SVM-based State Transition Framework for Dynamical Human Behavior Identification

Chen-Yu Chen[1], Jia-Ching Wang[2], Jhing-Fa Wang[2], and Li-Pang Shieh[2]

[1] Institute for Information Industry, 3F., No. 2, Fusing 4th Rd, Kaohsiung City 80661, Taiwan, R.O.C.
[2] Department of Electrical Engineering, National Cheng Kung University, No.1, University Road, Tainan City 701, Taiwan, R.O.C
cy.chen@iii.org.tw

ABSTRACT

This investigation proposes an SVM-based state transition framework (named as STSVM) to provide better performance of discriminability for human behavior identification. The STSVM consists of several state support vector machines (SSVM) and a state transition probability model (STPM). The intra-structure information and inter-structure information of a human activity are analyzed and correlated by the SSVM and STPM, respectively. The integration of the SSVM and the STPM effectively provides human behavior understanding. With a database consisting of five kinds of human behaviors: raising hand, standing up, squatting down, falling down, and sitting, the proposed algorithm has been demonstrated with a significant recognition rate of 88.6%.

Index Terms: Image processing, pattern recognition, user interface human factors

1. INTRODUCTION

Ambient intelligence environments allow people to friendly access information for interaction between users and devices. Human behavior identification is a core technique for ambient intelligent applications. In [1-4], model-based methods are used to achieve the goal of human behavior identification. To construct the behavior model, a human body would be divided into several parts to define poses and to capture human motions, etc. Meaningful human activities could further be detected with different combination of poses or motions.

To overcome the problem of tracking lost, two related research works respectively using model-based and HMM-based approaches have been proposed. Bregler [5] proposed a probabilistic decomposition of human dynamics at multiple abstractions, such as level of input image sequence, coherence blob hypotheses, simple dynamical categories, and complex movement sequences. The state space of the dynamical systems is constructed by correct definition of blob hypothesis and the translation and angular velocities of the blob hypothesis. In literature [6], coupled HMM is proposed for dynamic and complex action recognition. More states are adopted in the coupled HMM to enhance the ability for complex action recognition. However, it is difficult to propose an algorithm to decompose a human body into human body parts with motion, color, and spatial support region, etc. Moreovers, Yamato et al. [7] proposed a HMM-based behavior recognition model without geometric representation of the human body. The authors adopted mesh feature instead. Although the mesh features could be successfully applied to complex 2D patterns, some significant features like skin color and face characters would be ignored in the recognition process.

In this paper, the STSVM to identify significant human activities is proposed. The STSVM consists of several states, each of which is constructed by a single support vector machine (SVM), called State SVM (SSVM). Morphological (intra-structure) information of each body silhouette can be fed into the SSVM to evaluate its likelihood belonging to this state. Motion (inter-structure) information of a temporal silhouette sequence can be exploited to model the correlations among states by the state transition probability model (STPM). Hence, the presented STSVM has not only the property of static shape-based classification, but also dynamic correlation-based classification. Another main advantage of the STSVM is the ability to deal with the time-varying characteristics of a human behavior with the property of state transition.

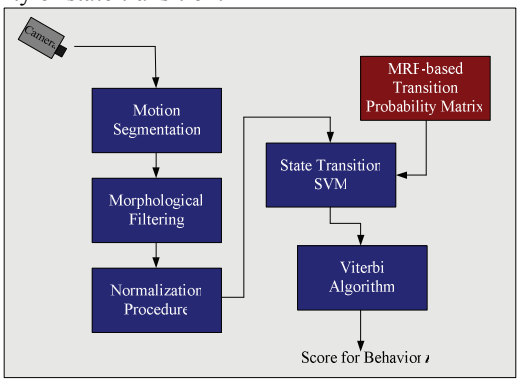

Fig. 1. Block diagram of the proposed framework

The block diagram of the proposed framework is shown in Fig. 1. The process begins with the camera image processing module (CIP Module) with motion segmentation, morphological filtering, and normalization procedure techniques. A person's silhouette sequence is extracted by the CIP module as inputs for the human behavior identification module (HBI Module), including MRF-based transition probability matrix, STSVM, and optimal path tracking procedure. The input human behavior is then determined using competitive model among different HBI modules.

2. METHODOLOGY

Tracking and analyzing human motion changes is a key technique for identification of behavior, however, it is very challenging to match an unknown silhouette sequence with a series of labeled reference sequences representing significant behaviors. For the purpose, STSVM consisting of SSVMs and the STPM is proposed in this study. The STSVM assumes the person's silhouette sequence of a human behavior is composed of several successively distinct states. We model each state by an individual 2-class SVM (called SSVM) trained by behavior state class and the competitive class. The competitive class is built by collecting a large number of various video events. The state transition of the STSVM is illustrated in Fig. 2.

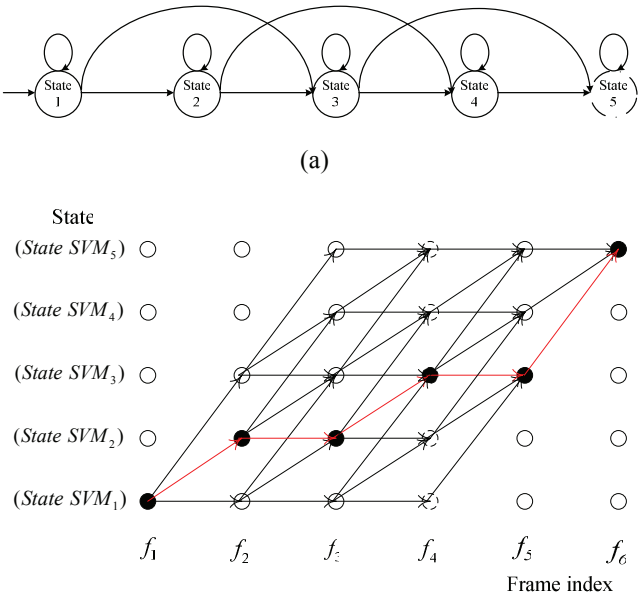

Fig. 2. Illustration of the proposed framework: (a) the state transition model for human behavior identification; (b) a state transition trellis generated from a video sequence.

As different human behaviors often sharing similar parts, the discriminability among them will decrease by using a fully connected state transition model. According to our experimental results, the fully connected model actually improves the recall rate, but decreases the precision rate. Considering the trade-off between flexibility and discriminability, eventually, the left-to-right model is adopted in this study. Figures 2(a) and 2(b) show that a state transition of our model possibly happens from state i to state j, while $j = i$, i+1, and i+2. Furthermore, we let the model begin and terminate at the first and the last state, respectively.

Based on the intra-structure information of an input person's silhouette, all the SSVMs generate the corresponding state probabilities. With the state probabilities and the MRF-based state transition probabilities, an optimal path with maximum accumulated probability can be decided. We construct the STSVM for each human behavior and choose the human behavior with highest accumulated probability as the identification result.

In the following, we address on how to compute the state probabilities and the state transition probabilities, respectively.

A. Generation of the State Probabilities

Assume the state number is N_{state} and the input person's silhouette number is T. Denote $\boldsymbol{I} = \{ f_t \}$, $t = 1, 2, \ldots, T$, as the feature sequence, i.e. the input person's silhouette sequence and represent $\boldsymbol{Q} = \{ q_t \}$, $t = 1, 2, \ldots, T$, as a state sequence for feature sequence $\boldsymbol{I}$. The state probability $\mathbf{B} = \{\Pr(\mathbf{f}_t \mid q_i)\}$, $i = 1, 2, \ldots, N_{State}$, $t = 1, 2, \ldots, T$, is generated by transforming conventional SVM output into a probabilistic score [8]. The following describes how to compute the probabilistic scores. Denote the human behavior class and the competitive class in i-th state as $C_{i,m}$, where $m = +1$ and -1, respectively. For an input person's silhouette ft classified into $C_{i,m}$, $m \in \{-1,+1\}$, the distance ratio of the distance between f_t and optimal hyperplane to the margin distance is defined by

$$R(\mathbf{f}_t) = \frac{\mathbf{w}\mathbf{f}_t + b}{\|\mathbf{w}\|} \Big/ \frac{1}{\|\mathbf{w}\|} = \mathbf{w}\mathbf{f}_t + b \cdot \tag{1}$$

The probabilistic score is then obtained by converting the distance ratio to a value between 0 and +1 through a sigmoid function

$$\Pr(\mathbf{f}_t \mid q_i) \equiv Score(\mathbf{f}_t \mid C_{i,m=+1}) = \frac{1}{1+e^{-R(\mathbf{f}_t)}} . \tag{2}$$

B. Generation of the State Transition Probabilities

Denote $\boldsymbol{A} = \{ a_{ij} \}$ as the state transition probability from state q_i to state q_j. Recall that the method of Hidden Markov Model (HMM) uses only occurrence counting of person's contours in calculating transition probability. This strategy is not always the most effective. For examples, there exist p person's contours in an unknown human action. We make

assumption the p person's contours to belong one of three categories labeled $p_{i,j}$, where i = 1,2, …,p and j = 1,2,3. In general, the transition probability aii would be much greater than a_{ij} ($i \neq j$). If more effective methods are used, the resulting transition probability calculation may perform better than the occurrence counting. In this investigation, we adopt the Markov random field (MRF) [9-10] to estimate the relationship (each a_{ij} in $\boldsymbol{A}$). The idea behind this method is as follows. Given a set of condition probability distributions which describe the probability distribution of a random variable (representing a person's contour) while the probability distributions of neighbor random variables (representing each two adjacent person's contours) are given, we then can represent the joint probability distribution (also called transition probability a_{ij}) of each pair of random variables. For instance, in Fig. 3, the MRF-based transition probability a_{23} (between the second and the third person's contours) is greater than a_{25}. It is also more reasonable in the real world if the second person's contour is followed by the third person's contour than followed by the fifth person's contour for a hand raising event.

Fig. 3. An example of a person's silhouette sequence for raising hand behavior.

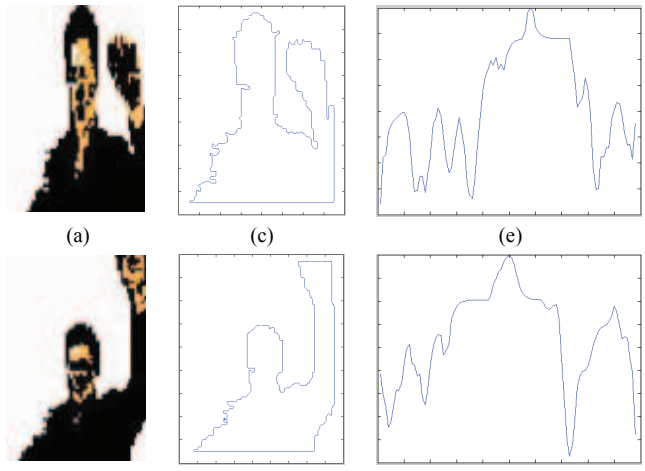

Fig. 4. Illustration of two localized contour sequences generated by the LCS approach. (a),(b): original images; (c),(d): the contour information of (a) and (b); (e),(f): the localized contour sequences of (c) and (d).

We model each transition probability aij with MRF-based inter-structure information of consecutive silhouettes. The state transition probability a_{ij} can be represented in formula (3) for all pairs of two successive f_t and f_{t+1} which belongs to state i and state j as

$$a_{ij} = \frac{1}{Z} \prod_{f_t \in q_i, f_{t+1} \in q_j} \varphi(x_{f_t, f_{t+1}}), \tag{3}$$

and the potential function $\varphi(x_{f_t, f_{t+1}})$ is defined as below

$$\varphi(x_{f_t, f_{t+1}}) = e^{-(D_m(f_t, f_{t+1}))}, \tag{4}$$

where $D_m(f_t, f_{t+1})$ denotes the difference between two consecutive person's contours represented by the localized contour sequence (LCS) [11]. This distance is calculated by

$$D_m(f_t, f_{t+1}) = \sum_{k=1}^{L} |lcs_t(k) - lcs_{t+1}((k+m))|, m = 0,1,...,(L-1), \tag{5}$$

where lcs_t denotes the i-th input silhouette of a test behavior, lcs_{t+1} represents a circular shift of m samples in lcs_{t+1}. The best match between lcst and lcs_{t+1} is given by

$$D^* = \min_m [D_m(f_t, f_{t+1})]. \tag{6}$$

Figure 4 presents an illustration of the adopted LCS approach to extract the contour information.

Furthermore, the sum of state transition probabilities from a fixed state to other states will be normalized to one (i.e. $a_{ij} = \frac{a_{ij}^{MRF}}{\sum_j a_{ij}^{MRF}}$, and then $\sum_j a_{ij} = 1$). With the state transition probability aij calculated by MRF theory, the optimal state transition paths of all training samples can be found. In the identification phase, the maximum accumulated probability along the optimal state transition path is found for each STSVM. Represent $\boldsymbol{\pi} = \{\pi_i\}$, where $\pi_i = \Pr(q_i \mid t = 1)$ as the initial state probabilities. For an input sequence $\boldsymbol{I}$ and an STSVM with $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{\pi}$ parameters, the normalized accumulated probability along the optimal path is defined as

$$-\frac{1}{T}\log(\Pr(\mathbf{I} \mid \mathbf{A}, \mathbf{B}, \boldsymbol{\pi})) = \underset{every\ possible\ q}{Max} \{-\frac{1}{T}\log(\Pr(\mathbf{I}, q \mid \mathbf{A}, \mathbf{B}, \boldsymbol{\pi})\}, \tag{7}$$

where the joint probability $\Pr(\boldsymbol{I}, q|\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{\pi})$ of an observation sequence and a path q is defined as

$$\Pr(\mathbf{I}, q \mid \mathbf{A}, \mathbf{B}, \boldsymbol{\pi}) = \Pr(\mathbf{I} \mid q, \mathbf{A}, \mathbf{B}, \boldsymbol{\pi}) \cdot \Pr(q \mid \mathbf{A}, \mathbf{B}, \boldsymbol{\pi}). \tag{8}$$

We use the Viterbi algorithm [12] to find out the optimal path. The unknown behavior will then be identified into a pre-defined behavior category with the highest accumulated log-likelihood, which should also be greater than a predefined threshold.

In this phase, several sets of STSVM models having different parameter sets are prepared. Using these models, the normalized accumulated probability scores are calculated for all sets, and a set of models having maximum score is selected. The testing sequence will be regarded as non event (still unknown) if the obtained score is less than the predefined threshold, which is defined based on different sets of models.

3. EXPERIMENTAL RESULTS

We collected 300 sequences from five different people, 60 of each activity: rising hand, standing up, squatting down, falling down, and sitting down. Each activity begins and ends with the corresponding poses. With the incrementally increasing behavior number, more states are necessary in STSVM. The more states we use in the STSVM model, the higher performance the recognition system provides in our investigation. However, the question now arises: the overhead of computation time with increasing state number. What has to be noticed is the trade-off between performance and complexity of the developed system. We developed the STSVM-based recognition system with three states, because of two reasons: providing enough discriminability for all behaviors that we consider and possibility for real-time computing.

The state transition support vector machines aims to identify significant human behaviors. We adopted the precision measure to assess its effectiveness again. Table I lists the precision values obtained by the single SVM, the multi-stage SVM, GMM-based HMM and the STSVM.

Table I. Evaluation of Human Behavior Identification on Using Single SVM, Multi-state SVM, GMM-based HMM and STSVM.

Human Behavior	Precision (%)			
	Single SVM	Multi-Stage SVM	GMM-based HMM	State Transition SVM
Raising hand	80	81	93	***95***
Standing up	74	77	81	***88***
Squatting down	68	71	72	***84***
Falling down	76	79	85	***85***
Sitting down	68	68	78	***91***
Average	***73.2***	***75.2***	***81.8***	***88.6***

According to the displayed experimental results, the performance of the proposed STSVM apparently outperforms the other two approaches. The single SVM classifier does not present its superior performance due to its weakness to deal with time-varying characteristics of an event. The multi-stage SVM is limited by how to associate input frames to the related stages. Besides SVM-based methods, HMM is one of the most often used algorithms for human behavior identification. In the last analysis, the main limitations of GMM-based HMM are the discriminability of GMM and the calculation of state transition probability for people poses.

4. CONCLUSIONS

The goal of this study presents an effective human behavior identification system for intelligent surveillance and ubiquitous computing. The STSVM is the core technique proposed in this study to develop the complete system, which allows users to define meaningful human behaviors. Each separated state contains one State SVM to determine the probability of a testing image belonging to. Moreover, the transition probability between states is modeled by MRF theory with the localized contour sequence (LCS) approach. The main advantages of the proposed method are shown as follows:

(a) The superiority in human pose recognition using State SVM instead of GMM

(b) The power functionality to deal with unknown length and non-deterministic content of event sequences through transition model

Therefore, STSVM can accomplish the purpose of human behavior identification well.

ACKNOWLEDGEMENTS

This study is conducted under the "Applied Information Services Development & Integration project" of the Institute for Information Industry which is subsidized by the Ministry of Economy Affairs of the Republic of China.

REFERENCES

[1] B. Stenger, P.H.S. Torr, and R. Cipolla, "Model-based hand tracking using a hierarchical Bayesian filter," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 28, no. 9, 2006.

[2] S. X. Ju, M. J. Black, and Y. Yacoob, "Cardboard people: A parameterized model of articulated motion," in *Proc. IEEE Conf. Automatic Face and Gesture Recognition*, 1996, pp. 38–44.

[3] H. Sidenbladh, M. Black, and D. Fleet, "Stochastic tracking of 3d human figures using 2d image motion," in *European Conf. Computer Vision*, 2000, pp. 702–718.

[4] K. Grauman, G. Shakhnarovich, and T. Darrell, "Inferring 3d structure with a statistical image-based shape model," in *Proc. Int. Conf. Computer Vision*, 2003, pp. 641-647.

[5] C. Bregler, "Learning and Recognizing Human Dynamics in Video Sequences", *In the Proc. of IEEE Computer Vision and Pattern Recognition*, 1997, pp. 568-573.

[6] M. Brand, N. Oliver, and A. Pentland, "Coupled hidden Markov models for complex action recognition," *In Proc. IEEE, Computer Vision and Pattern Recognition*, 1997, pp. 994-998.

[7] J. Yamato, J. Ohya, K. Ishii, "Recognizing Human Action in Time-Sequential Images Using Hidden Markov Model", *In the Proc. of IEEE Computer Vision and Pattern Recognition*, 1992, pp. 379-385.

[8] A. Ganapathiraju, J.E. Hamaker, and J. Picone, "Applications of support vector machines to speech recognition," *IEEE Transactions on Signal Processing*, vol. 52, no. 8, pp. 2348–2355. Aug. 2004.

[9] M. I. Jordon, editor, *Learning in graphical models*, MIT Press, 1999.

[10] C. M. Bishop, *Pattern recognition and machine learning*, Springer, 2006.

[11] L. Gupta and S. Ma, "Gesture-based interaction and communication: automated classification of hand gesture contours," *IEEE Trans. System, Man, and Cybernetics-Part C: Application and reviews*, vol. 31, no. 31, pp. 114–120, Feb. 2001.

[12] G. D. Forney, "The Viterbi algorithm," *Proceedings of the IEEE*, vol. 61, no. 3, pp. 268–-278, Mar. 1973.