# A Robust Wipe Detection Algorithm

**C. W. Ngo, T. C. Pong & R. T. Chin**
*Department of Computer Science*
*The Hong Kong University of Science & Technology*
*Clear Water Bay, Kowloon, Hong Kong*
*Email: {cwngo, tcpong, roland}@cs.ust.hk*

**ABSTRACT**

In this paper, we present an improved wipe detection algorithm by analyzing spatio-temporal slices of videos. The algorithm is based on the observation that the regional boundaries of various shapes in spatio-temporal slices give rise to different wipe patterns. Wipes can be detected by locating these boundaries. Our algorithm thus exploits the structural and color information in slices to detect wipes. Compared with other existing methods, our algorithm is efficient, capable of detecting various wipe patterns, and robust to object and camera motions.

**Keywords:** wipe detection, spatio-temporal slice model.

## 1. INTRODUCTION

A video is physically formed by shots. By decomposing videos into shots, we can facilitate the indexing, retrieval and browsing of video content. The boundary between two shots is called a camera break. Based on the transitional properties of camera breaks, we can categorize them as *cut*, *wipe* and *dissolve*. In the current literature, there are various algorithms for detecting cuts [4, 5, 7, 9], however, comparatively few algorithms for detecting wipes [1, 4, 8, 6] and dissolve [2, 4, 7, 9]. Wipes and dissolves involve gradual transitions with no drastic changes between two consecutive frames, and hence, are relatively difficult to identify. While cuts can be identified by comparing two adjacent frames, wipes and dissolves require the investigation of frames along a larger temporal scale.

In this paper, we discuss an improved algorithm for wipe detection. A wipe is generally defined as a moving transition of a line across the screen that enables one shot to gradually replace another [3]. Nonetheless, due to the advance of video production technology, various types of wipe transitions can easily be generated. Figures 1-3 show the examples of wipe transitions which cross about 30 frames. Detecting such transitions in a long video is a non-trivial problem. It requires the close inspection of low-level image features such as lines and motion vectors.

Recently, Wu *et. al.* [6] proposed the projected pairwise difference deviation to detect wipes. However, it can only handle very limited and simple wipe patterns. Yu & Wolf [8] employ frame differencing and edge detection to trace the moving boundary lines created by wipes. Unfortunately, this approach is sensitive to noise, as a result, tends to raise false alarms. Alattar [1] proposed a more general wipe detection algorithm by analyzing the statistical change in the mean and the variance of the wiped frames. Nevertheless, this statistical approach assumes there is only slight motion in shots so that the beginning and ending of wipe regions can be identified. As a result, it cannot detect wipes in videos with fast motions.

Recently, we have proposed work on the wipe detection based on the temporal slice analysis [4]. In this paper, we generalize our previous work so as to handle various wipe transitions. Compared with other approaches, our proposed approach offers the following advantages:

- *Efficiency.* Our approach reduces the wipe detection to spatio-temporal slice segmentation problems. Only partial data is required for analysis without seriously sacrificing detection accuracy.
- *Boundary problem.* It is possible to accurately detect the start and end time of various wipes.
- *Fault tolerant.* It is comparatively easy to distinguish the wipe transitions from motions.

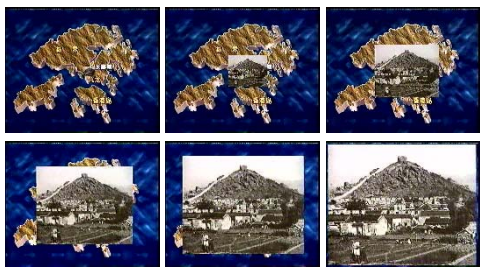


**Figure 1.** Zoom wipe transition (from top to bottom and left to right).

**Figure 2**. Motion wipe transition (from top to bottom and left to right).

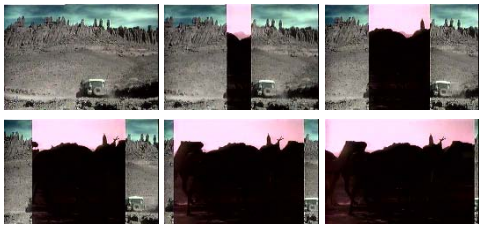


**Figure 3**. Barn door wipe transition (from top to bottom and left to right)..

| Pattern | H | V | D |
|---|---|---|---|
| Iris Round | | | |
| Iris Shape | | | |
| Radial | | | |
| Clock | | | |
| Flip over | | | |
| Spiral box | | | |
| Zig-zag | | | |
| Checker | | | |
| Motion | | | |
| Zoom | | | |
| Barn door | | | |
| Band | | | |
| Wipe | | | |
| Page turn | | | |

**Figure 4**. Spatio-temporal slices of various wipe transitions.

## 2. EMERGENCE OF WIPE PATTERNS IN SPATIO-TEMPORAL SLICES

We can view a video as a volume with $(x, y)$ image dimension and $t$ temporal dimension. The volume is formed by a set of spatio-temporal $2D$ slices each with dimension $(x, t)$ or $(y, t)$, for example. Each spatio-temporal slice is then a collection of $1D$ scans in the same selected position of every frame as a function of time. These slices contain both spatial and temporal information from which coherent regions are indicative of shots separated by camera breaks. Intuitively, segmenting these spatio-temporal slices into regions is equivalent to detecting camera breaks [4]. As in [4], we extract the horizontal **H**, vertical **V** and diagonal **D** spatio-temporal slices which cut through the center of volume for segmentation. Figure 4 shows the spatio-temporal slices of fourteen wipe transitions, while Figure 5 displays an example of each wipe. These wipe transitions create regional boundaries of various shape. Our task is to detect these boundaries and tell when the wipes start and end.

**Figure 5**. Various wipe patterns. From top to bottom and left to right: iris round, iris shape, radial, clock, flip over, spiral box, zig-zag, checker, motion, zoom, barn door, band, wipe, page turn.

It is worth noting that not all spatio-temporal slices will reflect the actual transition period of a wipe. For instance, the **V** of a barn door wipe will only create a vertical boundary line although the transition lasts for more than one frame. Investigating only one slice, in general, is not adequate to detect a wipe precisely. In principle, selecting more slices will improve the precision with the expense of computational cost. For our application, three slices $\{\mathbf{H},\mathbf{V},\mathbf{D}\}$ are chosen to cover most of the wipe transition period. Among the fourteen tested wipe patterns, four frames in the clock wipe, eight frames in the motion wipe, and three frames in the page turn are not covered by $\{\mathbf{H},\mathbf{V},\mathbf{D}\}$.

## 3. SPATIO-TEMPORAL SLICE MODEL

The spatio-temporal slice model captures the shape of regional boundary as *priori* knowledge and measures the changes of coherency in slices. These changes correspond to the appearance of camera breaks.

Denote $x_i = (h_i, v_i, d_i)$ as a pixel triple located at $\mathbf{H}(i,t)$, $\mathbf{V}(i,t)$ and $\mathbf{D}(i,t)$. The probability that a pixel triple $x_i$ is at the boundary $\xi$ of two connected regions is

$$\log p(x_i \in \xi | \mathbf{H}, \mathbf{V}, \mathbf{D}) \propto -\{U(h_i) + U(v_i) + U(d_i)\}$$

where $U(.)$ is the energy function of a pixel governed by the color-texture properties of its $3 \times 3$ neighboring pixels[1]. Considering all pixels in a scan $x$, we write $\sum_{i=1}^{N} \log\{p(x_i \in \xi | \mathbf{H}, \mathbf{V}, \mathbf{D})\} \propto -\sum_{i=1}^{N}\{U(h_i) + U(v_i) + U(d_i)\}$. Subsequently, we have

$$L(x \in \xi) \propto -\sum_{i=1}^{N}\{U(h_i) + U(v_i) + U(d_i)\} \qquad (1)$$

where $L(x \in \xi) = \log\{\prod_{i=1}^{N} p(x_i \in \xi | \mathbf{H}, \mathbf{V}, \mathbf{D})\}$. In other words, the likelihood of a regional boundary is dependent on the total energy value of the three corresponding scans in the selected slices.

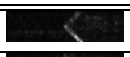
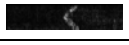
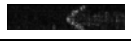
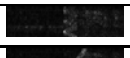
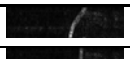
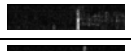
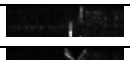
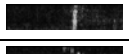
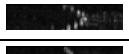
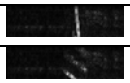
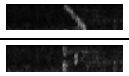
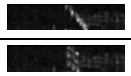
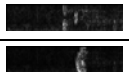
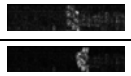
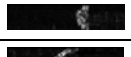
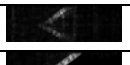
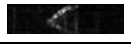
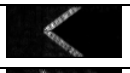
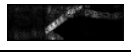
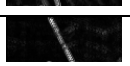
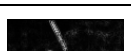
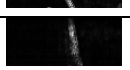
| Pattern | **H** | **V** | **D** |
|---|---|---|---|
| Iris Round |  |  |  |
| Iris Shape |  |  |  |
| Radial |  |  |  |
| Clock |  |  |  |
| Flip over |  |  |  |
| Spiral box |  |  |  |
| Zig-zag |  |  |  |
| Checker |  |  |  |
| Motion |  |  |  |
| Zoom |  |  |  |
| Barn door |  |  |  |
| Band |  |  |  |
| Wipe |  |  |  |
| Page turn |  |  |  |

**Figure 6**. Computed energy for the spatio-temporal images of various wipe patterns

We further classify $U(.)$ into three types of energy: $U_{cut}(.)$, $U_{wipe^-}(.)$ and $U_{wipe^+}(.)$. $U_{cut}(.)$ models the energy of a vertical boundary line; $U_{wipe^+}(.)$ models the energy of a slanted boundary line of positive gradient; while $U_{wipe^-}(.)$ models the energy of a slanted boundary line of negative gradient. To detect wipe, we let $\xi = wipe$ and re-write (1) as

$$L(x \in wipe) \propto -M \qquad (2)$$

where $M$ is the minimum value of the set

$$\left\{ \begin{array}{l} \sum_{i=1}^{N}\{U_{wipe}(h_i) + U_{wipe}(v_i) + U_{wipe}(d_i)\} \\ \sum_{i=1}^{N}\{U_{cut}(h_i) + U_{wipe}(v_i) + U_{wipe}(d_i)\} \\ \sum_{i=1}^{N}\{U_{wipe}(h_i) + U_{cut}(v_i) + U_{wipe}(d_i)\} \\ \sum_{i=1}^{N}\{U_{wipe}(h_i) + U_{wipe}(v_i) + U_{cut}(d_i)\} \end{array} \right\}$$

and

$$U_{wipe} = \sqrt{U_{wipe^-}^2 + U_{wipe^+}^2}$$

Figure 6 shows the $U_{wipe}$ of slices in Figure 4, the white lines indicate the presence of regional boundaries.

## 4. WIPE DETECTION ALGORITHM

Figure 7 depicts our wipe detection algorithm for a suspected wipe pattern illustrated in Figure 8. It starts by computing the energy of three spatio-temporal slices, and then locates the suspected wipe regions by considering the total value of $L(x \in wipe)$ in a group of five adjacent slices. The color histograms of the two neigbouring blocks (block P and Q as shown in Figure 8) of the suspected wipe regions in **H**, **V** and **D** are compared[2]. If the histogram difference is larger than an empirical threshold, Hough transform will be performed to locate the boundary lines formed by wipes. Only pixels whose values exceed $0.05\%$ of the total values in the Hough space are considered as peaks. If $10\%$ of the total pixels are peaks, the suspected wipe regions are regarded as object or camera motions.

The duration of a detected wipe range is obtained directly from a peak with the highest value in the Hough space. Let the detected wipe range in **H** as $\tau^h = [t_1^h, \ldots, t_p^h]$, **V** as $\tau^v = [t_1^v, \ldots, t_q^v]$, and **D** as $\tau^d = [t_1^d, \ldots, t_r^d]$. If $\tau^h \cap \tau^v \cap \tau^d \neq \emptyset$, the start of a wipe transition is detected as $\min(t_1^h, t_1^v, t_1^d)$, while the end of a wipe transition is detected as $\max(t_p^h, t_q^v, t_r^d)$. A wipe is also detected similarly if $\tau^h \cap \tau^v \neq \emptyset$ or $\tau^h \cap \tau^d \neq \emptyset$ or $\tau^v \cap \tau^d \neq \emptyset$. In addition, two wipes are connected if they are less than $15$ frames apart.

## 5. EXPERIMENT RESULTS

### 5.1. Ground-truth Results

We show the empirical results of the three major steps in our proposed algorithm. These steps are

- Step A: locate the suspected wipe ranges in a spatio-temporal image;

[1] The model is based on our previous approach in [4], interested readers can refer to [4] for more details.

[2] The sizes of wipe regions in **H**, **V** and **D** are not necessary to be equal, the sizes depend on the value of computed $U_{wipe}(.)$.

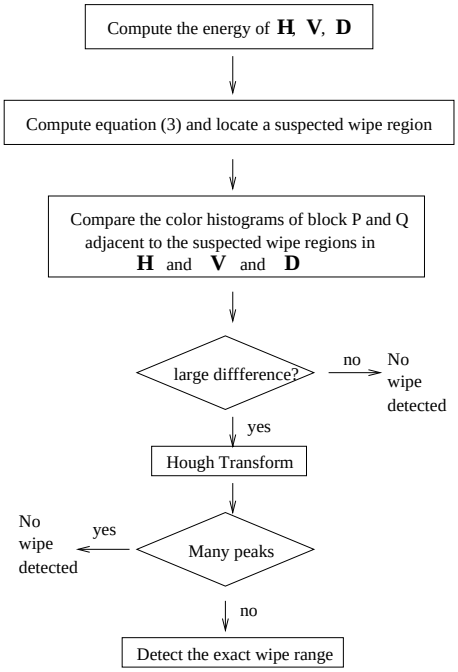



**Figure 7.** Wipe detection algorithm.

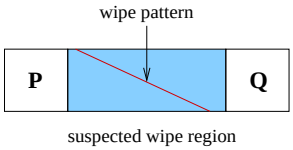



**Figure 8.** P and Q are the blocks (formed by five scans) adjacent to the suspected wipe regions.

- Step B: compare the color histograms of block P and Q of the suspected wipe regions;
- Step C: perform Hough transform on the suspected wipe ranges.

Tables 1 to 3 list the results of using three slices independently. Notice that the desired wipe ranges given in the tables is not the actual wipe transition periods, but the start and end point of the created regional boundaries in a slice. Throughout the experiments, step A can roughly locate all the correct wipe ranges, however, some false wipes are detected due to object or camera motions. The false wipes are eliminated after comparing the color histograms of their neighboring blocks. Nonetheless, a correct wipe range is also eliminated (page turn in **H**) due to the low color contrast. The precise wipe ranges are further located after performing Hough transform. Some wipe ranges are more precisely located (e.g., flip over in **V**), while some correct wipe ranges are eliminated (e.g., band wipe in **D**) due to the presence of many short lines. In most cases, the desired wipe ranges are successfully detected except that the begin and end of wipes cannot be detected precisely (e.g., radial wipe in **H**).

| Wipe Pattern | Desired wipe range | Detected Wipe Range | | |
|---|---|---|---|---|
| | | A | A+B | A+ B+C |
| Iris round | 80-101 | 80-114 | 80-114 | 80-102 |
| Iris shape | 80-109 | 77-97 | 77-97 | - |
| Radial | 88-109 | 84-116 | 84-116 | 85-106 |
| Clock | 80-101 | 84-91<br>99-116 | 84-91<br>99-116 | 86-103<br>86-103 |
| Flip over | 80-109 | 78-114 | 78-114 | 80-109 |
| Spiral box | 88-109 | 92-117 | 92-117 | 107-109 |
| Zig-zag | 95-100 | 91-102 | 91-102 | 95-99 |
| Checker | 80-109 | 79-116 | 79-116 | 82-93 |
| Motion | 88-109 | 85-112 | 85-116 | 86-109 |
| Zoom | 43-83 | 6-11<br>43-85 | 43-85 | 47-73 |
| Barn door | 34-63 | 31-59 | 31-59 | 32-54 |
| Band | 31-45 | 6-13<br>15-97 | 15-97 | 29-44 |
| Wipe | 41-63 | 38-68<br>75-82<br>84-100 | 38-68 | 41-62 |
| Page turn | 58-70 | 54-68 | - | - |

**Table 1.** Detection results based on the horizontal slices.

### 5.2. Comparative Studies

In this section, we employ the approach in Section 4 to consider the detected wipe ranges of **H**, **V** and **D** in Section 5.1 jointly. In addition, we compare the

| Wipe Pattern | Desired wipe range | Detected Wipe Range | | |
|---|---|---|---|---|
| | | A | A+B | A+ B+C |
| Iris round | 80-99 | 78-100 | 78-100 | 79-97 |
| Iris shape | 80-98 | 78-96 | 78-96 | 88-90 |
| Radial | 80-103 | 79-104 | 79-104 | 83-101 |
| Clock | 80-96 | 77-98 | 77-98 | 93-96 |
| Flip over | 94-96 | 88-113 | 88-113 | 95-97 |
| Spiral box | 82-109 | 78-112 | 78-112 | 85-110 |
| Zig-zag | 82-109 | 78-112 | 78-112 | 85-109 |
| Checker | 80-96 | 76-89<br>92-99 | 76-89 | - |
| Motion | 98-109 | 93-113 | 93-113 | 96-104 |
| Zoom | 43-80 | 6-11<br>43-85<br>108-108 | 43-85 | 47-72 |
| Barn door | 34-34 | 29-38 | 29-38 | 32-36 |
| Band | 37-37 | 6-110 | 6-110 | 37-37 |
| Wipe | 52-52 | 48-58<br>60-61<br>86-95 | 48-56 | 51-51 |
| Page turn | 54-63 | 52-68 | 52-68 | 61-62 |

**Table 2**. Detection results based on the vertical slices.

| Wipe Pattern | Desired wipe range | Detected Wipe Range | | |
|---|---|---|---|---|
| | | A | A+B | A+ B+C |
| Iris round | 80-109 | 80-114 | 80-114 | 81-103 |
| Iris shape | 85-107 | 83-113 | 83-113 | 90-111 |
| Radial | 94-97 | 93-114 | 93-114 | 96-97 |
| Clock | 92-105 | 88-114 | 88-114 | 90-106 |
| Flip over | 80-109 | 78-114 | 78-114 | 86-109 |
| Spiral box | 80-109 | 78-137 | 78-137 | 84-110 |
| Zig-zag | 80-109 | 83-114 | 83-114 | 94-108 |
| Checker | 85-96 | 79-114 | 79-114 | 84-95 |
| Motion | 96-109 | 93-137 | 93-137 | 96-102 |
| Zoom | 52-80 | 6-11<br>47-86<br>108-109 | 47-86 | 52-73 |
| Barn door | 34-60 | 30-58 | 30-58 | 32-57 |
| Band | 31-45 | 6-110 | 6-110 | - |
| Wipe | 41-58 | 39-67<br>75-99 | 39-67 | 41-56 |
| Page turn | 54-69 | 51-95 | 51-95 | 55-65 |

**Table 3**. Detection results based on the diagonal slices.

performance of our approach (namely slice coherency) with the statistical approach proposed by Alattar [1]. The statistical approach first detects the spikes in the second derivative of the mean and variance of frames. These spikes mark the end or the start and end of a potential wipe. The approach then investigates the average values of the first derivative of the mean and variance of a suspected wipe region. A wipe is detected if the average value is above a threshold.

Table 4 lists the wipe transitions along the actual and detected wipe frames by these two algorithms. To analytically measure the results, we compute recall-precision measures

$$recall = \frac{D}{A} \quad precision = \frac{D}{B}$$

where $A$ is the number of wiped frames in the fourteen tested sequence; $B$ is the number of detected wiped frames; $D$ is the number of correctly detected wiped frames. The value of $recall$ and $precision$ are in the interval of $[0, 1]$. Low recall values indicate the frequent occurrence of missed detections, while low precision values show the frequent occurrence of false alarms. For instance, if only $20$ frames are detected for a wipe of $30$ frames, then $precision = 1$ and $recall = \frac{2}{3}$, while if there are $40$ frames detected, then $precision = \frac{3}{4}$ and $recall = 1$.

As shown in Table 5, our proposed algorithm shows a significantly better performance than the statistical approach in term of recall and precision. The statistical approach fails in detecting some wipes because of the absence of sharp spikes in the wipe regions. Furthermore, it cannot accurately mark the boundary of a wipe when there are motions in the two connected shots.

### 5.3. Fault Tolerant Issue

We test the robustness of our algorithm on three sequences having camera panning, zooming and object motions but without wipe transition. These motions create random and short lines in slices. Figures 9-11 show the spatio-temporal slices and their $U_{wipe}$. In step A, some false wipe regions are detected due to the presence of random lines. Nonetheless, all the false wipe regions are removed after performing the steps B and C. In contrast, the statistical approach [1] generates false alarms for all the three tested sequences.

## 6. CONCLUSION

We have presented an improved wipe detection algorithm by detecting the regional boundaries of three temporal slices. The algorithm employs a spatio-temporal slice model that detects the change of coherency as camera break. To improve performance, the algorithm also employs color histogram comparison to remove false wipes, and Hough transform to precisely locate the boundaries. Experimental results demonstrate that the algorithm can detect various wipe

| Wipe Pattern | Actual Wipe Range | Detected Wipe Range | |
|---|---|---|---|
| | | Slice Coherency | Statistical Approach |
| Iris round | 80-109 | 79-103 | none |
| Iris shape | 80-109 | 88-111 | none |
| Radial | 80-109 | 83-106 | none |
| Clock | 80-109 | 86-106 | none |
| Flip over | 80-109 | 80-109 | 80-109 |
| Spiral box | 80-109 | 84-110 | none |
| Zig-zag | 80-109 | 81-107 | none |
| Checker | 80-109 | 82-95 | none |
| Motion | 80-109 | 86-106 | 101-109 |
| Zoom | 43-83 | 47-73 | none |
| Barn door | 34-63 | 32-57 | 34-72<br>82-89 |
| Band | 31-45 | 29-44 | 31-93 |
| Wipe | 41-63 | 41-62 | 40-69<br>77-81 |
| Page turn | 54-73 | 55-65 | 61-66 |

**Table 4.** Detection results on various wipe transitions.

| Approach | $Recall$ | $Precision$ |
|---|---|---|
| Slice Coherency | 0.77 | 0.97 |
| Statistical Approach | 0.28 | 0.60 |

**Table 5.** Recall and precision measures for various wipe transitions

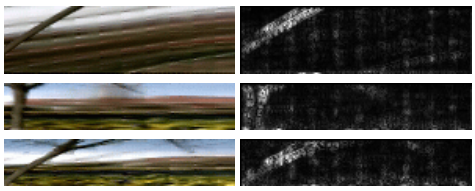


**Figure 9.** The spatio-temporal slice and $U_{wipe}$ of camera panning sequence. From top to bottom: **H**, **V**, **D**.

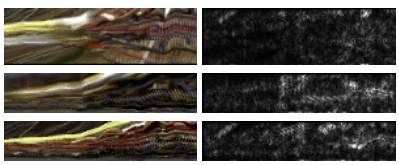


**Figure 10.** The spatio-temporal slice and $U_{wipe}$ of camera zooming sequence. From top to bottom: **H**, **V**, **D**.

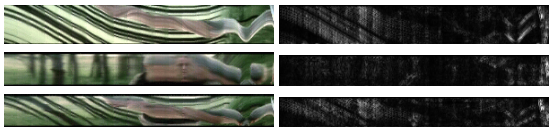


**Figure 11.** The spatio-temporal slice and $U_{wipe}$ of object and camera motion sequence. From top to bottom: **H**, **V**, **D**.

transitions, in addition, is tolerant to camera and object motions. In the current implementation, the algorithm operates in 38 frames/second on a Pentium II machine.

## Acknowledgements

This work is supported in part by RGC Grants HKUST661/95E and HKUST6072/97E.

### References

[1] A. M. Alattar, "Wipe Scene Change Detector for Use with Video Compression Algorithms and MPEG-7," *IEEE Transactions on Consumer Electronics*, vol. 44, no. 1, pp. 43-51, Feb 1998.

[2] A. M. Alattar, "Detecting and Compressing Dissolve Regions in Video Sequences with a DVI Multimedia Image Compression Algorithm," *IEEE Intl. Symposium on Circuits and Systems*, vol. 1, pp.13-16, 1993, New York.

[3] D. Bordwell & K. Thompson, *Film Art: an Introduction*, University of Wisconsin, 2nd edition, Random House, 1986.

[4] C. W. Ngo, T. C. Pong & R. T. Chin, "Detection of Gradual Transitions through Temporal Slice Analysis," *Proc. IEEE Int. Conf. on Computer Vision and Pattern Recognition*, vol.1, pp. 36-41, 1999.

[5] N. V. Patel & I. K. Sethi, "Compressed Video Processing for Cut Detection," *IEE Proc. Visual Image Signal Process*, vol. 143, no. 5, pp. 315-23, Oct 1996.

[6] M. Wu & W. Wolf & B. Liu, "An Algorithm for Wipe Detection," *IEEE Intl. Conf. on Image Processing*, vol. 1, pp. 893-7, 1998, Los Alamitos.

[7] B. L. Yeo & B. Liu, "Rapid Scene Analysis on Compressed Video," *IEEE Trans. on Circuits & Systems for Video Technology*, vol. 5, no. 6, pp. 533-44, 1995.

[8] H. Yu & W. Wolf, "A Multi-resolution Video Segmentation Scheme for Wipe Transition Identification," *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, vol. 5, pp. 2695-8, 1998, New York.

[9] H. J. Zhang, A. Kankanhalli, & S. W. Smoliar, "Automatic Partitioning of full-motion video," *ACM Multimedia System*, Vol. 1, No. 1, pp. 10-28, 1993.