Object-Based Accumulated Motion Feature for the Compressed Domain Human Action Analysis

Cheng-Chang Lien Chen-Yu Hong Yu-Ting Fu

Dept. of CSIE, Chung Hua University
Hsin-Chu, Taiwan, R.O.C.
Email: cclien@chu.edu.tw

Abstract

Generally, the object-based prominent motion features haven't been generated to analyze the human actions in the conventional compressed domain human action analysis systems such that the human actions are identified inaccurately. Here, we apply the video segmentation method to acquire the human region in the compressed domain and then use the macroblock tracking process to trace each macroblock within the human region. Then a new motion feature called object-based accumulative motion vector (OAMV) is generated to extract the prominent motion feature for the human actions. Finally, human actions are identified by using the Hidden Markov Models.

Keywords: Compressed video, Video segmentation, Object-based accumulative motion vector (OAMV), Hidden Markov Models.

1. Introduction

Recently, intelligent surveillance systems are widely applied to the public area security [1-4] and homecare applications [5]. Both applications are established on the basis of human action analysis. Human actions are frequently analyzed by using the uncompressed video data [6-7]. However, the motion features in the abovementioned researches are hard to generate in the compressed video. Furthermore, the amount of uncompressed video data is too large to transmit via the network with limited bandwidth. Human action analysis within the compressed video may not only extract the visual features directly from MPEG compressed video but also overcome the problem of limited network bandwidth.

In [9], the multi-resolution images are utilized to extract the human region and recognize the human activities in the MPEG video sequences. Firstly the region of human body is extracted by grouping the motion vectors with similar magnitude. Then the skin color detection and human action templates are utilized for analyze the activities. However, a lot of templates are necessary to be defined for the template matching process. In [10], the polar histogram of motion vectors is used to analyze the various kinds of human actions.

However, the prominent motion feature doesn't be generated to analyze the human actions in the above-mentioned human action analysis systems such that the human actions are identified inaccurately. Here, we apply the video segmentation method to acquire the human region in the compressed domain and then use the macroblock tracking process to trace each macroblock within the human region. Then a new motion feature called object-based accumulative motion vector (OAMV) is generated to extract the prominent motion feature for the human actions. Based on the observations of human actions, the movements of a complete action usually takes a few seconds, i.e., several GOPs within the compressed video are required to analyze the human actions. Therefore, the sliding window with length of six GOPs is applied to identify each human action. Finally, the human actions are identified by using the Hidden Markov Models. The block diagram of the proposed system is shown in Fig. 1.

2. Object-Based Accumulative Motion Vector (OAMV)

To extract the moving characteristics of various parts of the human body, we proposed a new motion feature called the object-based accumulative motion vector (OAMV) to describe the motion feature for each macroblock in the region of human body (legs, arms, body).

2.1. Object Segmentation

The main objective of object segmentation is to obtain the regions of moving human such that the human actions may be analyzed. Here, the method of recursive shortest spanning tree (RSST) is applied to segment the regions of moving objects including the human body.

The RSST segmentation method [11-12] is proposed to improve spiky phenomenon by using the global information. Given a frame with $m \times n$ macroblocks, the link weight w_{ij}^{c} in the spanning tree can be defined as the absolute difference value of the chrominance DC signals between adjacent blocks i and j, i.e.,

$$w_{ij}^{c} = \left| d_{Cr}^{i} - d_{Cr}^{j} \right| + \left| d_{Cb}^{i} - d_{Cb}^{j} \right|, \quad (1)$$

where d^i and d^j denote the DC values of the chrominance signals for the adjacent macroblocks i and j respectively. Starting from the least-weighted link, the RSST segmentation method merges the vertices of the least weighted link into one new vertex, and then save the link to a new spanning tree. The new vertex value is replaced by the mean of the two adjacent vertices. By searching the least-weighted link and perform the above process repeatedly, a new spanning tree is constructed with the saved links. Finally, the image is partitioned into N regions by cut the new spanning tree at the N-1 largest weighted links.

The image segmentation using the motion vectors may also be obtained by defining the link weight w_{ij}^{m} as:

$$w_{ij}^{m} = \left| m_{x}^{i} - m_{x}^{j} \right| + \left| m_{y}^{i} - m_{y}^{j} \right|, \quad (2)$$

where m^i and m^j denote the motion vectors for the adjacent blocks i and j respectively. By projecting the motion segmented regions onto the color segmented regions, we may acquire the regions of moving human. The segmentation method for extracting the human region is developed on the basis of the projection operator [13] that is defined as:

$$p(I_i, R^M) = M_m \quad (3)$$

where I_i denotes the color segmented region and R^M denotes the motion segmented region. In Eq. (3), the projection operation finds the color segmented regions that overlap with the motion segmented regions. Based on the projection operation, the group operator G [13] defined as Eq. (4) is used to collect all the regions founded in (3).

$$G(I_i, R^M, g) = \{ I_i : p(I_i, R^M) = M_g \} \quad (4)$$

Fig. 2 shows the segmented human region obtained by using the group operator.

2.2. OAMV Motion Feature

Generally, a complete human action often takes a few seconds. Hence, several GOPs are required to extract the visual features for recognizing the human action. To obtain the prominent motion feature for human action a new GOP-based motion descriptor called the object-based accumulative motion vector (OAMV) is proposed. Fig. 3 describes the concept of the OAMV motion feature. In Fig. 3, the jumping action is illustrated and the red line segments denote the motion vectors. It is obviously that the motion vectors between the adjacent frames have high correlations. If the motion vector for a tracked block on the successive frames is accumulated, then the trajectory of the tracked block may be constructed to describe the body movements.

By continuously tracking each block in the region of human body within each GOP, the OAMV motion feature is constructed with the following steps:

1. Initialize a starting frame and record the center position for each tracked block in the region of human body.
2. For the P-frame, by using the extracted motion vectors the new center position for each block is updated with the center position of the tracked block. Record the updated center position for each tracked block within the object region.
3. Repeat the process in step 2 until all the P frames in the GOP are searched.
4. For each GOP, construct the trajectory (OAMV motion feature) for each block in the region of human body.

In Fig. 4(a)-(b), the GOP-based OAMV features for the tumbling down action are illustrated. Each MB within the human region is tracked for time period of a GOP and then the OAMV feature is constructed.

2.3. Polar Histogram

In [10], the polar histogram is used to extract the distribution pattern for each human action such that the human actions may be identified. Here, the polar histogram is also applied to extract the distribution pattern for the OAMV motion feature in the compressed domain. The OAMV features for different human actions will distribute with different patterns in the polar histograms. Fig. 5 illustrates the OAMV feature on the polar coordinate for the running and walking actions. Although the running and walking actions looks alike, the OAMV feature still can discriminate the two similar actions. The polar histogram is generated by transforming the OAMV motion features from Cartesian coordinate to polar coordinate. There are totally 32 bins in the polar histogram. In order to enhance the polar histogram the squared polar histogram is applied.

3. Human Action Analysis

Generally, a human action often takes a few seconds (several GOPs) to complete. By using the polar histograms among the successive GOPs and the Hidden Markov Models, we may identify the various kinds of human actions. Since the human actions in the video shot may change irregularly, the ergodic HMM [14] is appropriate to be applied for identifying the human

actions.

In HMM, the methods of vector quantization or K-means are frequently used to generate the observation symbols by clustering the acquired features. By the careful observation of human action, a complete action will takes about the time period of six GOPs. Hence, the sliding window with length of six GOPs is applied to identify each human action. In general, HMM is expressed by a 3-tuple parameters $\boldsymbol{\lambda}=(\mathbf{A}, \mathbf{B}, \pi)$, where π is the initial state distribution, $\mathbf{A}=[a_{ij}]$ is a matrix of state transition probabilities with elements a_{ij} being the state transition probability from state i to state j, and $\mathbf{B}=\{b_j(k)\}$ is the observation symbol probability distribution for symbol o_k in state j. By training the parameters $(\mathbf{A}, \mathbf{B}, \pi)$ for each human action, the various kinds of actions may be identified.

In this paper, five kinds of human actions (walking, running, tumbling down, jumping and crouching) are identified by the HMM method. Based on the clustered polar histograms, the observation symbol sequence with length T denoted as $\mathbf{O}=[o_1, o_2, \ldots, o_T]$ may be constructed and then the parameters $(\mathbf{A}, \mathbf{B}, \pi)$ for the 4-state ergodic HMM may be acquired by using the Baum-Welch algorithm [14]. Given the observation symbol sequence $\mathbf{O}$, the observation probability $P(\mathbf{O}|\boldsymbol{\lambda})$ for each detected event may be computed via the forward-backward procedure. In [14], a forward variable $\alpha_t(i)$ is defined to calculate the probability of partial observing sequence of the state i at time t for the model $\boldsymbol{\lambda}$ denoted as $P(o_1 o_2 \ldots o_t, q_t=i|\boldsymbol{\lambda})$. By using the probability of partial observing sequence, $P(\mathbf{O}|\boldsymbol{\lambda})$ can be calculated as follows:

1) Initialization: $\alpha_i(i) = \pi_i b_i(o_1)$

2)Induction: $\alpha_{t+1} = [\sum_{i=1}^{N} \alpha_t(i) a_{ij}] b_j(o_{t+1}),\ 1 \le t \le T-1, 1 \le j \le N$

3) Termination: $P(O|\lambda) = \sum_{i=1}^{N} \alpha_T(i)$

Here, all the values of initial transition probabilities a_{ij} are equally distributed with value 0.25 and all the values of observation probabilities $b_j(O)$ for all states are also equally distributed with value of 0.1 since there are totally 10 observation symbols.

4. Experimental Results

In our experiments, the video format is MPEG-II with resolution 352×240 pixels and the frame rate is 30 frames per seconds. Based on the video segmentation method mentioned in section 2.1, Fig. 6 shows the segmented region of moving human for different actions. In Fig. 6, the images on the left side show the moving human with different actions and those on the right side illustrate the segmented regions of moving human respectively. Once the region of moving human is segmented, we may acquire the OAMV feature to identify the various kinds of human actions. By continuously tracking each block in the region of human body for each GOP, the OAMV motion feature may be obtained. The 2-D polar distributions for the OAMV features of two kinds of actions are illustrated on the left side of Fig. 7 and the corresponding polar histograms are shown on the right side. It is obvious that the OAMV features for different human actions will distribute with different patterns in the polar histograms. In Table 1, five actions are detected and the accuracy analysis of the action identification is given. There are total 75 video clips are used to train the Hidden Markov Model and 15 test video clips are used to analyze the accuracy of human action identification.

5. Conclusion

This paper proposed an effective and robust method to detect the rare behavior events within the compressed video directly. New motion feature called object-based accumulative motion vector (OAMV) is generated to extract a prominent motion feature and then polar histograms are used to describe the distribution patterns for each human action. The various kinds of human actions are identified by the HMM method. Experimental results show that the human actions may be identified accurately.

References

[1] M. Leo, T. D'Orazio, I. Gonoi, P. Spagnolo, A. Distante (2004) Complex human activity recognition for monitoring wide outdoor environments. In: Proc. 17th international conference on pattern recognition, Aug. 2004, 4: 913-916

[2] J. A. Freer, B. J. Beggs, H. L. Fernandez-Canque, F. Chevrier, A. Goryashko (1996) Automatic video surveillance with intelligent scene monitoring and intruder detection. In: Proc. 30th international Carnahan conference, Oct. 1996, pp. 89-94

[3] A. Divakaran, K. Vetro, K. Asal, H. Nishikawa (2000) Video browsing system based on compressed domain feature extraction. IEEE Trans. on Consumer Electronics, Aug. 2000, 46(3):637-644

[4] R. T. Collins, A. J. Lipton, T. Kanade, H. Fujiyoshi, D. Duggins, Y. Tsin, D. Tolliver, N. Enomoto, O. Hasegawa, P. Burt, L. Wixson (2000) A system for video surveillance and monitoring: VSAM final report. Technical Report CMU-RI-TR-00-12, Robotics Institute, Carnegie Mellon University, May 2000.

[5] D. Liu, P. C. Chung, Y. N. Chung (2004) Human home behavior interpretation from video streams. In: Proc. IEEE International conference on networking, sensing and control, Mar. 21-23, 2004, 1:192-197

[6] J. Yamato, J. Ohya, K. Ishii (1992) Recognizing human action in time-sequential images using hidden Markov model. In: Proc. Computer vision and pattern recognition, June 1992, pp. 379-385

[7] R. Rosales (1998) Recognition of human action using moment-based features. Technical Report BU 98-020, Boston University, Computer Science, 1998.

[8] M. Umeda (1982) Recognition of multi-font printed Chinese characters. In: Proc. 6th ICPR, 1982, pp. 793-796

[9] B. Ozer, W. Wolf, A. N. Akansu (2000) Human activity detection in MPEG sequences. In: Proc. Workshop on human motion, Dec. 7-8, 2000, pp. 61-66
[10] R. V. Babu, B. Anantharaman, K. R. Ramakrishnan, S. H. Srinivasan (2002) Compressed domain action classification using HMM. Pattern Recognition Letters, Aug. 2002, 23(10): 1203-1213
[11] J. Morris, M. J. Lee, A. G Constantinides (1986) Graph theory for image analysis: an approach base on the shortest spanning tree. In Proc. Inst. Elect. Eng., Apr. 1986, 133(2):146-152
[12] E. Tuncel, L. Onural (2000) Utilization of the recursive shortest spanning tree algorithm for video-object segmentation by 2-D affine motion modeling. IEEE Trans. Circuit and System for Video Technology, Aug. 2000, 10(5):776-781
[13] A. Alatan, E. Tuncel, L. Onural (1997) A rule-based method for object segmentation in video sequences. In: Proc. IEEE international conference on image processing ICIP, Oct. 1997, 2:522-525
[14] L. Rabiner, B.-H. Juang (1993) Fundamentals of speech recognition. Prentice-Hall

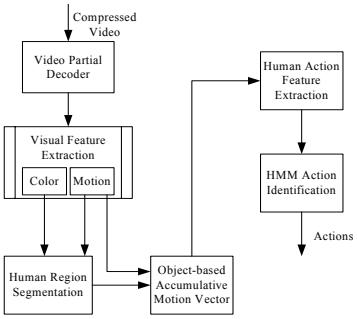


Fig. 1 Block diagram of the human action analysis system.

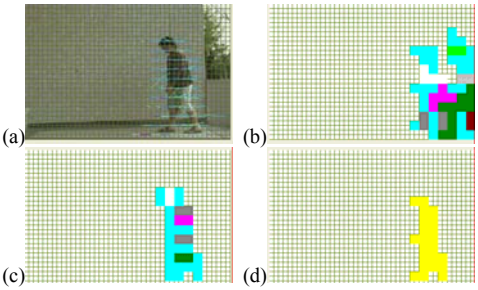


Fig. 2 (a) I-frame for a human walking video clip. (b) Color segmentation mask. (c) Motion segmentation mask. (d) Color segmented regions projected onto motion segmented regions.

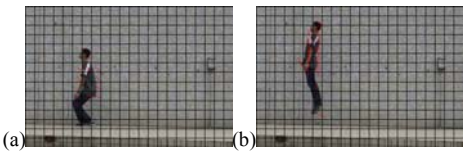
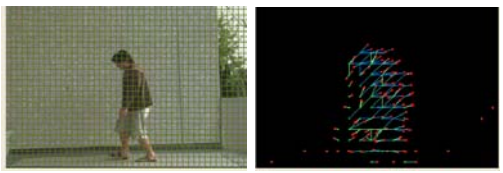
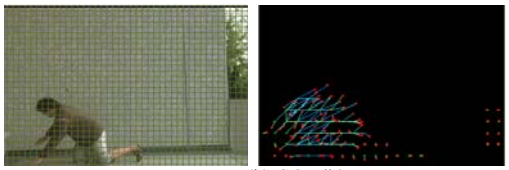


Fig. 3 Motion vectors for the jumping action on (a) frame #122 and (b) frame #131.

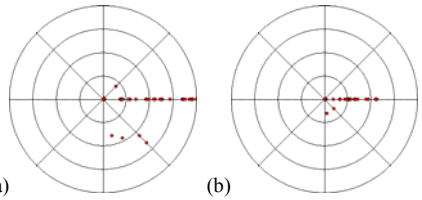


(a) GOP #4



(b) GOP #6

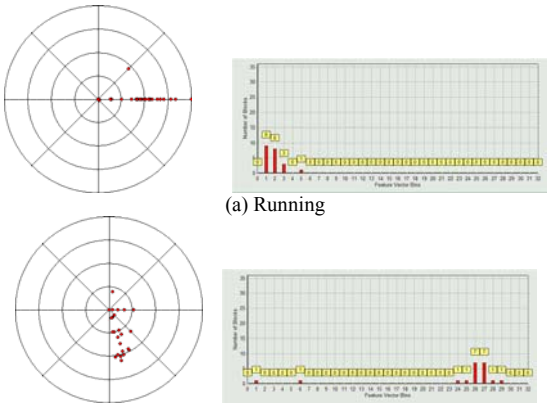
Fig. 4 GOP-based OAMV features for the tumbling down action are illustrated in (a) to (b). Each MB within the human region is tracked for time period of a GOP and then the OAMV feature is constructed.



(a) (b)

Fig. 5 The OAMV features on the 2-D polar coordinate for (a) running action and (b) walking action.

Fig. 6 Images on the left side show the moving human with different actions and images on the right side illustrate the segmented regions.



(a) Running

(b)Crouching

Fig. 7 The 2-D polar distributions for the OAMV features of two kinds of actions are illustrated on the left side and the corresponding polar histograms are shown on the right side.

Table 1 Accuracy analysis of action identification.

Actions	The number of test video clips	The number of false action classification	The correct recognition rate
Walking	15	3	80%
Running	15	1	93%
Jumping	15	3	80%
Crouching	15	2	87%
Tumbling down	15	2	87%