DETECTION OF WIPES AND DIGITAL VIDEO EFFECTS BASED ON A PATTERN-INDEPENDENT MODEL OF IMAGE BOUNDARY LINE CHARACTERISTICS

Kota Iwamoto and Kyoji Hirata

Common Platform Software Research Laboratories, NEC Corporation

ABSTRACT

This paper proposes detection of wipes and digital video effects (DVEs) in a video sequence based on a new pattern-independent model. This model is based on the characteristics of image boundary lines dividing the two image regions in the transitional frames. Wipes and DVEs are modeled as frame sequences where either (A) a single boundary line moves continuously in a time sequence, or (B) multiple boundary lines form a quadrilateral within a frame. The model is applied to the image boundary lines extracted from a video sequence to detect wipes and DVEs. An evaluation using news programs containing various patterns of wipes and DVEs shows that the proposed method achieves recall of 91.5% and precision of 60.7%, improving the conventional twin-comparison method by 29.6% in recall and 46.5% in precision.

Index Terms— wipe, DVE, gradual transition, video segmentation, image boundary line, optical flow

1. INTRODUCTION

Temporal segmentation of video sequences by their semantic structure is important for the efficient access of video contents. In edited video contents, wipes and digital video effects (DVEs), which are the gradual spatial transitions of a video scene, are often used at important semantic boundaries to emphasize semantic changes to viewers. For example, wipes are often used to imply changes in location or elapse in time. DVEs are often used to imply changes in topics, such as news stories in a news program. Almost all of the wipes and DVEs in Japanese news programs imply some kind of semantic boundary. Therefore, the detection of wipes and DVEs can be valuable for the semantic segmentation of video contents.

One basic method for detecting wipes and DVEs is to detect the gradual transition of visual features based on frame comparison. Zhang *et al*. [1] proposed a twin-comparison method [1], which uses two thresholds to detect transitions based on accumulated histogram differences of consecutive frames. Yeo *et al*. [2] proposed a method which detects plateau patterns generated by the differences of DC images. Although these methods can detect wide variety of wipe and DVE patterns, they also detect huge number of false positives caused by various motions in video sequences such as camera motions and object motions.

More sophisticated model-based methods specialized for the detection of wipes and DVEs have also been proposed [3]-[9]. Alattar *et al*. [3] proposed a method based on the statistical characteristics of intensity means and variances. Zabih *et al*. [4] proposed a method based on the spatial distribution of entering and exiting edge pixels. Fernando *et al*. [5], Yu *et al*. [6], and Wu *et al*. [7] proposed methods based on local image differences between consecutive frames. Ngo *et al*. [8] and Kim *et al*. [9] proposed methods based on edge patterns in spatio-temporal slice images. The problem of these conventional model-based methods is that they only cover limited patterns of wipes, mostly the major patterns such as vertical and horizontal wipes. They do not cover most of the numerous DVE patterns, which are much more difficult to detect because of the huge variety of patterns and heavy manipulation of images (details in chapter 2). DVEs account for more than half of all the spatial transitions in actual video contents such as news programs. Furthermore, DVEs are much more valuable than typical wipes since they usually imply topic changes. Therefore, the conventional model-based methods are inapplicable to actual video contents.

This paper proposes detection of wipes and DVEs based on a new pattern-independent model. This model is based on the characteristics of image boundary lines, which covers almost all of the wipe and DVE patterns contained in news programs.

2. WIPES AND DVES IN VIDEO CONTENTS

Wipes and DVEs are spatial gradual transitions in video sequences which are generated by video switchers in an editing process. The image regions of exiting and entering scenes coexist during the transitional frames. Wipes are transitions where the exiting video scene is removed to reveal the entering scene, while the image positions of both scenes are fixed. DVEs are transitions where the image of either the exiting or entering scenes are manipulated by shifting, scaling, rotation, shearing, projection, flip, twist, etc. Numerous DVE patterns can be generated by combining these manipulation operations (Fig. 1).

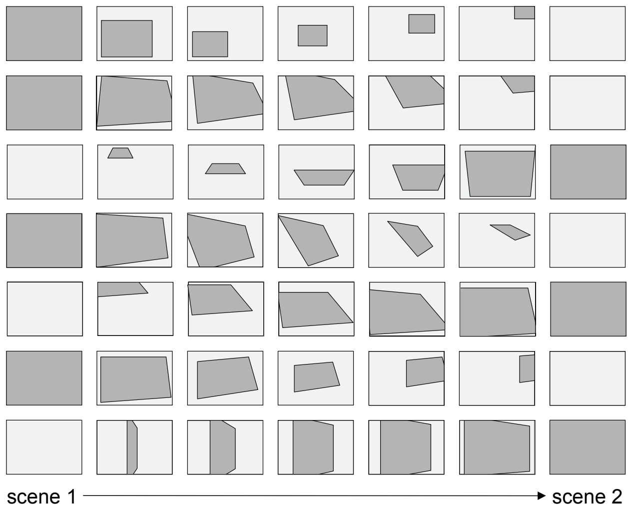


Fig. 1 Examples of various DVE patterns.

3. PATTERN-INDEPENDENT MODEL BASED ON IMAGE BOUNDARY LINE CHARACTERISTICS

The image boundary lines are lines formed on the boundary of the two image regions of the exiting and entering scenes in the transitional frames. They are formed by connected edges, and they move in a time sequence. Wipes and DVEs can be modeled as the frame sequence where the image boundary lines follow either of the following characteristics.

- **(A)** a single image boundary line moves continuously in a time sequence.
- **(B)** multiple image boundary lines form a quadrilateral within a frame.

Type (A) applies to majority of wipe patterns and some DVE patterns such as slide-ins, slide-outs and peels (Fig. 2). Type (B) applies mainly to DVE patterns (Fig. 3). The quadrilateral forms the rim of the manipulated image of either the exiting or entering scene. Since the borders of the manipulated image may lie on the exterior of the frame, image boundary lines do not always form a complete quadrilateral, but form the adjacent sides of the quadrilateral, as shown in the last two examples of Fig. 3. This model covers almost all of the wipe and DVE patterns in Japanese news programs.

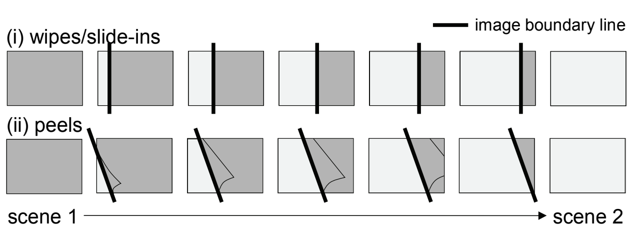


Fig. 2 Examples of type (A) characteristics.

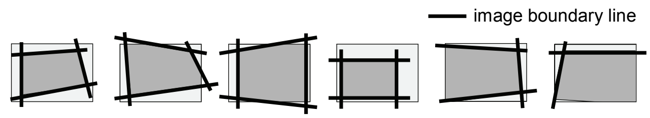


Fig. 3 Examples of type (B) characteristics.

4. DETECTION OF WIPES AND DVES BASED ON THE PROPOSED MODEL

Figure 4 shows the detection procedure of wipes and DVEs in a video sequence based on the proposed model. First, the video sequence pre-filtering detects the candidate frame sequences. Then, image boundary lines are extracted from each frame of the candidate sequences. Finally, the extracted image boundary lines are analyzed to detect the frame sequences where they satisfy the characteristics of type (A) or type (B).

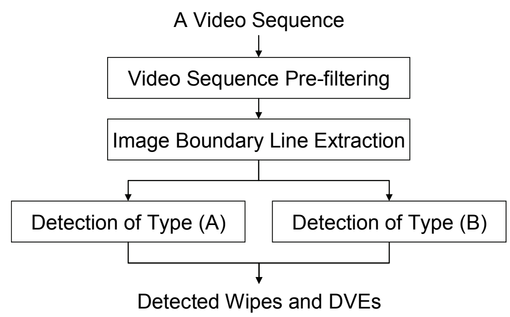


Fig. 4 Detection procedure of wipes and DVEs.

4.1. Video Sequence Pre-filtering

The pre-filtering suppresses the number of possible false positives and reduces the computation cost. It uses frame comparison of HSV histograms to detect the frame sequences of gradual transitions. By using two thresholds, it detects the frame sequences where the differences of consecutive frames exceed the lower threshold, while the difference of the frames before and after the sequence exceeds the higher threshold. The sequences including cuts are excluded. The thresholds are set at relatively low values so as not to eliminate any true sequences of wipes or DVEs.

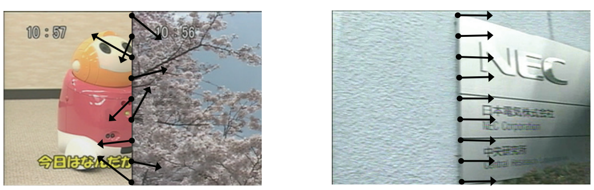
4.2. Image Boundary Line Extraction

Image boundary lines can be distinguished from lines formed by an object in an image based on analysis of optical flow. Since a true image boundary line lies on the border of two different images with independent motion, the optical flow on an image boundary line is undefined. Therefore, the estimated optical flow vectors of the multiple points on an image boundary line are more likely to diffuse in all directions. On the other hand, the optical flow vectors on an object line which moves in the image by camera motions or object motions are less likely to diffuse (Fig. 5). This assumption of optical flow diffusion is used to extract image boundary lines.

The extraction of image boundary lines is carried out in three steps i.e., edge pixel extraction, Hough transform, and line selection based on optical flow diffusion. First, pixel-wise frame difference extracts the changing regions, and the Canny edge detector [10] is applied to the regions to detect edge pixels. Then, the improved Hough transform proposed by O'Gorman *et al.* [11] is applied to the edge pixels to extract lines. The lines are represented in the parametric form $r = x\cos\theta + y\sin\theta$, where r is the perpendicular distance from the origin to the line and θ is the angle of this perpendicular with respect to the x-axis. Finally, for each extracted line, the optical flow vectors of the multiple points on the line are calculated using the block matching method. The diffusion of the optical flow vectors is calculated as the sum of the variances of horizontal and vertical components of the vectors. Let $\{(u_1,v_1),\dots,(u_N,v_N)\}$ denote the optical flow vectors, where N is the number of the vectors, and $(\bar{u},\bar{v})$ denote the mean vector. The diffusion D is calculated as,

$$D = \frac{1}{N}\sum_{n=1}^{N}(\bar{u}-u_n)^2 + \frac{1}{N}\sum_{n=1}^{N}(\bar{v}-v_n)^2 \quad . \tag{1}$$

If the diffusion D exceeds a certain threshold, the line is selected as an image boundary line. Figure 6 shows an example of the extracted image boundary lines.



(a) Optical flow vectors are more likely to diffuse on an image boundary line

(b) Optical flow vectors are less likely to diffuse on an object line

Fig. 5 Optical flow diffusion.

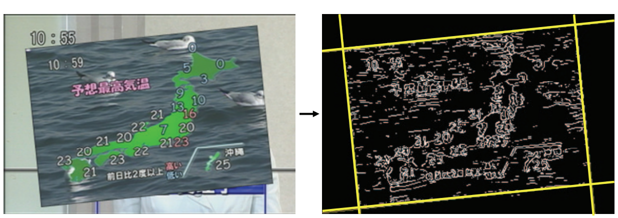


Fig. 6 Extracted image boundary lines (in bold lines).

4.3. Detection of Type (A)

To detect frame sequences where an image boundary line moves continuously in a time sequence, the extracted image boundary lines are tracked in the (r,θ) feature space, where each line is represented as a feature point in the space. The feature points of the moving image boundary line in a type (A) transition plot a continuous trajectory in the feature space. Figure 7 depicts the feature point trajectory of the moving image boundary line in a slide-out DVE. This continuous trajectory is tracked and detected as type (A) characteristics.

Each extracted image boundary line is projected to a feature point in the (r,θ) feature space, and these feature points are tracked by feature point prediction. An predicted feature point $(\hat{r}_i,\hat{\theta}_i)$ in frame i is calculated as $(\hat{r}_i,\hat{\theta}_i) = (2r_{i-1} - r_{i-2}, \theta_{i-1})$, where (r_{i-1},θ_{i-1}) and (r_{i-2},θ_{i-2}) denote the tracked feature points in the preceding two frames. The prediction error between the predicted feature point and the actual extracted feature point is calculated to determine the continuity of the tracking. If the length and the frame number of the tracked feature point trajectory exceed certain thresholds, the HSV histograms of the frames just before and after the trajectory are compared. If the difference exceeds a certain threshold, the frame sequence of the trajectory is detected as a transition satisfying the type (A) characteristics.

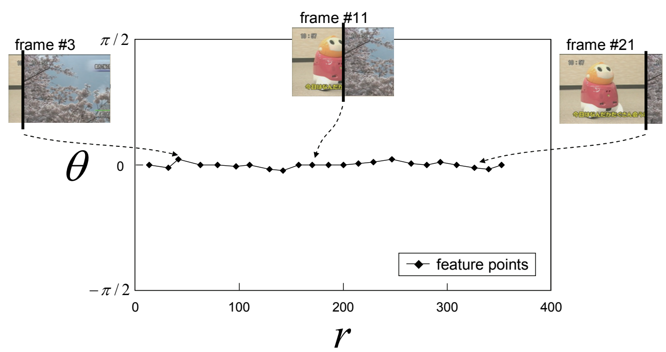


Fig. 7 Feature point trajectory of type (A).

4.4. Detection of Type (B)

To detect frame sequences where multiple image boundary lines form a quadrilateral within a frame, pairs of extracted image boundary lines forming the adjacent sides of the quadrilateral are searched within a frame. Square windows centered at the four corners of the frame are set as search regions. A pair of image boundary lines is searched which intersects within the search regions, while the intersecting angle and the orientations of the lines fall within predefined ranges. For the searched pair, an edge tracker is applied to track edges along both lines starting from the intersecting point (if the intersecting point lies on the exterior of the frame, start from the frame border). If edges continue for a certain length in both directions without breaking up, the pair is determined to be the adjacent sides of a quadrilateral. The frame sequence containing these pairs is detected as a transition satisfying the type (B) characteristics.

5. EVALUATION

The proposed method was evaluated using Japanese news programs. For comparison, the twin-comparison method [1] was also evaluated.

5.1. Evaluation Conditions

Table 1 lists the specifications of the video sequences used in the evaluation. They are composed of 14 different Japanese news programs from 5 different TV channels. They contain 59 wipes and 177 DVEs, which can be categorized into 41 different patterns.

The parameters of the proposed method were empirically set using training video sequences. The two thresholds for the twin-comparison method were also set using the sequences. Both methods used a post-filtering which filters out multiple outputs within a window of fixed time length to suppress successive false positives.

Table 1 Video sequences used in the evaluation.

# of video sequences	41 (from 14 different programs)
Total time length	15 hours 24 minutes
# of wipes and DVEs	236 (59 wipes, 177 DVEs)
# of different patterns	41
Coding properties	MPEG2, 352×240 pixels, 3Mbps

5.2. Results and Discussion

Recall and precision of the detection were evaluated. Their definitions are

$$\text{recall} = \text{TP}/(\text{TP} + \text{FN}), \tag{2}$$

$$\text{precision} = \text{TP}/(\text{TP} + \text{FP}), \tag{3}$$

where TP, FN, and FP denotes the number of true positives (correct detection), false negatives (miss-detection), and false positives (over-detection), respectively.

Table 2 shows the results. The proposed method achieved recall of 91.5% and precision of 60.7%. It improves the twin-comparison method by 29.6% in recall and 46.5% in precision. This result shows that the proposed method can detect wide variety of wipes and DVEs, while sufficiently suppressing the number of false positives. Although a large portion of the DVEs had a very fast and complicated movement of the manipulated images, the proposed method successfully detected those DVEs without causing much false positives. This is because the type (B) is a characteristic closed within a single frame and unaffected by the movement. The proposed method was successful in reducing the number of false positives since the characteristics of type (A) and type (B) are unique to wipes and DVEs. The analysis of optical flow diffusion also contributed to the reduction of false positives. The false positives were caused by moving object lines which the optical flow analysis failed to filter out and have the characteristics similar to the proposed model. The computation time was nearly equal to the time length of the video sequence when using a PC with 3.2GHz CPU, which shows that the proposed method is highly feasible for the detection of wipes and DVEs in actual video contents.

Table 2 Evaluation results.

Method	Recall			Precision
	Wipe	DVE	Total	
Twin-Comparison [1]	50.8%	65.5%	61.9%	14.2%[1]
Proposed	93.2%	91.0%	91.5%	60.7%

6. CONCLUSION

Detection of wipes and DVEs in a video sequence based on a new pattern-independent model has been proposed. The model is based on the characteristics of image boundary lines dividing the two image regions in the transitional frames. Wipes and DVEs are modeled as frame sequences where either (A) a single boundary line moves continuously in a time sequence, or (B) multiple boundary lines form a quadrilateral within a frame. The model is applied to the image boundary lines extracted from a video sequence to detect wipes and DVEs. An evaluation using news programs containing various patterns of wipes and DVEs shows that the proposed method achieves recall of 91.5% and precision of 60.7%, improving the conventional twin-comparison method by 29.6% in recall and 46.5% in precision.

REFERENCES

[1] H. J. Zhang *et al.*, "Automatic Partitioning of Full-Motion Video", Mutimedia Systems, Vol. 1, pp. 10-28, 1993.
[2] B.-L. Yeo *et al.*, "Rapid Scene Analysis on Compressed Video", IEEE Trans. on Circuits and Systems for Video Technology, Vol. 5, No. 6, pp. 533-544, 1995.
[3] A. M. Alattar, "Wipe Scene Change Detector for Use with Video Compression Algorithms and MPEG-7", IEEE Trans. on Consumer Electronics, Vol. 44, No. 1, pp. 43-51, 1998.
[4] R. Zabih *et al.*, "A Feature-Based Algorithm for Detecting and Classifying Production Effects", Multimedia Systems, Vol. 7, pp.119-128, 1999.
[5] W. A. C. Fernando *et al.*, "Wipe Scene Change Detection in Video Sequences", Proc. of ICIP'99, Vol. 3, pp. 294-298, 1999.
[6] H.-H. Yu *et al.*, "A Multi-resolution Video Segmentation Scheme for Wipe Transition Identification", Proc. of ICASSP'98, Vol. 5, pp. 2965-2968, 1998.
[7] M. Wu *et al.*, "An Algorithm for Wipe Detection", Proc. of ICIP'98, Vol. 1, pp. 893-897, 1998.
[8] C. W. Ngo *et al.*, "Detection of Gradual Transitions through Temporal Slice Analysis", CVPR'99, Vol. 1, pp. 36-41, 1999.
[9] H. Kim *et al.*, "Processing of Partial Video Data for Detection of Wipes", SPIE Storage and Retrieval for Image and Video Databases, Vol. 7, pp. 280-289, 1999.
[10] J. Canny, "A Computational Approach to Edge Detection", IEEE Trans. on PAMI, Vol. 8, No. 6, pp.679-698, 1986.
[11] F. O'Gorman *et al.*, "Finding Picture Edges Through Collinearity of Feature Points", IEEE Trans. on Computers, Vol. C-25, No. 4, pp. 449-456, 1976.

[1] Dissolves and fades were counted out from the false positives.