VIDEO SHOT BOUNDARY DETECTION USING INDEPENDENT COMPONENT ANALYSIS

Jian Zhou, Xiao-Ping Zhang

Department of Electrical and Computer Engineering, Ryerson University
350 Victoria Street, Toronto, Ontario, Canada, M5B 2K3
E-mail: {jzhou, xzhang}@ee.ryerson.ca

ABSTRACT

Video shot boundary detection is an important early stage of content-based video analysis. In this paper, a new method for shot boundary detection using independent component analysis (ICA) is presented. By projecting video frames from illumination-invariant raw feature space into low dimensional ICA subspace, each video frame is represented by a two-dimensional compact feature vector. An iterative clustering algorithm based on adaptive thresholding is developed to detect cuts and gradual transitions simultaneously in ICA subspace. Experimental results successfully validate the new method and show that it can effectively detect both abrupt transitions and gradual transitions.

1. INTRODUCTION

Content analysis of video has been an area of active research in recent years. As the building block of video projects, a video shot is a fundamental abstraction level for video analysis. Detecting shot boundaries means temporally segmenting a video into its constituent shots and thus recovering the elementary units of a video. The research of video parsing focuses on the detection of two types of transitions: abrupt transition (cut) and gradual transition (fade/dissolve).

Many automatic techniques have been developed to detect video boundaries. In early works [1], pair-wise pixel comparison, likelihood ration and histogram comparison were used for cut detection. In [2], scene changes were detected by using pixel difference and luminance histograms based on DC-images in compressed domains. In [3], edge changes were used as a feature for shot detection. Shot detection techniques were reviewed in detail in [4], and a statistical detector based on motion feature was proposed. Most of the above techniques can achieve good performance on hard cut detection.

Gradual transitions, especially dissolves are generally more difficult to detect. One of the main reasons is that it is difficult to define and capture the visual discontinuities. Therefore, new features, such as edge changes and intensity variance, have been introduced to detect dissolves. In [5], a twin threshold mechanism based on histogram difference was used to detect gradual transitions. The most commonly used feature for dissolve detection is intensity variance. The intensity variance curve forms a downwards-parabolic shape during a dissolve and it has been used in many dissolve detectors [4] [6] [7]. In a previous work, we introduced a new feature based on skewness, and dissolves were detected by a combined analysis of mean-variance-skewness [8]. Most of these existing techniques require careful selection of thresholds to achieve good performance. Such parameter tuning is undesired, especially for the video data from different genres.

In this paper, we present a data-driven feature extraction for shot detection based on independent component analysis (ICA) model. The same feature is used for both cut detection and gradual transition detection. Since the features learned from ICA can automatically adapt to data, the configurable parameters are expected to be more robust to different data, compared with those for manually selected features. In the new method, illumination-invariant chromaticity histogram from each video frame is created to form raw features. By performing ICA, two independent components (ICs) are generated and chosen as features. In the low dimensional ICA subspace, a dynamic clustering algorithm based on adaptive thresholding is developed to detect shot boundaries. Experimental results successfully show that the new method can effectively detect both abrupt transitions and gradual transitions.

2. ICA REPRESENTATION

ICA is a recently developed technique using high-order statistics [9]. ICA is a linear non-orthogonal transform which blindly separates the independent source signals from their linear mixtures without knowing the mixture matrix. ICA is an extension of classical principal

component analysis (PCA). PCA is optimal in terms of reconstruction error in Euclidean space. The features produced by PCA are mutually uncorrelated. However, ICA not only decorrelates the data but also reduces higher-order statistical dependence [9] of data. Denote the N-dimensional random vector by $\vec{x} = [x_1, x_2, \cdots, x_N]^T$, and the M statistically independent non-Gaussian source signals by $\vec{s} = [s_1, s_2, \cdots, s_M]^T$ with $M \le N$. The noiseless ICA model can be expressed as follows,

$$\vec{x} = A \cdot \vec{s}, \quad (1)$$

where $A = [\vec{a}_1\ \vec{a}_2 \cdots \vec{a}_M]$ is an $N \times M$ invertible mixing matrix with linearly independent columns. The ICA task is to find an unmixing matrix $W \approx A^{-1}$ using only the observed signal $\vec{x}$. The matrix W is applied on the observed signal such that the output signals denoted by $\vec{y}$ are as statistically independent as possible,

$$\vec{y} = W \cdot \vec{x} = W \cdot A \cdot \vec{s}. \quad (2)$$

The rows of the output signals are ICs. The basis functions learned by ICA form the columns of matrix A. Rows of W can be considered as filters. In this paper, we employ the FastICA [10] algorithm to estimate the unmixing matrix and the individual ICs.

3. NEW SHOT DETECTION METHOD BASED ON ICA

The new method has the following major steps: (i) Raw feature generation from illumination-invariant chromaticity histograms; (ii) ICA feature extraction; (iii) Dynamic clustering for shot detection. Each step is described in the following subsections.

3.1. Illumination-invariant Chromaticity Histogram

Illumination changes and object/camera motion are the key factors that affect the performance of shot detection. Since histograms do not carry spatial information, they are expected to be robust to object and camera motion. However, histograms are generally sensitive to lighting changes. Therefore, in the new method, the normalized chromaticity histograms [11] are chosen as raw features. Based on 3D RGB color space, the 2D illumination-invariant normalized chromaticity (r, g) is defined as,

$$r = R/(R+G+B),\ g = G/(R+G+B). \quad (3)$$

Histograms with 256 bins are generated as features in the normalized chromaticity color space for each frame of video. In this work, only r component is used. Thus, the dimension of raw feature vector is N=256.

3.2. ICA Feature Extraction

ICA has been used for applications such as blind source separation, compression and denoising. In [12], ICA model is used to extract basis functions from natural images. Such basis functions could be used as features since two different classes of images tend to have different basis functions. In the new method, the ICA model is applied in feature domain. Each video frame (raw feature vector) is processed as one observation that can be considered as a linear combination of hidden basis functions. Since the time course is only associated with the ICs, we select the most two significant ICs as the new features instead of the basis functions. The temporal characteristics of ICs are explored by a clustering algorithm to detect shot boundaries.

For the i-th video frame, let $\vec{h}_i$ denote the raw feature vector created from the normalized chromaticity histogram. Using $\vec{h}_i$ as a column vector, the observed N-dimensional (N=256) signal is constructed in matrix form as,

$$X = [\vec{h}_1\ \vec{h}_2 \cdots \vec{h}_p], \quad (4)$$

where p is the total length of the video sequence. Each frame is represented by a column of X. ICA learning method is performed to generate the unmixing matrix W and the independent sources. We reduce the dimension and only keep the two most significant projecting directions (M=2). The two-dimensional output ICs are given by the product of matrices W and X. Thus, the data are projected onto an ICA subspace spanned by two basis functions. Each IC gives the coordinates for one projection direction. A video frame is represented by a point in the ICA subspace. The frames within one shot tend to form a compact cluster.

3.3. Dynamic Clustering for Shot Detection

Based on video frame distribution in the ICA subspace, a dynamic clustering algorithm is developed to classify video frames into shots and detect the shot boundaries. Euclidean distance is used as dissimilarity measure between two points, $\vec{x}_i$ and $\vec{x}_j$, in ICA subspace.

$$d(\vec{x}_i, \vec{x}_j) = \left\| \vec{x}_i - \vec{x}_j \right\|_2. \quad (5)$$

Given the (i+1)-th sample $\vec{x}_{i+1}$, the sample mean vector $\vec{\mu}$ can be iteratively updated as

$$\vec{\mu}_{i+1} = \vec{\mu}_i + \frac{(\vec{x}_{i+1} - \vec{\mu}_i)}{i+1}. \quad (6)$$

Denote a vector by $\vec{\phi} = [\sigma^2_{(1)}, \cdots, \sigma^2_{(M)}]^T$ where the n-th element $\sigma^2_{(n)}$ ($1 \le n \le M$) is the sample variance in the n-th dimension of the feature vector in ICA subspace.

The n-th element of vector $\vec{\phi}$ at time (i+1) is iteratively updated as

$$\sigma^2_{i+1,(n)} = \left(\frac{i-1}{i}\right)\cdot\sigma^2_{i,(n)} + (i+1)\cdot(\mu_{i+1,(n)} - \mu_{i,(n)})^2 , \quad (7)$$

Due to camera motion and noise, intra-shot variations may cause the cluster center to gradually float away. In order to reduce the contributions from old samples, we introduce a decay factor α with $0 < \alpha < 1$. Denote the weighting vector by $\vec{w} = [\alpha^{i-1} \ldots \alpha^1 \ 1]^T$. For a given vector $\vec{x}$ with size of i, its weighted sample mean can be calculated as:

$$\mu = \frac{\vec{w}^T \cdot \vec{x}}{\|w\|_1} . \quad (8)$$

If the sample size i is large, the weighted sample mean can be iteratively estimated as:

$$\vec{\mu}_{i+1} = \alpha \cdot \vec{\mu}_i + (1-\alpha)\cdot \vec{x}_{i+1} . \quad (9)$$

In practice, (9) is used instead of (6) to calculate the cluster center. We still use (7) to approximate the sample variance, since we are not interested in estimating a true and unbiased weighted variance.

During clustering process, for a new sample $\vec{x}_{i+1}$, we calculate the distance between this new sample and the cluster center. The adaptive threshold T_a is defined as

$$T_a = \beta \cdot \left\|\vec{\phi}\right\|_1 , \quad (10)$$

where β is a predefined parameter to control how big the intra-shot variations are allowed. The new sample is classified into the current cluster if the following condition holds

$$d(\vec{\mu}_i, \vec{x}_{i+1}) < T_a . \quad (11)$$

Then (9) and (7) are used to update the sample mean and sample variance. Otherwise, if the distance is larger than T_a, we create a new cluster initialized with the sample $\vec{x}_{i+1}$. The time index of $\vec{x}_{i+1}$ is saved as a shot boundary. Since the condition adapts to the density of the points in ICA subspace, this mechanism essentially introduces an adaptive thresholding.

Two techniques are developed to improve performance. The first technique is outlier removal. If the distance between one sample and the current cluster center is larger than T_a, we check whether or not the next sample satisfies the condition (11). If the next sample can be classified into the current cluster, the previous sample is considered as an outlier and discarded. The other technique is to improve the performance for detecting gradual transitions. Once the recent samples are found to "move away" from the current cluster, a new cluster is formed. But this new cluster might be within the transition period when dealing with gradual transitions. A special property for those points within a transition period is that they are sparsely distributed in ICA subspace as shown in Figure 1. To capture this property, we use a temporal window of size *K (K=30)*. Let J denote the average variation of sample variance within the temporal window. We define a measure of cluster compactness as:

$$J = \left(\sum_{k=1}^{K-1} \left\|\vec{\phi}_{k+1} - \vec{\phi}_k\right\|_1\right) \Big/ (K-1) . \quad (12)$$

The above criterion is used to distinguish gradual transitions from cuts. If J is larger than a predefined threshold, a gradual transition is declared. Otherwise, the boundary is detected as a cut. It is worth mentioning that this evaluation is checked once at the beginning only when a new cluster is formed.

The clustering algorithm is summarized as follows:

- **Initialization:**
  Get the first *P (P=5)* samples and calculate the sample mean and sample variance directly. In extreme cases such as "freeze" frames, a minimum value T_b is used to initialize the variance if the calculated sample variance is less than T_b.
- **Iterative clustering:**

1. Get a new sample $\vec{x}_{i+1}$ and check condition (11).
2. Update mean and variance by (9) and (7) if condition (11) is satisfied. Otherwise, check outlier removal rule.
3. Repeat step 1 and 2 until a sample can not be classified into the current cluster.
4. Create a new cluster, and use (12) to check the boundary type, and set the new cluster as the current cluster.

4. EXPERIMENTAL RESULTS

In the experiments, we have collected TV shows and documentary video sequences as the test data. The test video data is carefully selected to include as many effects as possible. The documentary video sequences contain many editing effects such as zoom-ins, zoom-outs, and camera panning. The experimental results are shown in Table 1. Precision and recall for cuts were obtained as 95% and 97.4% respectively. Precision and recall for gradual transitions were 85.7% and 89.3%. The false positives for TV shows were caused by fast camera motion. The water scenes in documentary video created some false positive for gradual transition detection. One gradual transition in documentary video joins two similar scenes. That was missed in our method. Even though we have chosen the difficult test data, the proposed method still had good performance. The results show that the

algorithms are effective for both cut detection and gradual transition detection.

Test Data	Total (C)	Total (G)	Missed (C)	Missed (G)	False (C)	False (G)
TV show	53	3	0	1	2	1
Documentary	64	44	3	4	4	6
Total	117	47	3	5	6	7

Table 1: Detection results for cuts (C) and gradual transitions (G).

5. CONCLUSIONS

In this paper, we present a new method for shot boundary detection. Raw features are formed by normalized chromaticity histograms that are illumination-invariant. By performing ICA, two ICs are generated. Unlike typical image feature extraction using ICA, which uses basis functions as features, we choose ICs as features and explore their temporal characteristics. By projecting the high dimensional raw features into low dimensional ICA subspace, video shots are represented as separable compact clusters. A dynamic clustering algorithm using adaptive thresholding is developed to detect both cuts and gradual transitions at one pass. The simulations show that the method achieved good performance for detecting both abrupt transitions and gradual transitions. Our future work is to extend the method to extract key frames based on cluster centers and apply ICA to video retrieval.

6. REFERENCES

[1] H. Zhang, A. Kankanhalli, and S. W. Smoliar, "Automatic partitioning of full-motion Video," *Multimedia Systems*, vol. 1, pp. 10-28, 1993.

[2] B. Yeo and B. Liu, "Rapid scene analysis on compressed video," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 5, pp. 533-544, Dec. 1995.

[3] R. Zabih, J. Miller, and K. Mai, "A feature-based algorithm for detecting and classifying scene breaks," in *Proc. ACM Multimedia 95*, San Francisco, CA, pp. 189-200, Nov. 1995.

[4] A. Hanjalic, "Shot-boundary detection: unraveled and resolved?," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 12, no. 2, pp. 90-104, Feb. 2002.

[5] H. J. Zhang, C. Y. Low, S. W. Smoliar, and J. H. Wu, "Video parsing, retrieval and browsing: An integrated and content-based solution," *Proc. of ACM Multimedia'95*, Nov. 1995.

[6] 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*, vol. 1, pp. 13-16, May 1993.

[7] B. T. Truong, C. Dorai, and S. Venkatesh, "New enhancements to cut, fade, and dissolve detection processes in video segmentation," in *Proceedings of the 8th ACM International Conference on Multimedia*, pp. 219-227, Nov. 2000.

[8] J. Zhou and X.-P. Zhang, "A web-enabled video indexing system," in *Proceedings of ACM MIR'04*, New York, Oct. 2004.

[9] P. Comon, "Independent component analysis – a new concept?," *Signal Processing*, vol. 36, pp. 287-314, 1994.

[10] A. Hyvärinen, "Fast and robust fixed-point algorithms for independent component analysis," *IEEE Transactions on Neural Networks*, vol. 10, no. 3, pp. 626-634, 1999.

[11] M. S. Drew, J. Wei, and Z.-N. Li, "Illumination-invariant color object recognition via compressed chromaticity histograms of color-channel-normalize images," *ICCV'98*, pp. 533-540, 1998.

[12] A. J. Bell, and T. J. Sejnowski, "The 'independent components' of natural scenes are edge filters," *Vision Research*, 37: 3327-3338, 1997.

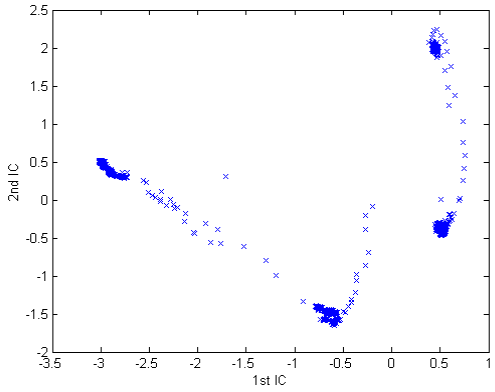

Figure 1: Cluster patterns formed during dissolves in the ICA subspace.

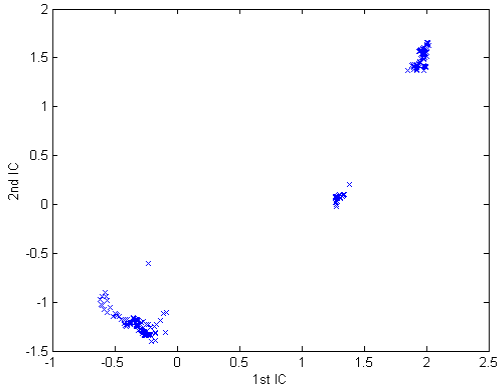

Figure 2: Cluster patterns for cuts in the ICA subspace.