

Visual Hand Gestures Classification Using Temporal Motion Templates and Wavelet Transforms

Sanjay Kumar, Dinesh Kant Kumar, Arun Sharma, Neil McLachlan
 School of Electrical and Computer Engineering
 RMIT University, GPO Box 2476V Melbourne, Victoria, Australia 3001
 Phone:00-61-3-99253025 Fax-00-61-3-99252007
 E-mail: s2003383@student.rmit.edu.au

Abstract

This paper presents a novel technique for classifying human hand gestures based on stationary wavelet transform (SWT). This approach uses a cumulative image-difference technique where the time between the sequences of images is implicitly captured in the representation of action. This results in the construction of Temporal History Templates (THT). These THT's are decomposed into 4 sub images using SWT, a average image (f_{ii}), and three detail images (f_{ih} , f_{hb} , f_{hh}) respectively. The average image (f_{ii}) is fed as the global image descriptors to the ANN for classification. The preliminary experiments show that such a system can classify human hand gestures with a classification accuracy of 97%.

1.Introduction

The work reported in this paper is view-based approach for the representation and classification of pre-defined hand gestures using characteristics of the fine image motion of hand-gestures from particular view direction. The technique is based on the use of Temporal History Templates (THT)[1], which is a scalar gray scale intensity image. This paper reports the research based on multiresolution analysis and decomposition of THT by using two-dimensional SWT of THT, which results in the four sub images of THT, average (f_{ii}) and three detail images (f_{ih} , f_{hb} , f_{hh}). The (f_{ii}) images are then fed to the two layers in the multiplayer perceptron neural network for classification.

2.Theory

Let $V(x, y, n)$ be an image sequence & let $DIFF(x, y, n) = |V(x, y, n) - V(x, y, n-1)|$
 Where $V(x, y, n)$ is the intensity of each pixel at location (x, y) in the n th frame and $DIFF(x, y, n)$ is the difference of consecutive frames representing regions of motion.

Binarisation of the difference image $DIFF(x, y, n)$ over a threshold τ , is $DOF(x, y, n)$

$$DOF(x, y, n) = \begin{cases} 1 & \text{if } DIFF(x, y, n) > \tau \\ 0 & \text{otherwise} \end{cases}$$

Then THT ($T_N(x, y)$) is:

$$THT(T_N(x, y)) = \text{Max}_{n=1}^{N-1} \{ DOF(x, y, n) * n \}$$

Where N represents the duration of the time window used to capture the motion.

2.Method

The video data was recorded for the predefined hand gestures and the THT was generated and then SWT of the THT was performed to decompose THT's of hand representation. in four sub images of THT, namely as average image (f_{ii}) "Figure 1", and three detail images (f_{ih} , f_{hb} , f_{hh}). For the purpose of classification the average sub image (f_{ii}) "Figure 1", each action computed from THT representation was the input to the ANN for classification.

3.Result, Discussions and Conclusion

Tests results yielded that neural network can classify hand gestures with an accuracy of 97 %. This method is more faster and accurate and less computational expensive.




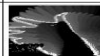

Action Class	Start Frame	Middle Frame	End Frame	Template	Wavelet Template at L-1
LEFT					

Figure 1

4.References

[1] Aaron F. Bobick, J.W.D., "The Recognition Of Human Movements Using Temporal Templates", IEEE - Pattern Analysis and Machine Intelligence, 23 No 3, pp. 257-267, 2001