

Camera Motion Analysis Towards Semantic-Based Video Retrieval in Compressed Domain

Ying Weng and Jianmin Jiang

School of Informatics, University of Bradford, BD7 1DP, UK
{Y.Weng, J.Jiang1}@bradford.ac.uk

Abstract. To reduce the semantic gap between low-level visual features and the richness of human semantics, this paper proposes new algorithms, by virtue of the combined camera motion descriptors with multi-threshold, to automatically retrieve the semantic concepts, i.e., close-up, and panorama, directly in MPEG compressed domain based on camera motion analysis. Extensive experiments illustrate that the proposed algorithms provide promising retrieval results under real-time application scenario and without human intervention.

Keywords: Camera motion analysis, MPEG compressed domain, semantic-based video retrieval.

1 Introduction

The ability to retrieve visual information by semantic terms in compressed domain will bring new opportunities for efficient utilisation of multimedia material in improved ways. While considerable research has been conducted in image/video retrieval [1]-[3], only few solutions have been given to this challenging task with limited decoding of MPEG videos. To this end, our work focuses on using low-level descriptors to automatically retrieve high-level semantic concepts, i.e., close-up, and panorama, directly in MPEG compressed domain based on camera motion analysis.

Previous research on close-up detection is mostly in sports video [5]. To improve the coverage of close-up retrieval and implement it extensively, zoom-in descriptor is extracted directly in compressed domain and when its value is larger than a certain two-level threshold, the frame is retrieved as a close-up.

Panorama is very common to everyone, carries more information and provides a larger field of view, thus panorama retrieval is significant for applications [4]. We investigate automatic panorama retrieval in compressed domain exploiting the combined descriptors, i.e., zoom-out and pan, with two-layered threshold.

2 Close-Up Retrieval

The camera motion descriptor is utilised, i.e., zoom-in, with two-level threshold to automatically retrieve close-ups. Let the class labels for thresholding be denoted as TH_S and TH_M ; the former represents the threshold for a single detected zoom-in

frame, and the latter represents the threshold for continuous multiple detected zoom-in frames. Then the thresholding operation is as follows: (i) If only one single zoom-in frame is detected, then the threshold is set to TH_S . Moreover, on the condition that its zoom factor μ has a higher value than TH_S , this frame is retrieved as a close-up; (ii) If continuous multiple zoom-in frames are detected, then the threshold is set to TH_M . Meanwhile, the maximum zoom factor μ_{max} and the minimum zoom factor μ_{min} for these continuous frames are obtained. If the ratio of μ_{max} to μ_{min} is larger than TH_M , these frames are retrieved as close-ups.

The above can be summarised in (1) and Fig. 1:

$$C_F = \begin{cases} F_S, & \text{if } \mu > TH_S \\ F_M, & \text{if } \frac{\mu_{max}}{\mu_{min}} > TH_M \end{cases} \quad (1)$$

where F_S denotes a single detected zoom-in frame, F_M denotes continuously multiple detected zoom-in frames, and C_F denotes retrieved close-ups.

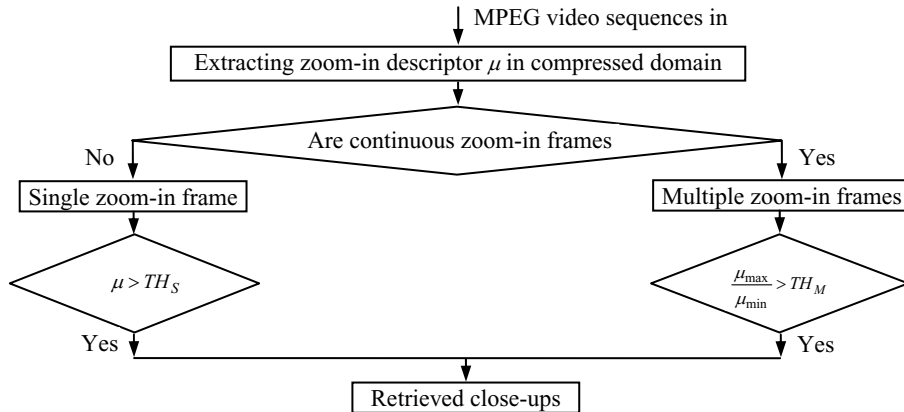


Fig. 1. Overview of the proposed automatic close-up retrieval in MPEG compressed domain

3 Panorama Retrieval

Automatic panorama retrieval exploits the combined camera motion descriptors, i.e., zoom-out and pan, with two-layered threshold is investigated. Let the class labels for thresholding be denoted as TH_Z and TH_P ; the former represents the threshold for a detected zoom-out frame, and the latter represents the threshold for panning. The approach is divided into two steps: (i) On the condition that a zoom-out frame is detected, the first threshold is set to TH_Z , and while its zoom factor μ is less than TH_Z , this frame is selected as a candidate; (ii) If a candidate frame is further detected with

panning, the second threshold is set to TH_p , and if its pan rate $f\alpha$ has a higher absolute value than TH_p , this frame is retrieved as a panorama frame.

The above can be summarised in (2) and Fig. 2:

$$\begin{aligned} F_C &= F_Z, & \text{if } \mu < TH_Z; \\ P_F &= F_C, & \text{if } |f\alpha| > TH_p. \end{aligned} \quad (2)$$

where F_Z denotes a detected zoom-out frame, F_C denotes a candidate frame, and P_F denotes a retrieved panorama.

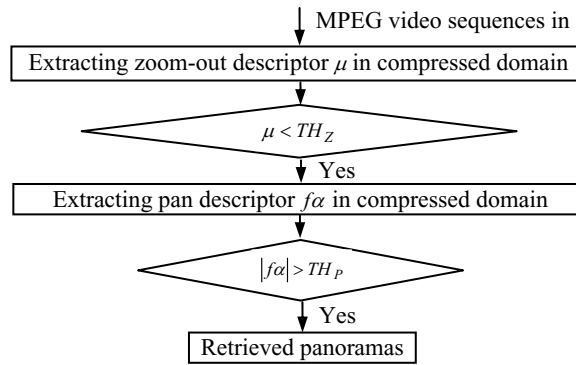


Fig. 2. Overview of the proposed automatic panorama retrieval in MPEG compressed domain

4 Experimental Results and Evaluations

The proposed automatic close-up and panorama retrieval algorithms are tested on a database with different MPEG video clips, including well-known TREC2001 documentary video sequences, movies, sports, and news. TH_S is set to 1.001, TH_M is set to 4, TH_Z is set to 0.998, and TH_p is set to 4. The recall and precision rates for close-up and panorama retrievals are listed in Table 1. The results demonstrate that the proposed algorithms are computationally efficient, also achieve superior performances in terms of both recall and precision rates. The samples of retrieval results from the test video clips are shown in Fig. 3 and 4, respectively.

Table 1. Summary of experimental results for close-up and panorama retrieval

Close-up	Videos	NAD32	NAD55	Movie1	Movie2	Sports1	News1	Average
	Recall	86.67%	95.00%	90.63%	92.86%	93.42%	94.74%	92.22%
	Precision	89.66%	87.69%	81.69%	78.00%	92.21%	94.74%	87.33%
Panorama	Videos	BOR08	BOR11	BOR19	Movie3	Sports2	Sports3	Average
	Recall	89.74%	90.20%	86.02%	88.46%	87.50%	91.84%	89.60%
	Precision	86.42%	85.19%	91.95%	82.14%	90.32%	93.75%	88.89%



(a) from NAD32 (b) from NAD55 (c) from Movie1 (d) from Sports1 (e) from News1

Fig. 3. Samples of close-up retrieval results from the test database



(a) from BOR08 (b) from BOR11 (c) from BOR19 (d) from Movie3 (e) from Sports2

Fig. 4. Samples of panorama retrieval results from the test database

5 Conclusions

In this paper, we build close-up retrieval via zoom-in descriptor with two-level threshold, and panorama retrieval via combined descriptors, i.e., zoom-out and pan, with two-layered threshold. The whole process is under real-time application scenario and without human intervention. Extensive experiments show that computational complexity and retrieval performance of the proposed algorithms are well balanced.

Acknowledgments. The authors would like to acknowledge the financial support under European IST FP6 Integrated Project: LIVE (Contract No. IST-4-027312).

References

1. Djordjevic, D., Izquierdo, E.: An object- and user-driven system for semantic-based image annotation and retrieval. *IEEE Trans. Circuits Syst. Video Technol.* 17(3), 313–323 (2007)
2. Jiang, J., Weng, Y., Guo, B., Feng, Y.: Robust-to-rotation texture descriptor for image retrieval in wavelets domain. *Journal of Electronic Imaging* 15(1), 1–14 (2006)
3. Hsieh, J.W., Yu, S.L., Chen, Y.S.: Motion-based video retrieval by trajectory matching. *IEEE Trans. Circuits Syst. Video Technol.* 16(3), 396–409 (2006)
4. Sun, X., Foote, J., Kimber, D., Manjunath, B.S.: Region of interest extraction and virtual camera control based on panoramic video capturing. *IEEE Trans. Multimedia* 7(5), 981–990 (2005)
5. Liu, L., Ye, X., Yao, M., Zhang, S.: A semantic description scheme of soccer video based on MPEG-7. In: Aizawa, K., Nakamura, Y., Satoh, S. (eds.) *PCM 2004*. LNCS, vol. 3332, pp. 298–305. Springer, Heidelberg (2004)
6. Tan, Y.P., Saur, D.D., Kulkarni, S.R., Ramadge, P.J.: Rapid estimation of camera motion from compressed video with application to video annotation. *IEEE Trans. Circuits Syst. Video Technol.* 10(1), 133–146 (2000)