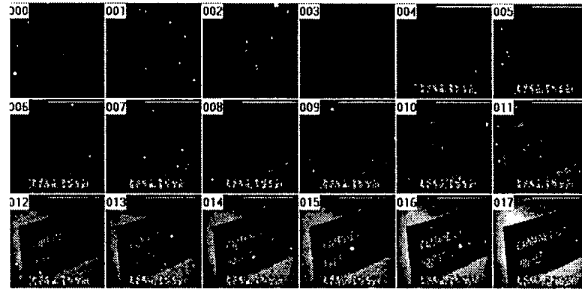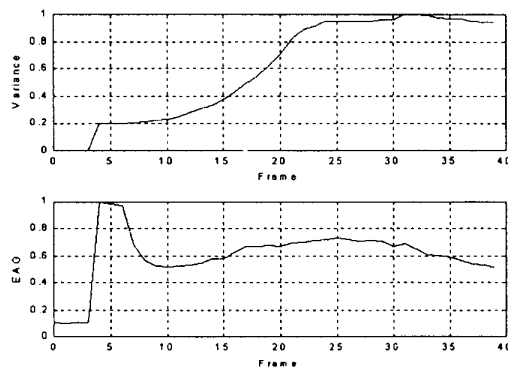

Figure 6. Video clip of fade-in

Figure 7. Variance and EAG sequence of Figure 6

To detect dissolve, we select ten sequences from movie "Casablanca". Each sequence contains several abrupt scene changes and one dissolve. One sequence from movie "CONAN- the barbarian" containing 9 dissolves and 5 abrupt changes is also selected. There are a number of camera transitions (pan, zoom) and object motions in these sequences. Two dissolve clips are shown in Figure 8 and Figure 9 as examples. The statistics of dissolve detection are listed in Table 1. The results are quite satisfactory.
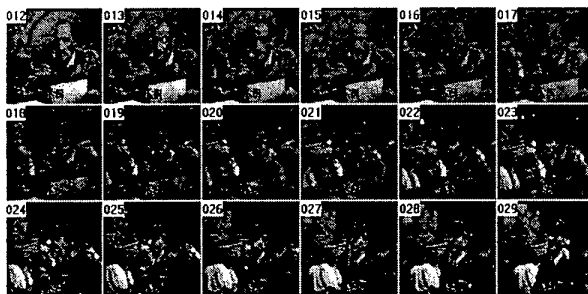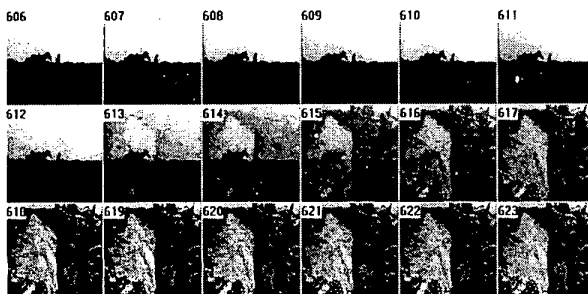

Figure 8. Dissolve clip from movie "Casablanca"

Figure 9. Dissolve clip from "CONAN- the barbarian"

Table 1. Experimental results of dissolve detection

Movie	"Casablanca"	"CONAN"
Total	10	9
Detected	11	10
Correct	9	9
Recall	90%	100%
Precision	81%	90%

4. Conclusion

A gradual scene change detection method based on video edit model is discussed and tested for DC sequences. Several image features: the variance, gradient magnitude, effective average gradient and double chromatic difference (DCD) of image sequence derived from video edit model are discussed and compared. The variance and DCD are used, the former for possible dissolve position detection and the latter, which can identify camera motion and object motion from gradual transition, for confirmation. Experiments show the method is fast and has high performance.

Acknowledgements

References

[1] J.Boreczky, L.Rowe, "Comparison of video shot boundary detection techniques", SPIE, Vol.2670, pp.122-128, 1996.

[2] H.J.Zhang, A.Kankanhalli, and S.W.Smoliar, "Automatic partitioning of full-motion video", ACM Multimedia Systems 1, 1993, pp.10-28.

[3] Boon-Lock Yeo, Bede Liu, "Rapid Scene Analysis on Compressed Video", IEEE Trans. On Circuits and Systems for Video Technology, Vol.5, pp533-544, 1995

[4] Ramin Zabih, Justin Miller, Kevin Mai. "A feature-based algorithm for detecting and classifying scene video", Multimedia'95, pp. 189-200, 1995.

[5] H.B.Lu, Y.J.Zhang, "Detecting abrupt scene change using neural network", Prcceeding of Third International Conference, VISUAL'99, pp.291-298, 1999.

[6] Y.J.Zhang, J.J.Gerbrands, "Transition region determination based thresholding", Pattern Recognition Letters, 12, pp13-23, 1991.

[7] H. Yu, G.Bozdage, S.Harrington, "Feature-based video segmentation", ICIP'97, pp.498-501; 1997.