Recognition of Human Actions Using Moment Based Features and Artificial Neural Networks

Arun Sharma, Dinesh K Kumar, Sanjay Kumar and Neil McLachlan
School of ECE, RMIT University, GPO Box 2476V Melbourne, Victoria, Australia 3001
e-mail:s9900960@student.rmit.edu.au

Abstract

This paper presents performance of view-based approach in automated recognition of pre-defined hand and gross body actions using artificial neural network. This approach represents motion by a static grey scale image template computed by collapsing the temporal components into the cumulative image-difference of frames. The seven invariant Hu moments are used as the feature vectors the performance of the system is tested, in real time, using feed forward multilayer perceptron (MLP) based on back propagation.

1. Introduction

Classification of human gestures is a very challenging problem. This research proposes to recognize gestures, without using intrusive devices, in real time so that the benefits can be extended to normal living and working environments thus enabling strong applications in computer surveillance and control, helping disabled people etc. This technique is computationally inexpensive as it reduces the dimensionality of the data space representing motion without loss of motion features. In these experiments motion is captured from a single view. The method can be extended to multiple views to take care of occlusion.

2. Explanation of the approach and Method

Actions can be recognized by the description of the appearance of motion [1, 2]. The motion is represented by the spatio-temporal template $H_n(x,y.$ For an intensity an image sequence I (x, y, n), difference of consecutive frames representing regions of motion D (x, y, n), and the motion template $H_n(x,y)$, for N frames is given by

$$H_n(x,y) = \max \left\{ \sum_{n=1}^{N-1} B(x,y,n) \times n \right.$$

where, $B(x,y,n) = \begin{cases} 1 & \text{if } D(x,y,n) > \Gamma \\ 0 & \text{Otherwise} \end{cases}$, and

$$D(x,y,n) = \left| I(x,y,n) - I(x,y,n-1) \right| .$$

The feature vectors representing Hn*(x,y)* are seven Hu functions [3] based on normalized centralized image moments which are invariant to scale, translation and rotation. This feature vector is used to train the neural network and classify the actions. For each set of actions a different two–layer feed forward preceptron (MLP) system with two hidden layers has been adopted. Motion was captured with the camera axis perpendicular to the plane of action. Six subjects performed ten times, each of the four pre-defined actions from each set of gross body and hand actions. Fig. 1 shows an example from each set.

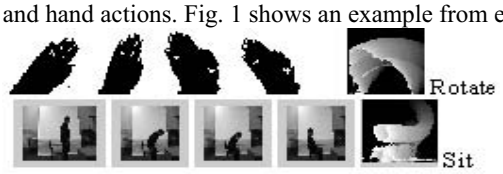

Fig.1 Hand and Gross body actions with motion template

3. Results and Discussions

The recognition accuracies for gross body and for hand actions are 87.5% and 97.5% respectively. The higher rate of recognition for hand actions can be attributed to the image differencing technique being very sensitive to secondary motion of the body part (eg. loose clothes), which may not essentially be a part of the action. Among the hand actions there is no unwanted movement to generate noise. The results show that the method looks at the appearance of motion and is independent of the person performing the action i.e. invariant to factors such as different height, weight and gender of subjects

4. References

[1] J. W. Davis, "Representing and recognizing Human Motion: From Motion templates to Movement Categories," *Intern. Conf on Intelligent Robots and Sys., Maui, Hawaii*, October 29 2001.
[2] A. Sharma, D. K. Kumar, S. Kumar , and N. McLachlan, "Classification of Human Actions using Temporal Templates, Histograms And Orthogonal Moments," presented at 7th World Multiconference (SCI 2003), Orlando, USA, 2003, July 27-30,.
[3] M.-K. Hu, "Visual Pattern Recognition by Moment Invariants," *IEEE Transaction on Information Theory*, vol. 8, pp. 179-187, 1962.

Proceedings of the 10th International Multimedia Modelling Conference (MMM'04)
0-7695-2084-7/04 $ 20.00 © 2004 IEEE