ISSN: 2321-7782 (Online)
Volume 2, Issue 4, April 2014
International Journal of Advance Research in Computer Science and Management Studies
Research Article / Paper / Case Study
Available online at: www.ijarcsms.com

Performance Evolution of Eye and Hand Fusion for Diagonal Movement Gesture Recognition

Yogesh C. Pathak[1]
Mtech 2 Year
G.H.R.C.E., Nagpur
India

S. A. Chhabria[2]
Hod Of It Dept.
G.H.R.C.E. Nagpur
India

Abstract: ***Human–computer Interaction (HCI) involves the study, planning, and design of the interaction between people (users) and computers. It is often regarded as the intersection of computer science, behavioral, design and several other fields of study. HCI plays a very important role in interaction with the computers. This will be the new era in computation.***

The project will compare different fusion techniques and use optimized fusion technique for fusing eye and hand gestures. It will be working on the movements towards diagonally right upwards, diagonally left upwards, diagonally right downwards, and diagonally left downwards.

The mechanism of the project is to make performance evolution of the fusion techniques; these fusion techniques will evaluate the ongoing performance using the testing phases and will show the results of fusion techniques for eye and hand gestures.

I. Introduction

Human–computer Interaction (HCI) involves the study, planning, and design of the interaction between people (users) and computers. It is often regarded as the intersection of computer science, behavioural, design and several other fields of study. HCI plays a very important role in interaction with the computers. This will be the new era in computation.

The project will compare different fusion techniques and use optimized fusion technique for fusing eye and hand gestures. It will be working on the movements towards diagonally right upwards, diagonally left upwards, diagonally right downwards, and diagonally left downwards.

The mechanism of the project is to make performance evolution of the fusion techniques; these fusion techniques will evaluate the ongoing performance using the testing phases and will show the results of fusion techniques for eye and hand gestures.

II. Problem Definition

The problem definition is to evaluate the performance of the fusion technique for eye and hand diagonal movements and to see which technique is the best suited related with the speed, accuracy, performance etc.

The mechanism of the project is to make performance evolution of the fusion techniques; these fusion techniques will evaluate the ongoing performance using the testing phases and will show the results of fusion techniques for eye and hand gestures.

III. LITERATURE SURVEY

For fusion:

Rule-based fusion methods

The rule-based fusion method includes a variety of basic rules of combining multimodal information. These include statistical rule-based methods such as linear weighted fusion (sum and product), MAX, MIN, AND, OR, majority voting. The work by Kittler et al. [3] has provided the theoretical introduction of these rules. In addition to these rules, there are custom-defined rules that are constructed for the specific application perspective. The rule-based schemes generally perform well if the quality of temporal alignment between different modalities is good. And these are more and better result proving technique.

Linear weighted fusion

Linear weighted fusion is one of the simplest and most widely used methods. In this method, the information obtained from different modalities is combined in a linear fashion. The information could be the low-level features (e.g. color and motion cues in video frames) [6] or the semantic-level decisions [4] (i.e. occurrence of an event) .To combine the information, one may assign normalized weights to different modalities. In literature, there are various methods for weight normalization such as min–max, decimal scaling, z score, tanh-estimators and sigmoid function [5]. Each of these methods has pros and cons. The min–max, decimal scaling and z score methods are preferred when the matching scores (minimum and maximum values for min–max, maximum for decimal scaling and mean and standard deviation for z score) of the individual modalities can be easily computed. But these methods are sensitive to outliers. On the other hand, tan h normalization method is both robust and efficient but requires estimation of the parameters using training. Note that the absence of prior knowledge of the weights usually equals the weight assigned to them. The general methodology of linear fusion can be described as follows. Let Ii, 1 B i B n be a feature vector obtained from ith media source (e.g. audio, video etc.) or a decision obtained from a classifier.1 Also, let wi, 1 B i B n be the normalized weight assigned to the ith media source or classifier. These vectors, assuming that they have the same dimensions, are combined by using sum or product operators and used by the classifiers to provide a high-level decision [1].

Majority voting

Majority voting is a special case of weighted combination with all weights to be equal. In majority voting based fusion, the final decision is the one where the majority of the classifiers reach a similar decision [7] . For example, Radova and Psutka [8] have presented a speaker identification system by employing multiple classifiers. Here, the raw speech samples from the speaker are treated as features. From the speech samples, a set of patterns are identified for each speaker. The pattern usually contains a current utterance of several vowels. Each pattern is classified by two different classifiers. The output scores of all the classifiers were fused in a late integration approach to obtain the majority decision regarding the identity of the unknown speaker [1].

Support vector machine

Support vector machine (SVM) [9] has become increasingly popular for data classification and related tasks. More specifically, in the domain of multimedia, SVMs are being used for different tasks including feature categorization, concept classification, face detection, text categorization, modality fusion, etc. Basically SVM is considered as a supervised learning method and is used as an optimal binary linear classifier, where a set of input data vectors are partitioned as belonging to either one of the two learned classes. From the perspective of multimodal fusion, SVM is used to solve a pattern classification problem, where the input to this classifier is the scores given by the individual classifier. The basic SVM method is extended to create a non-linear classifier by using the kernel concept, where every dot product in the basic SVM formalism is replaced using a non-linear kernel functions. Many existing literature use the SVM-based fusion scheme. Adams et al. [10] adopted a late fusion approach in order to detect semantic concepts (e.g. sky, fire-smoke) in videos using visual, audio and textual modalities [1].

Bayesian inference

The Bayesian inference is often referred to as the 'classical' sensor fusion method because it has been widely used and many other methods are based on it [11]. In this method, the multimodal information is combined as per the rules of probability theory [12]. The method can be applied at the feature level as well as at the decision level. The observations obtained from multiple modalities or the decisions obtained from different classifiers are combined, and an inference of the joint probability of an observation or a decision is derived [13].

Dempster–Shafer theory

Although the Bayesian inference fusion method allows for uncertainty modelling (usually by Gaussian distribution), some researchers have preferred to use the Dempster– Shafer (D–S) evidence theory since it uses belief and plausibility values to represent the evidence and their corresponding uncertainty[14]. Moreover, the D–S method generalizes the Bayesian theory to relax the Bayesian inference method's restriction on mutually exclusive hypotheses, so that it is able to assign evidence to the union of hypotheses [15].

Neural networks

Neural network (NN) is another approach for fusing multimodal data. Neural networks are considered a non-linear black box that can be trained to solve ill-defined and computationally expensive problems [15] .The NN method consists of a network of mainly three types of nodes—input, hidden and output nodes. The input nodes accept sensor observations or decisions (based on these observations), and the output nodes provide the results of fusion of the observations or decisions. The nodes that are neither input nor output are referred to hidden nodes. The network architecture design between the input and output nodes is an important factor for the success or failure of this method. The weights along the paths, that connect the input nodes to the output nodes, decide the input–output mapping behaviour. These weights can be adjusted during the training phase to obtain the optimal fusion results [16].This method can also be employed at both the feature level and the decision level [1].

Maximum entropy model

In general, maximum entropy model is a statistical classifier which follows an information-theoretic approach and provides a probability of an observation belonging to a particular class based on the information content it has. This method has been used by few researchers for categorizing the fused multimedia observations into respective classes [1]

Features

The main purpose of using features instead of raw pixel values as the input to a learning algorithm is to reduce/increase the in-class/out of class variability compared to the raw input data, and thus making classification easier. Features usually encode knowledge about the domain, which is difficult to learn from a raw and finite set of input data.

Cascade of Classifiers:

A cascade of classifiers is a degenerated decision tree where at each stage a classifier is trained to detect almost all objects of interest (frontal faces in our example) while rejecting a certain fraction of the non-object patterns Each stage was trained using one out of the three Boosting variants. Boosting can learn a strong classifier based on a (large) set of weak classifiers by re-weighting the training samples. Weak classifiers are only required to be slightly better than chance. Our set of weak classifiers is all classifiers which use one feature from our feature pool in combination with a simple binary thresholding decision. At each round of boosting, the feature-based classifier is added that best classifies the weighted training samples. With increasing stage number the number of weak classifiers, which are needed to achieve the desired false alarm rate at the given hit rate, increases. [2]

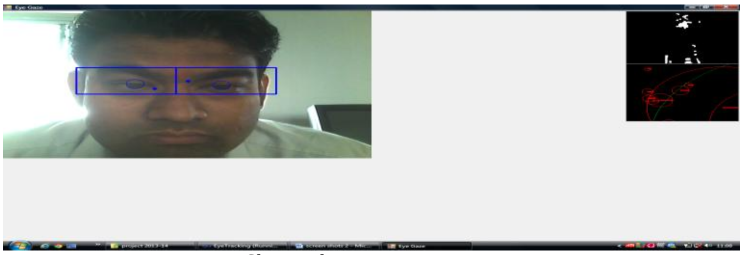

Fig 1: Shows The Eye Region Detection

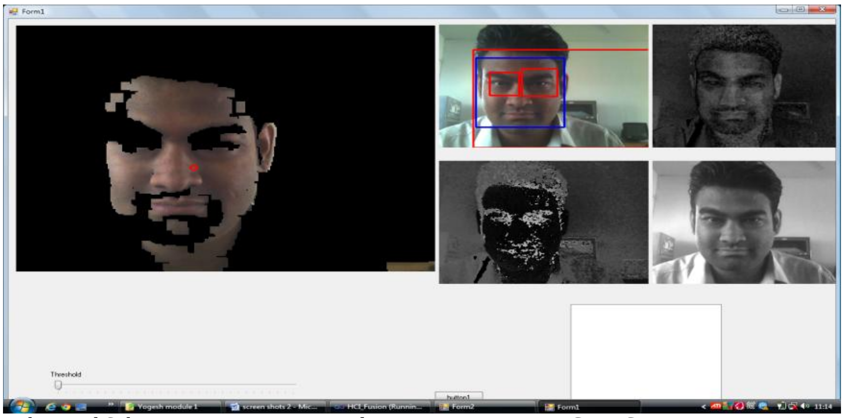

Fig 2: Different Frames Where Background Substraction, Face Region and Eye Region Detection, Grey, Contrast, Negative Images Are Shown.

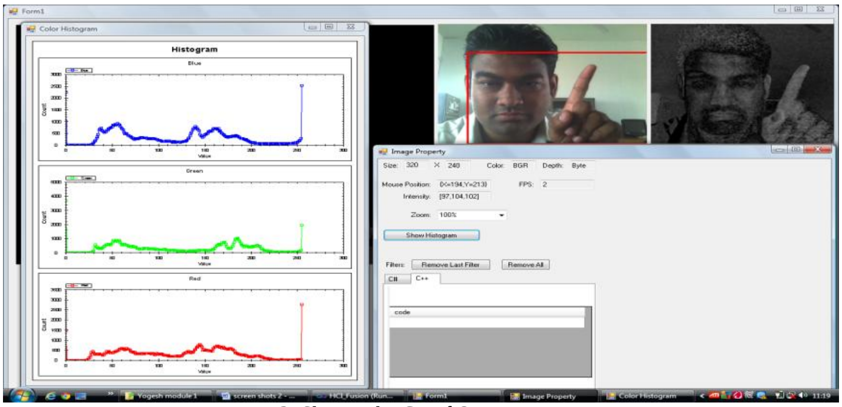

Fig 3: Shows The Grapf Or Histogram

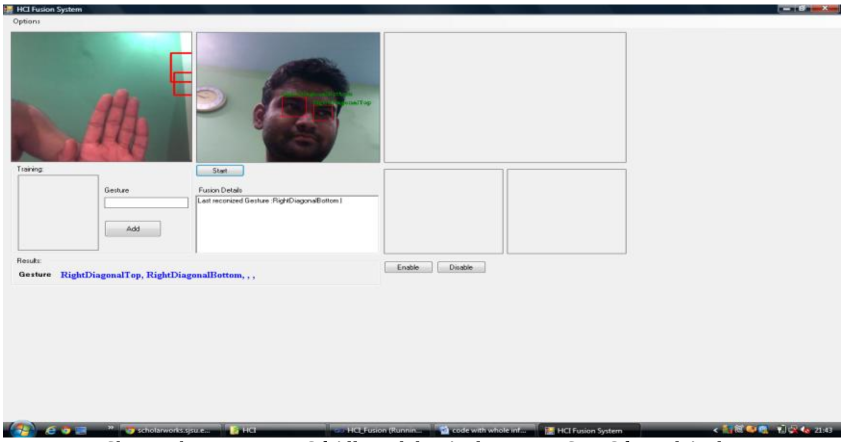

Fig 4: Shows The Integration Of All Modules And Training Sets Of Hand And Eyes.

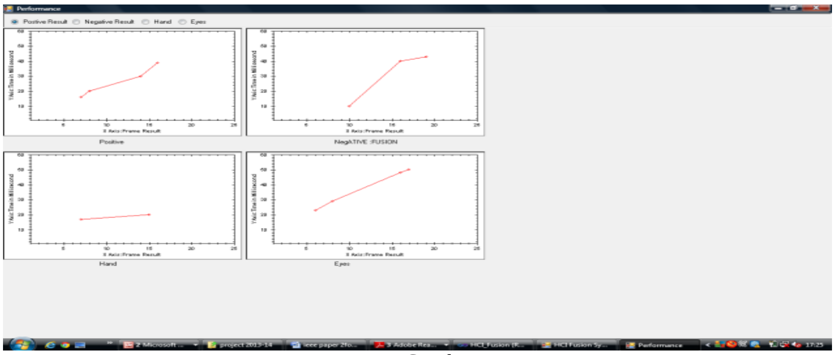

Fig 5: Graphs

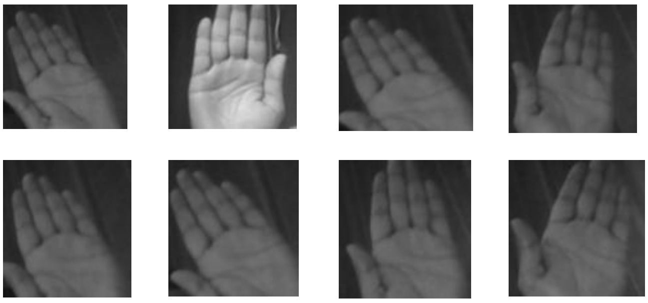

Fig 6: Shows Training Data Set For Hand.

Fig 7: Shows Training Data Set For Eyes.

IV. Application

It could be used by paralyzed persons i.e. to move the wheel chair, or to operate the specific devices or even full computer operations.

Could be further used by the patients or handicap persons to play certain games.

V. Future Scope

By developing this project it could be in future be implemented to help a disabled person or any handy cap person to move his or her wheel chair or even could be used to play certain games . This project will provide a different touch to the world of human computer interaction.

References

1. Pradeep K. Atrey , M. Anwar Hossain , Abdulmotaleb El Saddik , Mohan S. Kankanhalli: Multimodal fusion for multimedia analysis: a survey, Published online: 4 April 2010 Springer-Verlag 2010, Multimedia Systems (2010) 16:345–379 DOI 10.1007/s00530-010-0182-0
2. Rainer Lienhart, Alexander Kuranov, Vadim Pisarevsky Empirical Analysis of Detection Cascades of Boosted Classifiers for Rapid Object Detection Microprocessor Research Lab, Intel Labs Intel Corporation, Santa Clara, CA 95052, USA, MRL Technical Report, May 2002, revised December 2002.
3. Kittler, J., Hatef, M., Duin, R.P., Matas, J.: On combining classifiers. IEEE Trans. Pattern Anal. Mach. Intell. 20(3), 226– 239 (1998)
4. Nock, H.J., Iyengar, G., Neti, C.: Assessing face and speech consistency for monologue detection in video. In: ACM International Conference on Multimedia. French Riviera, France (2002)
5. Jain, A., Nandakumar, K., Ross, A.: Score normalization in multimodal biometric systems. Pattern Recognit. 38(12), 2270– 2285 (2005)
6. Wang, J., Kankanhalli, M.S., Yan, W.Q., Jain, R.: Experiential sampling for video surveillance. In: ACM Workshop on Video Surveillance. Berkeley (2003)
7. Sanderson, C., Paliwal, K.K.: Identity verification using speech and face information. Digit. Signal Process. 14(5), 449–480 (2004)
8. Radova, V., Psutka, J.: An approach to speaker identification using multiple classifiers. In: IEEE International Conference on Acoustics, Speech, and Signal Processing, 2, 1135–1138. Munich, Germany (1997)
9. Burges, C.J.C.: A tutorial on support vector machines for pattern recognition. Data Min. Knowl. Discov. 2(2), 121–167 (1998)
10. Adams, W., Iyengar, G., Lin, C., Naphade, M., Neti, C., Nock, H., Smith, J.: Semantic indexing of multimedia content using visual, audio, and text cues. EURASIP J. Appl. Signal Process.2003(2), 170–185 (2003).
11. Hall, D.L., Llinas, J.: An introduction to multisensor fusion. In: Proceedings of the IEEE: Special Issues on Data Fusion, vol. 85, no. 1, pp. 6–23 (1997).
12. Luo, R.C., Yih, C.C., Su, K.L.: Multisensor fusion and integration: Approaches, applications, and future research directions. IEEE Sens. J. 2(2), 107–119 (2002).
13. Rashidi, A., Ghassemian, H.: Extended dempster–shafer theory for multi-system/sensor decision fusion. In: Commission IV Joint Workshop on Challenges in Geospatial Analysis.
14. Reddy, B.S.: Evidential reasoning for multimodal fusion in human computer interaction (2007). MS Thesis, University of Waterloo, Canada
15. Wu, H.: Sensor data fusion for context-aware computing using dempster–shafer theory. Ph.D. thesis, The Robotics Institute, Carnegie Mellon University, Pittsburgh, PA (2003)
16. Brooks, R.R., Iyengar, S.S.: Multi-sensor Fusion: Fundamentals and Applications with Software. Prentice Hall PTR, Upper Saddle River, NJ (1998)