Automatic Key Posture Selection for Human Behavior Analysis

Duan-Yu Chen [a], Hong-Yuan Mark Liao [a]

[a] Institute of Information Science
Academia Sinica, Taiwan
{dychen, liao}@iis.sinica.edu.tw

Hsiao-Rang Tyan [b], and Chia-Wen Lin [c]

[b] Department of Information and Computer Engineering
Chung-Yuan Christian University, Taiwan
[c] Department of CSIE
National Chung-Cheng University, Taiwan

***Abstract*—A novel human posture analysis framework that can perform automatic key posture selection and template matching for human behavior analysis is proposed. The entropy measurement, which is commonly adopted as an important feature to describe the degree of disorder in thermodynamics, is used as an underlying feature for identifying key postures. First, we use cumulative entropy change as an indicator to select an appropriate set of key postures from a human behavior video sequence and then conduct a cross entropy check to remove redundant key postures. With the key postures detected and stored as human posture templates, the degree of similarity between a query posture and a database template is evaluated using a modified Hausdorff distance measure. The experiment results show that the proposed system is highly efficient and powerful.**

Keywords*—*human behavior analysis; key posture selection; exponential entropy

Topic area*—*Multimedia Databases - indexing and retrieval

I. Introduction

Human posture analysis is one of the most important steps towards successful human behavior analysis. The difficulty of human posture analysis is twofold. First, the movement of a human body is an articulated motion. Therefore, it is obvious that the issue to be addressed is a problem with high dimensionality and complexity. Second, characterization of human behavior is equivalent to dealing with a sequence of video frames that contain both spatial and temporal information. The most challenging issue is how to properly characterize spatial-temporal information and then facilitate subsequent comparison/retrieval tasks. The objective of this study is to propose a methodology for systematic human posture analysis, not simply conducting a thorough human behavior analysis. Therefore, the detailed descriptors, such as the movement of hands, feet or torso, are not the main subjects of this research. The posture classification systems proposed in the past can be categorized into two classes, i.e., 2-D based or 3-D based, depending on the types of human body model adopted [1]. In [2], Haritaoglu et al. proposed the W4 system which computed the vertical and horizontal projections of a silhouette to determine the global posture of a person (standing, sitting, bending and lying). Bobick and Davis [3] proposed a temporal template built by stacking a set of consecutive frames. The proposed temporal template characterized human motion by using motion energy images (MEI) and motion intensity images (MHI). Moment based features were extracted from MEI and MHI and they used these moment based features to conduct template matching. It is also possible to perform human behavior analysis based on human silhouette analysis [6-8]. In [7], the projection histograms of each person were computed and compared with the probabilistic projection maps stored for each posture during the training phase. For the purpose of posture classification reliability, the obtained posture was further validated exploiting the information extracted by a tracking module. In addition to 2-D model-based systems introduced above, there are also some existing 3-D model-based systems. Boulay et al. [4] first computed projections of moving pixels on a reference axis and learned 2-D posture appearances through PCA. Then, they employed a 3-D model of posture to make the projection-based method independent of the camera position. Zhao et al. [5] used a 3-D human model to verify whether a moving region detected is a person or not. The verification process was done by walking recognition using an articulated human walking model. However, due to the need of developing low-cost systems, complex computations and expensive 3-D solutions will not be considered in this study. In order to achieve the goal of building an efficient and automated human behavior analysis system, the first important step is to identify the significant postures from a human behavior video sequence systematically and automatically. To the best of our knowledge, none of the previous researches did try to deal with automatic key posture selection and comparison problems.

In this paper, we propose a systematic way to conduct automatic key posture selection. It is well known that a period of human behavior can be recorded by a sequence of video frames. Since the data amount is huge, it is not feasible to characterize the video sequence frame by frame. Under these circumstances, an automatic key posture selection algorithm that can systematically determine some representative "key" postures from the video sequence is indispensable. A good key posture selection algorithm must satisfy two criteria. First, the number of selected key postures cannot be too few. Otherwise, the set of key postures is not enough to "completely describe" the original video sequence. Second, the number of selected key postures cannot be too many because it will cause the redundancy problem and the subsequent characterization process will be time-consuming. Therefore, we propose to use cumulative entropy change as an indicator to select an appropriate set of key postures from a human behavior video sequence. Then we conduct a cross entropy check to remove some redundant key postures that were selected in the previous step. For every human action (a period of video that contains human behavior) recorded as a video sequence, we use the above strategy to choose a set of key postures from it. With the set of key postures stored as the elements of a codebook, we are able to "characterize" (or "encode") the original video sequence into a number sequence. For different human action video sequences, one can always encode them into corresponding number sequences. In the matching process, the problem is simplified into a number sequence comparison problem.

The rest of this paper is organized as follows. In Section 2 we shall introduce the theory of entropy. The proposed approach will be described in Section 3. Experiment results will be shown in Section 4. Conclusions will be drawn in Section 5.

II. ENTROPY

Entropy is a distribution-based parameter that is commonly used to describe the degree of disorder of a system. In thermodynamics, the entropy of a system is usually defined as follows [9]:

$$S = k \ln \Omega , \tag{1}$$

where Ω is the probability of thermodynamics. k and S, respectively, represent a constant and the entropy. Entropy is an indicator showing the degree of disorder. When the internal energy of a system is increased, its corresponding entropy is increased too. Usually, stability of a system will be reached when the molecules within the system are uniformly distributed. Under these circumstances, the value of entropy reaches its maximum. Shannon's entropy, which is commonly used in multimedia signal processing, is based on the concept that the gain in information from an event is inversely proportional to the probability of occurrence p_i . Therefore, Shannon's entropy can be defined as follows [10]:

$$H = -\sum_{i=1}^{n} p_i \log p_i \tag{2}$$

Assume that the probability distribution is of the Shannon type, i.e. $\sum_{i=1}^{n} p_i = 1$. Thus, the probability of every component must be greater than zero. If a probability approaches zero, its corresponding information gain, $\Delta I(p_i) = -\log(p_i)$, cannot be defined. Here, we adopt the so-called exponential entropy defined by Pal and Pal [11]. In their approach, the gain information is defined as

$$I = \exp(1 - p_i) . \tag{3}$$

Furthermore, the gain information that can be derived from an event is inversely proportional to its probability of occurrence p_i . In addition, the exponential entropy can be defined as the expected value of the gain function, i.e.

$$H = E(I) = \sum_{i=1}^{n} p_i \exp(1 - p_i). \tag{4}$$

From (4), it is obvious that the maximum entropy will be reached when all the constituent probabilities are equal. In addition, the exponential entropy has a firm upper bound for a uniform distribution. When the probabilities of occurrence are uniformly distributed, (4) becomes:

$$H(1/n, 1/n, ..., 1/n) = \exp\left(1 - \tfrac{1}{n}\right), \tag{5}$$

and n becomes infinite, we have

$$\lim_{n\to\infty} H(1/n, 1/n, ..., 1/n) = \lim_{n\to\infty} \exp\left(1 - \tfrac{1}{n}\right) \cong e \quad . \tag{6}$$

III. THE PROPOSED METHOD

In this section, we shall first extract the spatial features from video objects and then characterize the shape of video objects using the probabilities calculated from each block. Then, we use an entropy-based method to detect the key postures from a video sequence that records human behavior and then remove redundancy from the detected postures. Finally, the degree of similarity between a query posture and a database key posture is evaluated using a modified Hausdorff distance measure.

A. *Characterizing Shape by Block Probabilities*

To characterize the shape of a video object while maintaining low computation cost, we decompose a video object into blocks and characterize the corresponding posture by the density of active pixels residing in each block. First, we compute the aspect ratio of a video object. Assume the height to width ratio of the bounding rectangle of the above video object is *h/w*. If we partition the rectangle into $h \times w$ blocks as shown in Fig. 1(a), then the probability of each block can be calculated as $p_i = N_i / N$, where N_i is the number of active pixels in the *i*th block of the bounding box as illustrated in Fig. 1(b), and N indicates the total number of pixels in a block. Therefore, the exponential entropy H of a video object can be defined as

$$H(t) = \sum_{i=1}^{h\times w} p_i \exp(1 - p_i) \cdot \tag{7}$$

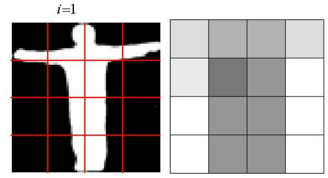

Fig. 1 (a) Partition of bounding box (b) corresponding block probability represented by different intensity levels

B. *Automatic Key Posture Selection Using Entropy*

To make the system efficient enough for real-time application, we use the entropy property, which is based on the concept that if a bounding box contains more blocks having connected components, then its corresponding entropy is larger. The proof of why an entropy value is proportional to the number of connected component-occupied blocks is as follows. Let the total number of blocks in a bounding box be n. Consider two different cases: in case 1, there are n' blocks containing connected components, and in case 2 there are n'' blocks containing connected components. If $n' < n'' < n$ is true, then based on the Schur-concavity property [12] the entropy calculated for case 2 will be larger than that calculated for case 1. Let the probability distributions of case 1 and case 2 be $P = (p(1), p(2), p(3), ..., p(n'), ..., p(n))$ and Q = $(q(1), q(2), q(3), ..., q(n''), ..., q(n))$, respectively. In the P sequence, the components after $p(n')$ are all zeros. On the other hand, the components before $p(n'+1)$ are all non-zeros but have no order. In the case of the Q sequence, the components preceding $q(n''+1)$ are non-zeros and are also non-ordered. According to the Schur-concavity property, one can always pick up non-zeros components without repetition from P and Q to satisfy

$$\sum_{i=1}^{k} q(i) \le \sum_{i=1}^{k} p(i) = 1 \text{ for } k = 1,2,...,n. \tag{7}$$

If Eq.(7) holds, then we can say that Q is majorised by P [12], and that it can be represented by $Q \prec P$. Since the exponential entropy H is also Schur-concave, it will make the above-mentioned P and Q functions satisfy

$$Q \prec P. \tag{8}$$

This implies that

$$H(P) \prec H(Q). \tag{9}$$

In the proposed method, we use the exponential entropy to characterize video objects, that is, to calculate the shape information using exponential entropy. In addition, we define a distance measure between two successive postures as follows:

$$D_{posture}(t,t+1) = \left|H(t+1) - H(t)\right|. \tag{10}$$

A posture is identified as a key posture if the cumulative value of $D_{posture}(t,t+1)$, which measures the exponential entropy difference between the current posture and the previous posture, is larger than a preset threshold. The value of cumulative exponential entropy change between two arbitrary postures, P and Q, can be calculated as follows:

$$C(P_P,P_Q) = \sum_{t=P}^{Q-1} D_{posture}(t,t+1), \tag{11}$$

where P_P denotes a posture detected at time instance p. With the cumulative exponential entropy change value defined by Eq.(11), one can detect key postures from a video sequence. For an arbitrary video sequence to be characterized, we choose the first posture as a candidate key posture. Then, we calculate the cumulative exponential entropy change value from this starting key posture. If the cumulative entropy change exceeds a preset threshold, we choose the posture at that time instance as the second key posture. The process is repeated until the video sequence is complete. Fig. 1(a) shows the detected 18 key postures out of 200 successive ones. Fig. 1(b) shows the exponential entropy values indicated by circles corresponding to the 18 key postures shown in Fig.1(a). The horizontal axis of Fig.1(b) indicates the frame number, and the vertical axis represents the corresponding exponential entropy value.Equations

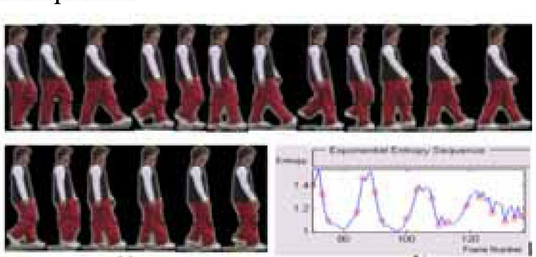

Fig. 1 (a) A set of detected key postures; (b) the exponential entropy values corresponding to the set of detected key postures shown in (a).

C. Redundancy Elimination Using Cross Entropy

In the previous section, we have systematically determined a set of key postures by judging the cumulative entropy change. However, the number of selected key postures is too many to be accepted. Therefore, we have to execute a redundancy removal process to make the number of selected key postures reasonable. To eliminate some key postures chosen from the previous step, we calculate the cross exponential entropy to measure the dissimilarity between key postures. A set of redundant key postures will be eliminated when the degree of dissimilarity $d(P_P,P_Q)$ is larger than a preset threshold. The degree of dissimilarity $d(P_P,P_Q)$ is defined as

$$d(P_P,P_Q) = \sum_{i=1}^{h\times w} P_P(i)\exp\left(\left|P_P(i) - P_Q(i)\right|\right) + \sum_{i=1}^{h\times w} P_Q(i)\exp\left(\left|P_Q(i) - P_P(i)\right|\right). \tag{11}$$

Fig. 2 illustrates the 18 key postures identified in the first step is now reduced to 5 key postures (highlighted in the rectangle). Using the degree of dissimilarity as a screening tool, the extracted key postures are more representative than those detected without redundancy removal. In addition, the matching process based on this reduced model set will be much more efficient.

Fig. 2 The key postures highlighted in the rectangle are the key postures without redundancy.

D. Template Matching Using Modified Hausdorff Distance

In this section, we describe how the matching process operates. In order to compute the degree of similarity between a query and a template T in a key posture database, we use the modified Hausdorff distance, $D(Q,T)$, to do this job. The definition of $D(Q,T)$ is as follows:

$$D(Q,T) = \sum_{i=1}^{h\times w} Q(i)\exp\left(\left|Q(i) - T(i)\right|\right) + \sum_{i=1}^{h\times w} T(i)\exp\left(\left|T(i) - Q(i)\right|\right). \tag{12}$$

The best matched key posture T_{match} can be determined by

$$T_{match} = \arg\min_{\lambda}\left(D(Q,T_\lambda)\right), \tag{13}$$

where $\lambda \in \{1,2,...,k\}$ and T_λ indicates the λth key posture in the database.

IV. EXPERIMENT RESULTS

A series of experiments were conducted to test the effectiveness of the proposed method. The data set used in the experiments was a real video sequence consisting 33 shots (6100 frames). Using the exponential entropy and cross entropy measure proposed in this paper, we were able to automatically choose 44 key postures from the original 6100 frames. Fig.3 illustrates the set of 44 key postures extracted from the original test video. The upper-left of Fig.3 shows an instance of a walking person and the posture corresponding to this walking person at this instance is key posture #7 (highlighted by a bold rectangle). The top of Fig. 3 illustrates a sequence of numbers indicating how the posture of the walking person evolved during the walking sequence. Fig. 4(a) shows the complete sequence of the walking process. Fig. 4(b) shows the sequence of the matched key postures. Note that the first three frames were matched to key posture #5 and the fourth and fifth frames of the walking sequence were matched to key posture #6. Fig. 4(c) shows only 6 key postures (#5, #6, #7, #8, #38, #39) were matched during the key posture matching process if the walking sequence shown in Fig. 4(a) was adopted. Using the simple encoded sequence shown on the top of Fig.3, one can encode a continuous sequence of action easily. Fig.5 shows the encoding process of a lying-to-standup sequence. Fig.5(a) shows the whole human action process from lying to stand up. Fig.5(b) shows the corresponding matched key postures. And Fig.5(c) shows the encoded key posture sequence.

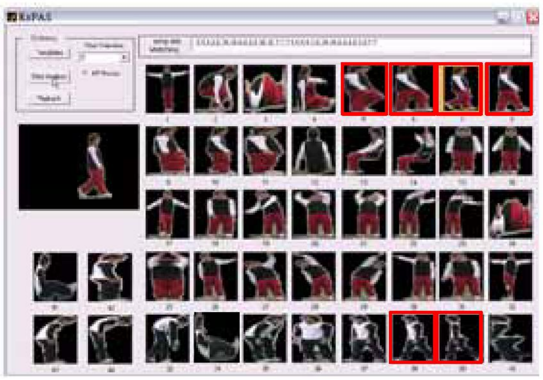

Fig. 3 The set of key postures chosen from the test video.

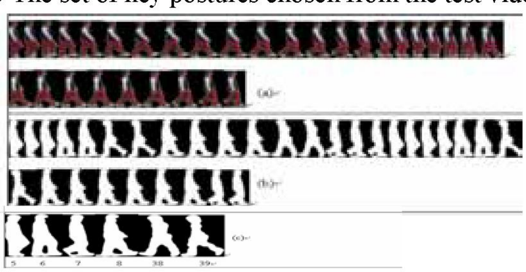

Fig. 4 Demonstration of template matching using a walking sequence (a) original video sequence (b) corresponding matched key posture sequence and (c) the key postures used in the matching process.

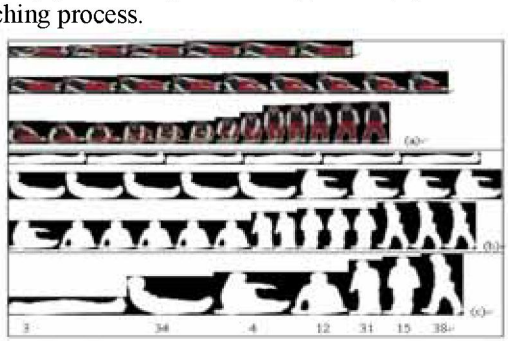

Fig. 5 Demonstration of human behavior encoding using a lying-to-standup sequence (a) original video sequence (b) corresponding matched key postures and (c) encoded key posture sequence.

V. CONCLUSIONS

A novel human posture analysis framework that can perform automatic key posture selection and template matching for human behavior analysis has been developed in this work. The proposed framework has three special features: 1) the cumulative entropy change has been employed as an indicator to select an appropriate set of key postures from a human behavior video sequence; 2) a cross entropy check has been conducted to remove some redundant key postures; 3) with the key postures detected and stored as human posture templates, the degree of similarity between a query posture and a database template has been evaluated using a modified Hausdorff distance measure. The experiment results show that the proposed system is indeed efficient and powerful.

ACKNOWLEDGEMENT

This work is supported by the Ministry of Economic Affairs under contract no. 94-EC-17-A-02-S1-032

REFERENCES

[1] T. Moeslund and E. Granum, "A Survey of Computer Vision-based Human Motion Capture," *Computer Vision and Image Understanding*, Vol.81, no.3, pp.231-268, Mar. 2001.

[2] I. Haritaoglu, D. Harwood, and L. Davis, "Ghost: A Human Body Part Labeling System Using Silhouettes," in *Proc. Int. Conf. Pattern Recognition*, Vol.1, pp.77-82, 1998.

[3] A. F. Bobick and J. W. Davis, "The Recognition of Human Movement Using Temporal Templates," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol.23, no 3, Mar. 2001.

[4] B. Boulay, F. Bremond, and M. Thonnat, "Human Posture Recognition in Video Sequence," *in Proc. IEEE Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance*, pp. 23-29, 2003.

[5] T. Zhao, R. Nevatia and F. Lv, "Segmentation and Tracking of Multiple Humans in Complex Situations," in *Proc. IEEE Int. Conf. on Computer Vision and Pattern Recognition*, Vol.2, pp.08-12, 2001.

[6] H. Fujiyoshi and A. Lipton, "Real-Time Human Motion Analysis by Image Skeletonization," in *Proc. IEEE Workshop on Application Computer Vision*, pp.15-21, 1998.

[7] I. C. Chang and C. L. Huang, "The Model-based Human Body Motion Analysis System," *Image Vision Computation*, Vol.18, no.14, pp.1067-1083, 2000.

[8] R. Cucchiara, c. Grana, A. Prati, and R. Vezzani, "Probabilistic Posture Classification for Huamn-Bahavior Analysis," *IEEE Trans. on Systems, Man, and Cybernetics-Part A: Systems and Humans*, Vol.35, no.1, pp.42-54, Jan. 2005.

[9] D. V. Schroeder, "An introduction to thermal physics," Addison Wesley, San Francisco, USA, 2000.

[10] T. O. Kvalseth, "On exponential entropies," *Proc. IEEE Int. Conf. on System, Man, and Cybernetics*, pp. 2822-2826, 2000.

[11] N. R. Pal, and S. K. Pal, "Entropy: a new definition and its applications," *IEEE Trans. Syst., Man Cybern.*, Vol. 21, no. 5, pp. 1260-1270, 1993.

[12] A. W. Marshall and I. Olkin, "Inequalities: Theory of majorization and its applications," Academic Press, San Diego, CA, USA, 1979.