

# Adaptive Video-Dissolve Detection Method Based on Correlation Between Two Scenes

Jong-Un Won<sup>1</sup>, Jae-Gark Choi<sup>2</sup>, Yoon-su Chung<sup>3</sup> and Kil-Houm Park<sup>1</sup>

<sup>1</sup> School of Electrical Engineering and Computer Science Kyungpook National University, 1370 Sankyuk-Dong, Taegu 702-701, KOREA, Tel. +82-53-940-8849, Fax.: +82-53-940-5505

<sup>2</sup> Division of Computer Application Engineering, Dongeui University, Gaya-Dong, Busanjin-Ku, Busan, 614-714, KOREA

<sup>3</sup> Electronics and Telecommunications Research Institute, 161 Gajeong-dong, Yuseong-gu, Daejeon, 305-350, KOREA

e-mail : [bestwon@palgong.knu.ac.kr](mailto:bestwon@palgong.knu.ac.kr), [cjg@dongeui.ac.kr](mailto:cjg@dongeui.ac.kr), [voonsu@etri.re.kr](mailto:voonsu@etri.re.kr), [khpark@knu.ac.kr](mailto:khpark@knu.ac.kr)

**Abstract:** In this paper, we propose a new adaptive dissolve detection method based on the analysis of a dissolve modeling error that is the difference between an ideally modeled dissolve curve without any correlation and an actual variance curve with a correlation. The dissolve modeling error is determined based on a correlation between two scenes and variances for each scene. First, Candidate regions are extracted by using the characteristics of a parabola that is downward convex, then the candidate region will be verified based on a dissolve modeling error. If a dissolve modeling error on a candidate region is less than a threshold that is defined by a dissolve modeling error with a target correlation, the candidate region should be a dissolve region with a correlation less than the target correlation. The threshold is adaptively determined based on the variances between the candidate regions and the target correlation. By considering the correlation between neighbor scenes, the proposed method is able to be a semantic scene-change detector. The proposed algorithm was tested on various types of data and its performance proved to be more accurate and reliable when compared with other commonly used methods

## 1. Introduction

Scene change detection is important technology for video indexing, browsing and retrievals. Generally, scene changes include both abrupt transitions and gradual transitions, such as fade and dissolve. Many methods have already been proposed for abrupt scene change detection [5,6]. Whereas, there have been relatively few reported studies on detecting gradual transitions [1-4,6,7], even though many video sequences, especially movies, contain a lot of gradual transitions. Consequently, the robust detection of gradual transitions is still an open issue. Zhang *et al.* [1] proposed a twin-comparison algorithm to detect abrupt and gradual transitions from color histogram differences between successive frames. Although efficient, the selection of appropriate thresholds is usually application-dependent. Alattar [2] and Meng *et al.* [3] used variance as an indication for detecting a dissolve. In their cases, dissolves are ascertained from peaks and spikes using a threshold that is heuristically determined based on experience and the assumption that neighboring scenes are independent. However, in most real cases a certain correlation exists between different scenes, thereby affecting the detection of peaks and spikes in a dissolve region. Consequently,

dissolves can be missed in a video sequence that has a high correlation or low variance between adjacent scenes, plus scenes including object and camera motion can be falsely detected as a dissolve.

Accordingly, the current paper proposes a new adaptive dissolve detection method based on the analysis of a dissolve modeling error. The proposed algorithm consists of two steps. First, the candidate dissolve regions are extracted using the characteristics of the first and second derivative of a variance curve. In the second step, the candidate regions are verified based on a dissolve modeling error. If the dissolve modeling error for a candidate region is less than a threshold defined by a dissolve modeling error with a target correlation, the candidate region is determined as a dissolve region with a lower correlation than the target correlation, which can be given application-dependently by user or can be used as a control factor of video parsing. The proposed algorithm was tested on a variety of data and the performance proved to be more accurate and reliable when compared with other commonly used methods.

## 2. Dissolve model and its characteristics

We model a dissolve as following: Let  $p(x, y, t)$  be a previous scene,  $q(x, y, t)$  be a last scene and  $f(x, y, t)$  be the scene composed from  $p(x, y, t)$  and  $q(x, y, t)$  with a gradual transition between them. Assuming  $p(x, y, t)$  and  $q(x, y, t)$  are ergodic random processes, then they are represented as  $p(x, y)$  and  $q(x, y)$ , respectively, therefore their variances  $\sigma_p^2(t)$  and  $\sigma_q^2(t)$  become constants. This dissolve region starts at  $t_1$  and ends at  $t_2$ . The dissolve model is expressed as the following equation

$$f(x, y, t) = \begin{cases} p(x, y) & t < t_1 \\ \alpha(t)p(x, y) + \beta(t)q(x, y) & t_1 \leq t \leq t_2 \\ q(x, y) & t > t_2 \end{cases} \quad (1)$$

where  $x$ ,  $y$  and  $t$  are continuous variables that represent the horizontal, vertical and temporal dimensions, respectively.

The parameter  $\alpha(t)$  is set equal to  $(t_2 - t)/(t_2 - t_1)$  and is a decreasing function during the gradual scene change. Also the parameter  $\beta(t)$  is an increasing function within the

dissolve region and is defined as  $(t_2 - t)/(t_2 - t_1)$ . The sum of two parameters  $\alpha(t)$  and  $\beta(t)$  is always one. The variance of  $f(x, y, t)$  within the dissolve region can be expressed as the following equation

$$\begin{aligned} \sigma_f^2(t) &= E[(f - \bar{f})^2] \\ &= E[(\alpha(t)p(x, y) + \beta(t)q(x, y) \\ &\quad - \alpha(t)\overline{p(x, y)} - \beta(t)\overline{q(x, y)})^2] \\ &= \alpha^2(t)\sigma_p^2 + \beta^2(t)\sigma_q^2 \\ &\quad + 2\alpha(t)\beta(t)E[(p(x, y) - \overline{p(x, y)})(q(x, y) - \overline{q(x, y)})] \end{aligned} \quad (2)$$

If  $p(x, y)$  and  $q(x, y)$  are assumed to be statistically independent with variances  $\sigma_p^2$  and  $\sigma_q^2$ , respectively, then the covariance between  $p(x, y)$  and  $q(x, y)$  is zero. Therefore Eq. (2) is approximated as following

$$\begin{aligned} \sigma_f^2(t) &\approx \alpha^2(t)\sigma_p^2 + \beta^2(t)\sigma_q^2 \\ &= [\sigma_p^2 + \sigma_q^2]\alpha^2(t) - 2\sigma_q^2\alpha(t) + \sigma_q^2 \end{aligned} \quad (3)$$

Eq. (3) shows that the variance  $\sigma_f^2(t)$  for dissolve is parabola that is convex downward. From Eq. (3), at the center of dissolve  $\alpha(t) = 0.5$ , the variance of ideal parabola model  $\tilde{\sigma}_{center}^2$  is defined as following

$$\tilde{\sigma}_{center}^2 = \frac{\sigma_p^2 + \sigma_q^2}{4} \quad (4)$$

The variance curve  $\sigma_f^2(t)$  of ideal dissolve model and its characteristics are affected by variance of two scenes and also it includes the approximation error that means the third term of Eq. (2) since the parabola model is based on approximation model expressed by Eq. (3). Therefore to detect dissolve exactly, we should consider the approximation error.

### 3. Effect of correlation

This section demonstrates the effect of a correlation between neighbor scenes in a dissolve model. Early research detected a dissolve using the characteristics of a parabola based on an ideal model without any correlation between neighbor scenes. However, in real sequences, there is often a correlation between neighbor scenes, which affects the dissolve detection. As such, this correlation must be considered for the precise detection of a dissolve.

When an ideal dissolve model curve is given for the region  $[p, q]$ , the difference between the actual variance curve and the ideal dissolve model curve is called the dissolve modeling error. Let  $\sigma_{real}^2(t)$  be the actual variance curve including a correlation and  $\sigma_{ideal}^2(t)$  be the ideal dissolve model curve without any correlation in the region  $[p, q]$ . The actual variance curve can be expressed by Eq. (2), while the ideal dissolve model is given from Eq. (3). As a result, the dissolve modeling error can be given by Eq. (5)

$$\begin{aligned} D_{pq}(t) &= 2\alpha(t)\beta(t)E[(p - \bar{p})(q - \bar{q})] \\ &= 2\alpha(t)\beta(t)\mu_{pq} \end{aligned} \quad (5)$$

where  $\mu_{pq}$  is the covariance between scene  $p$  and scene  $q$ .

The covariance  $\mu_{pq}$  is normalized by the standard deviations at  $p$  and  $q$ , and given by the following equation

$$\rho_{pq} = \frac{\mu_{pq}}{\sigma_p \sigma_q} \quad (6)$$

where  $\rho_{pq}$  is the covariance normalized by  $\sigma_p$  and  $\sigma_q$ , and  $0 \leq |\rho_{pq}| \leq 1$ . If Eq. (6) is substituted for Eq. (5), the dissolve modeling error can be represented as follows

$$D_{pq}(t) = 2\alpha(t)\beta(t)\sigma_p\sigma_q\rho_{pq} \quad (7)$$

The dissolve modeling error is proportional to the correlation. At the center of a dissolve,  $\alpha(t) = 0.5$ , the maximum dissolve modeling error  $D_{max}$  can be defined as follows:

$$D_{max} = \frac{\sigma_p\sigma_q\rho_{pq}}{2} \quad (8)$$

When the correlation is  $c$  in the region  $[p, q]$ , the maximum dissolve modeling error  $D_{max_c}$  becomes  $\frac{\sigma_p\sigma_q c}{2}$ . If the maximum dissolve modeling error  $D_{max}$  is less than  $D_{max_c}$  in any other region, this region can be identified as a dissolve with a correlation of less than  $c$ . Therefore, the maximum dissolve modeling error  $D_{max_c}$  with the correlation  $c$  becomes the threshold for detecting a dissolve with a correlation of less than  $c$ . As a result, the threshold is automatically determined by the characteristics of each region. For the region  $[p, q]$ , the threshold is given by Eq. (9)

$$D_{max_c} = \frac{\sigma_p\sigma_q c}{2} \quad (9)$$

where  $c$  is a target correlation. A target correlation can be given application-dependently by user or can be used as a control factor of video parsing. The maximum dissolve modeling error  $D_{max_c}$  with a target correlation  $c$  is in a word called a target modeling error.

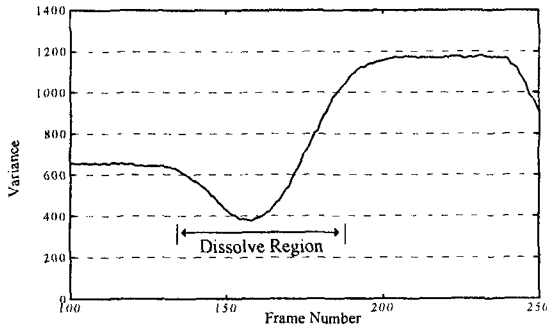
### 4. Adaptive dissolve detection algorithm

The proposed method consists of two steps. First, candidate dissolve regions are extracted based on the characteristics of the first and second derivatives of the

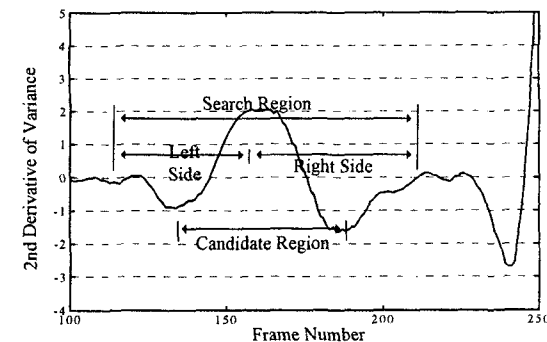
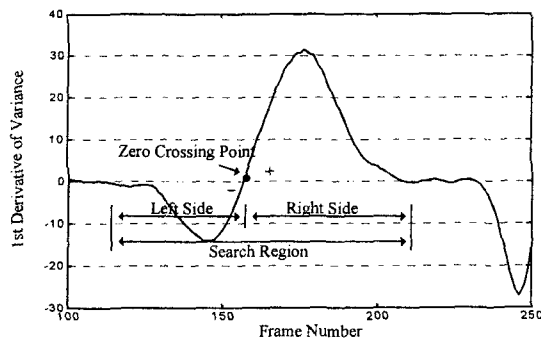
variance curve. In the second step, the candidate regions are then verified using the proposed adaptive threshold.

#### 4.1 Extraction of candidate dissolve regions

In this paper, a candidate dissolve region is identified using the characteristics of the first and second derivatives of the variance curve. Fig. 1 shows the procedure used to extract a candidate region. The candidate region extraction algorithm starts by identifying a search region in the first derivative of the variance curve. A candidate region is then defined by locating the largest negative spike within the search region.



(a) Variance curve



(c) Second derivative of variance curve

Fig. 1 Example of variance curve and its 1st and 2nd derivatives

To determine the search region in the first derivative of the variance curve, the zero crossing point from negative to positive is first identified and used as the center of the search region. Thereafter, the start point of the search region is determined as the first position where the value of

the first derivative is zero to the left of the zero crossing point as the center, while the end point of the search region is determined as the first position where the value of the first derivative is zero to the right of the zero crossing point. The region between the zero crossing point and the start point of the search region is referred to as the left side of the search region, and the region between the zero crossing point and the end point of the search region is referred to as the right side of the search region. Fig. 1 (b) shows the search region.

A candidate region is obtained from search region using the second derivative curve. The minimum valued position of the second derivative curve within the left side of the search region is set as the start point of a candidate region, while the end point of the candidate region is set as the minimum valued position of the curve within the right side of the search region. Fig. 1 (c) shows a candidate region.

#### 4.2 Verification of candidate region

As the candidate regions are only identified based on an analysis of the characteristics of the first and second derivatives of the variance curve of a parabola corresponding to a dissolve, such candidate regions also may include parabolas corresponding to false dissolves caused by object and camera motion. Therefore, a parabola corresponding to a true dissolve should be further distinguished using other dissolve characteristics. Accordingly, the current paper verifies the candidate regions based on the modeling error introduced in section 3.

Fig. 2 shows a flow chart for verifying a dissolve. For each candidate region, the maximum dissolve modeling error  $D_{max\_c}$  between a dissolve model with a given target correlation  $c$  and an ideal dissolve model with no correlation is estimated by Eq. (9) with variances at the start and end points of each candidate region and the given target correlation  $c$ . Then  $D_{max\_c}$  becomes the threshold to verify each candidate region as a dissolve.

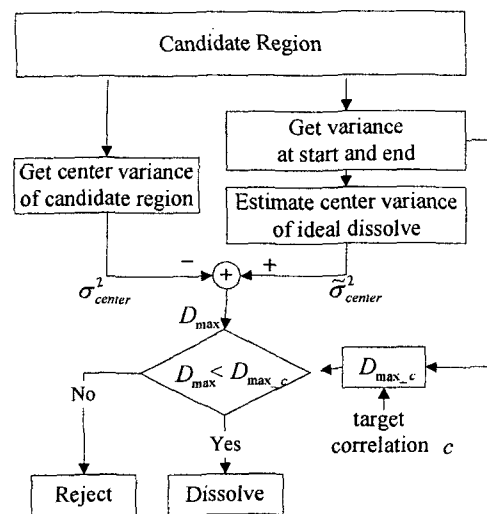


Fig. 2 Flow chart for verifying dissolve region

The maximum dissolve modeling error  $D_{max}$  in each candidate region is defined by the difference between the variance  $\sigma_{center}^2$  at the center of each candidate region and the variance  $\tilde{\sigma}_{center}^2$  at the center of an ideal dissolve model estimated by Eq. (4). If the maximum dissolve modeling error  $D_{max}$  in the current region is less than the target modeling error  $D_{max,c}$ , the region is determined to be a dissolve region.

### 5. Experimental Results

Several video image sequences of different characteristics were used to test the dissolve detector. Some of these sequences contain several genuine dissolve of various lengths, while others contain synthesized dissolve. The length and correlation of dissolves is 28 ~ 75 frames, and 0.01 ~ 0.42, respectively. Each frame of the sequence is in SIF(352×240) format.

Table 1. Dissolve detection relative to target correlation  $\rho$  (detection: O, failure: X).

Seq. Name	Dissolve region	$\rho$	Target correlation			
			0.15	0.25	0.35	0.45
Autumn	27 – 82	-0.01	O	O	O	O
	156 – 210	0.09	O	O	O	O
	324 – 368	0.42	X	X	X	O
Sunrise	12 – 87	-0.11	O	O	O	O
	135 – 186	-0.02	O	O	O	O
	242 – 270	0.16	X	O	O	O
	370 – 396	-0.09	O	O	O	O
Toy	137 – 164	0.06	O	O	O	O
	216 – 249	0.18	X	O	O	O
	366 – 393	0.32	X	X	O	O

Table 1 shows the result of dissolve detection according to a correlation  $\rho$  between neighbor scenes. When a correlation was varied from 0.1 to 0.4, all regions with smaller correlation than a given target correlation were completely detected, but other regions with larger correlation than given target correlation were not detected. When a correlation between two scenes is larger than a target correlation, the modeling error between an ideal dissolve model curve and an actual dissolve curve is to be larger than the modeling error between an ideal dissolve model and a dissolve model with a target correlation. If target correlation is 0.3, in an autumn sequence, the third dissolve wasn't detected. However it is detected with target correlation 0.4. If target correlation is increased near 1, the detected regions may be not only a true dissolve region but also false dissolve region caused by the motion of camera or object. Therefore the target correlation must be set circumspectly for detecting dissolve exactly with considering the correlation.

Table 2 shows the experimental results of proposed method and conventional methods. Target correlation is set to 0.5. The total number of dissolve used in our experiment is 17 and is presented Total. The number of detected dissolve is presented Detect and the number of true dissolve among detected dissolve is presented Correct. Recall and Precision rates are for use in measurement to compare each method. Perfect Recall and Precision rate are achieved in our method.

Table 2. Experimental results for dissolve detection.

	Proposed method	Alattar's method	Meng's method
Total	17	17	17
Detected	17	20	26
Correct	17	16	15
Recall	100%	94.1%	88.2%
Precision	100%	80.0%	57.6%

### 6. Conclusions

In this paper, we have demonstrated the effect caused by a correlation between two scenes on dissolve regions and introduced the dissolve modeling error that is determined based on the correlation and variances for each region and was proportional to the correlation and variance. And then we proposed the adaptive dissolve detection method based on dissolve modeling error. The proposed method is able to reliably detect dissolves with considerable object motion because it is based on the dissolve characteristics. Results show that the proposed method is effective in detecting various dissolves regardless of variances and length of a dissolve and it was necessary that we must consider the correlation between two scenes on dissolve region.

### References

- [1] H. J. Zhang, A. Kankanhalli, and S. W. Smoliar, "Automatic partitioning of full-motion video", ACM Multimedia Systems 1, pp.10-28, 1993
- [2] A. M. Alattar, "Detecting and compressing dissolve regions in video sequences with DVI multimedia image compression algorithm," ISCAS, pp. 13-16, 1993.
- [3] J. Meng, Y. Juan, S. F. Chang, "Scene change detection in a MPEG compressed video sequence," IS&T/SPIE Symposium, Proceedings, vol. 2419, Feb. 1995.
- [4] Boon-Lock Yeo, Bede Liu, "Rapid Scene Analysis on Compressed Video", IEEE Trans. on Circuits and Systems for Video Technology, vol.5, pp.533-544, 1995
- [5] J.Boreczky, L.Rowe, "Comparison of video shot boundary detection techniques", SPIE, Vol.2670, pp.122-128, 1996
- [6] H.B.Lu Y.J.Zhang, Y.R.Yao, "Robust Gradual Scene Change Detection", IICIP 99. Proceedings, Vol.3, pp.304-308, 1999
- [7] R. Lienhart, "Reliable Dissolve Detection", Proc. SPIE 4315, pp.219-230, Jan. 2001