# PCA-based approach for video scene change detection on compressed video

L. Gao, J. Jiang, J. Liang, S. Wang, S. Yang and Y. Qin

An automatic, real-time detection approach to video scene change detection is presented. Owing to the high correlation of two consecutive video frames, it is proposed that only the eigenvector corresponding to the largest eigenvalue is retained in the principal component analysis (PCA) for video data. A one-dimensional PCA feature of video data is then generated from the PCA. It shows superior performance compared to the histogram feature and the pixel feature. The detection algorithm based on this PCA feature is then designed to detect both abrupt and gradual transitions. The proposed approach is tested on the TREC video test repository to validate its performance.

*Introduction:* Video scene change detection is a fundamental and crucial step in automatic content analysis algorithms [1]. There are two types of scene transitions: abrupt and gradual. Various algorithms have been proposed [1]. However, it is generally difficult to detect the scene change, owing to fast object or camera motion, camera flashlight, fast illumination changes and MPEG errors [1]. The existing algorithms perform relatively well for abrupt transition [2]. But, up to now, gradual transition detection remains a challenging problem, mainly because it is hard to be distinguished from fast object or camera motion [1, 2]. The precision and recall for gradual transition are still at a low level [1, 2].

In this Letter, the design of a new detection algorithm is presented. Its input is the new feature extracted by the principal component analysis (PCA). The PCA feature shows superior performance compared to the histogram feature and the pixel feature [1, 2] used in other algorithms. This new detection algorithm can further improve precision and recall, especially for gradual transition detection.

*PCA for video data:* A training set of the DC images [2] that is updated at the beginning of each video scene is used to train the eigenspace. Given a DC image of size $h \times w$, the image vector $X_k$ of length $N = h \times w$ is obtained by concatenating the rows. The data matrix $P$ for the training set of DC images is an $N \times M$ matrix ($N \gg M$), that is $P = [X_1, \ldots, X_M]$, where $M$ is the size of the training set. The PCA finds eigenvectors from the estimated correlation matrix $C$, that is $C = PP^T$. Here, $C$ is an $N \times N$ matrix, and its size is very large, which would result in computationally intensive operations. As described in [3], the implicit matrix, $\tilde{C} = P^T P$, is considered for negotiating this problem. Here, $\tilde{C}$ is an $M \times M$ matrix, which is much smaller than $C$. The $L$ ($L \leq M$) largest eigenvalues $\lambda_i$, and corresponding eigenvectors $e_i$ of $C$ can be found from the $L$ largest eigenvalues $\tilde{\lambda}_i$ and $\tilde{e}_i$ of $\tilde{C}$ [3], that is $\lambda_i = \tilde{\lambda}_i$ and $e_i = \tilde{\lambda}_i^{-(1/2)} P \tilde{e}_i$, where $i = 1, 2, \ldots, L$.

In this Letter, for the training set of video data, we propose that only the eigenvector $e_1$ corresponding to the largest eigenvalue $\lambda_1$ is retained to constitute a new eigenspace. We find that the power of other eigenvalues decreases very fast relative to that of the largest eigenvalue. This is due to very high correlation of consecutive frames in the videos. Consequently, the primary characteristics of the image are centralised on the eigenvector $e_1$ in the new eigenspace. We find that the largest eigenvalue generally accounts for over 95% of the total eigenvalues in the PCA, even if fast object or camera motion occurred. As shown in Table 1, there are eight eigenvalues corresponding to eight consecutive images of video Lanc.m2v (frame 1300–1307). The largest eigenvalue accounts for 99.8% of the total eigenvalues.

**Table 1:** Eight eigenvalues corresponding to eight consecutive video images

| $\lambda$ | Eigenvalue |
|---|---|
| $\lambda_1$ | $4.5231 \times 10^9$ |
| $\lambda_2$ | $5.082 \times 10^6$ |
| $\lambda_3$ | $8.2356 \times 10^5$ |
| $\lambda_4$ | $1.478 \times 10^5$ |
| $\lambda_5$ | 75532 |
| $\lambda_6$ | 46433 |
| $\lambda_7$ | 20170 |
| $\lambda_8$ | 18544 |

*One-dimensional PCA feature:* The $N \times 1$ eigenmatrix $\phi$ is then generated, the column of which is the eigenvector $e_1$. Using this matrix $\phi$, each original DC vector $X_k$ of the video can be represented by a one-dimensional PCA feature $Y_k$ form (1) in the new eigenspace:

$$Y_k = \phi^T X_k \tag{1}$$

This feature $Y_k$ shows superior performances compared to the histogram feature [2] used in other algorithms. First, the feature $Y_k$ changes dramatically when the eigenmatrix $\phi$ obtained in the current video scene is applied to the DC vector $X_k$ of the next video scene from (1), as shown in Fig. 1, using the video NAD55.mpg of the TREC video test repository as example.

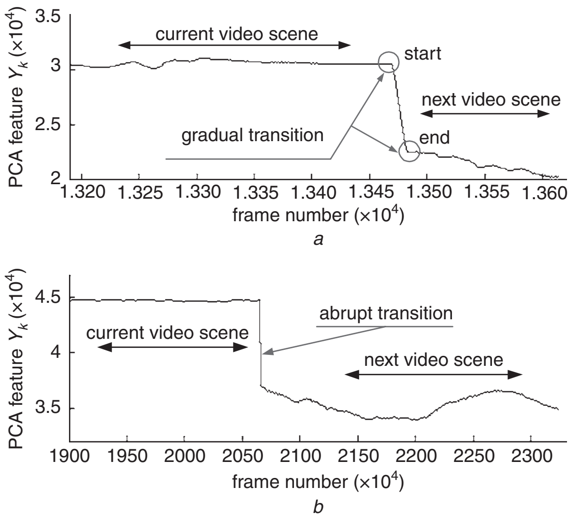


**Fig. 1** *PCA feature $Y_k$*

*a* Feature $Y_k$ in gradual transition of video NAD55.mpg (frame: 13190–13615)
*b* Feature $Y_k$ in abrupt transition of video NAD55.mpg (frame: 1900–2325)

Secondly, in traditional algorithms, the difference $h_k(H_k, H_{k+1})$ between histogram feature $H_k$ and $H_{k+1}$ of two consecutive frames is calculated [1, 4] as the input of the detection algorithm. As shown in Fig. 2, the level of peaks in true gradual transition is hard to distinguish from that in fast object or camera motion. However, in this Letter, the difference $d_k(Y_k, Y_{k+1})$ between the PCA feature $Y_k$ and $Y_{k+1}$ is calculated too. The curve $d_k$ has a dramatic peak at the time of abrupt transition. Additionally, it also remains at a relatively high level through the duration of the gradual transition, as shown in Fig. 3, which is obviously distinguished from fast object or camera motion.

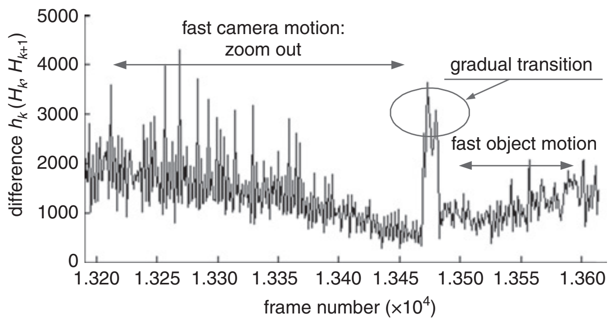


**Fig. 2** *Difference $h_k(H_k, H_{k+1})$ for histogram feature in gradual transition of video NAD55.mpg (frame 13190–13615, same frames as in Fig. 1a)*

Lastly, compared with the histogram feature with 256 bins generally, the one-dimensional PCA feature $Y_k$ can both save memory space and reduce the execution time. For example, we use the Lanc.mpg (6000 frames) as the test sequence, running on an Intel Pentium IV 2.0 GHz and using the Matlab 6.1 processing language. The execution time of the detection algorithm based on the histogram feature is 108.9 s, while that based on the PCA feature is 25.2 s.

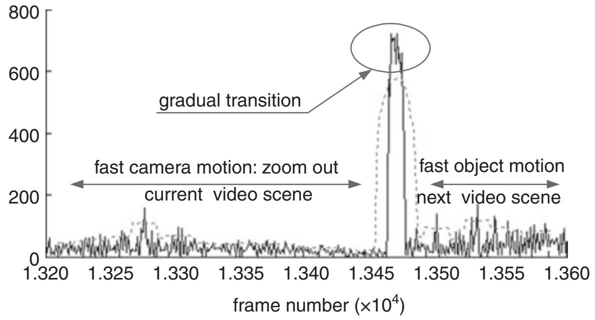


**Fig. 3** *Difference $d_k(Y_k, Y_{k+1})$ for PCA feature in gradual transition of video NAD55.mpg (frame 13190–13615, same frames as in Fig. 1a) and new test feature $b_k$ corresponding to difference $d_k(Y_k, Y_{k+1})$*

—— difference $d_k(Y_k, Y_{k+1})$
- - - - new test feature $b_k$

*Video scene change detection:* First, the difference $d_k(Y_k, Y_{k+1})$ between feature $Y_k$ and $Y_{k+1}$ of two consecutive video frames is calculated, that is $d_k = |Y_{k+1} - Y_k|$. Secondly, the abrupt scene change detection is implemented. A one-dimensional temporal window of size $N_a = 11$ proposed in [4] is selected. There are two criteria used to determine whether abrupt transition occurs or not in this sliding window. If the current frame meets these two criteria, the abrupt transition is detected.

Otherwise, a gradual scene change detection is implemented when abrupt transition does not occur in the current frame. Another one-dimensional temporal window of size $N_g = 21$ proposed in this Letter is selected for gradual transition. A lowpass filter is applied to smooth the difference $d_i$ ($i = k - n + 1, \ldots, k - 1, k, k + 1, \ldots, k + n - 1$) within this symmetric sliding window. The largest value $b_k$ of the lowpass filter is chosen as the new test feature of the current frame from (2), as shown in Fig. 4. It is used to detect the gradual transition:

$$b_k = \max\{B_i | i = k - n + 1, \ldots, k - 1, k, k + 1, \ldots, k + n - 1\} \quad (2)$$

where $N_g = 2n - 1$, and $B_i$ is the lowpass filter smooth result.

There are also two elimination criteria defined in this sliding window for gradual transition: first, the new test feature $b_k$ of the current frame needs to be larger than a predefined threshold $T_g$ (200–300); secondly, the feature $b_k$ of the current frame is the largest one within the three consecutive frames, including the previous frame, current frame and next frame, that is $b_k > b_{k-1}$ and $b_k > b_{k+1}$. If the two criteria can be satisfied, one gradual transition is detected.

*Experiments:* To evaluate the performance of the proposed algorithm, the algorithm was tested on the TREC video test repository [5, 6]. The comparison between the algorithm's results and ground truth relies on the well-known recall and precision figures of merit [1, 5, 6]. However, a good detector should have both high precision and high recall. $F_1$ is a commonly used metric that combines precision and recall [5, 6], that is

$$F_1 = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \quad (3)$$

The method in [5] achieves an $F_1$ performance of 0.94 and 0.69, for abrupt and gradual transition, respectively. In 2005TREC video evaluation report [6], the best performing system achieves an $F_1$ performance of 0.95 and 0.789, respectively. Our unsupervised algorithm achieves an $F_1$ performance of 0.95 and 0.805 for abrupt and gradual transition, respectively. The total execution time of our scene change detection algorithm is 0.017s per frame, running on an Intel Pentium IV 2.0 GHz, and using the C programming language.

*Conclusions:* An automatic approach based on PCA for video scene change detection has been presented. Our experiments prove the superior performance of the new PCA feature. Then our change detection algorithm is based on the PCA feature to detect the scene change. To evaluate the performance of the approach proposed in this Letter, experiments were carried out on the TREC video test repository. The results show that the recall and precision for scene change detection are further improved via our detection algorithm.


*7 August 2006*
*Electronics Letters* online no: 20062175
doi: 10.1049/el:20062175

L. Gao, J. Liang, S. Wang and S. Yang (*Department of Integrated Digital System, Institute of Acoustics, Chinese Academy of Sciences, 21 Beisihuanxi Road, Haidian, Beijing 100080, People's Republic of China*)

E-mail: future_gao@hotmail.com

J. Jiang (*Department of Electronic Imaging and Media Communications, University of Bradford, Bradford, United Kingdom*)

Y. Qin (*Department of Information Science and Technology, Qingdao University of Science and Technology, Qingdao 266061, People's Republic of China*)

## References

1 Koprinska, I., and Carrato, S.: 'Temporal video segmentation, a survey', *Signal Process., Image Commun.*, 2001, **16**, (5), pp. 450–477
2 Bescos, J.: 'Real-time shot change detection over online MPEG-2 video', *IEEE Trans. Circuits Syst. Video Technol.*, 2004, **46**, (3), pp. 475–484
3 Murakami, H., and Kumar, V.: 'Efficient calculation of primary images from a set of images', *IEEE Trans. Pattern Anal. Mach. Intell.*, 1982, **PAMI-4**, (5), pp. 511–515
4 Yeo, B., and Liu, B.: 'Rapid scene analysis on compressed video', *IEEE Trans. Circuits Syst. Video Technol.*, 1995, **5**, (6), pp. 533–544
5 Qi, Y., Hauptmann, A., and Liu, T.: 'Supervised classification for video shot segmentation'. Proc. IEEE Conf. on Multimedia Expo (ICME), 2003, Vol. 2, pp. 689–692
6 'TREC Video retrieval evaluation', 2005 [Online]. Available: http://www.nlpir.nist.gov/projects/trecvid/