

Visual Hand Gestures Classification Using Temporal Motion Templates

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 method for hand gesture classification using a view-based approach for representation and Artificial Neural Network for classification. This approach uses a cumulative image-difference technique in which 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 images are used to compute the 7 Hu image moments, which are invariant to scale, rotation and translation. The classification is then performed using back propagation based multilayer perceptron (MLP) Artificial Neural Network (ANN). The preliminary experiments show that such a system can classify human hand gestures with a classification accuracy of 96%. Motivation of the work is to build a system for person identification based on this technique.

1.Introduction

As the technology is advancing the Human Computer Interaction is becoming an increasingly important part of our daily lives. The present work is view-based approach for the representation and classification of pre-defined 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]. This research has combined the use of THT with the image moment technique proposed by Hu [2] for use with finer hand movements.

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 pre defined hand gestures and the THT was generated. From the THT representation of each action 7-Hu moments were computed [2]. The hu-moments from each THT were fed to two layer of MLP neural network for classification.

3.Result, Discussions and Conclusion

The results of the testing show with this described method can classify the five gesture classes with 96% accuracy and is more accurate and less computational expensive.





Action Class	Start Frame	Middle Frame	End Frame	Template
HOLD				

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.
- [2] Hu, "Visual Pattern Recognition By Moment Invariants", IEEE - Pattern Transaction On Information Theory, 8(2), pp.179-187, 1962.