IIOAB JOURNAL

ARTICLE

VISION BASED GESTURE RECOGNITION: A COMPREHENSIVE STUDY

A Balasundaram[1*], C Chellappan[2]

[1]Research Scholar, Department of CSE, G.K.M. College of Engineering and Technology, Chennai, INDIA
[2]Principal, G.K.M. College of Engineering and Technology, Chennai, INDIA

ISSN 0976-3104

ABSTRACT

Computer vision involves identifying the behavior, interpreting background scene and understanding the scenario much similar to the manner in which human brain thinks and reacts. This intelligence in tuning the computer to understand objects under study and their behavior is highly essential across various domains such as security, health care, medical science, educational systems, etc., . As part of this work, a comprehensive study of various works done towards hand and face based gesture recognition has been listed and the concepts and technical aspects used behind have been compared. Also, an architecture for designing a vision based intelligence system which has the intelligence to understand the situation or scene and react to the same is proposed.

KEY WORDS
Gesture Recognition; Hand Gesture; Face Gestures; Computer Vision; GCAIS

INTRODUCTION

An intelligence system essentially tracks the entire activities which include actions, behaviors, interactions and happenings. But, the potential of this system can be maximized by inducing some intelligence and training the system to be context aware and capable of making decisions and in turn perform a sequence of alerting mechanisms. This process of making the system intelligent is easier said than implementable. Multiple steps are involved in acquiring this intelligence which starts with preprocessing of images to capture the scene or scenario at different levels of abstraction, detecting the objects of interest and tracking the same in the given scene or sequence, training the system to understand multiple contexts and able to apply the same in real time and make educated decisions (predictions) and perform corresponding actions. Detection and tracking of objects involves studying the motion or movement of the object. The features of motion of object such as movement pattern, trajectory, inter-object distance etc., [1] are highly significant in the study of motion of object. In general, any intelligent system requires subsystems to analyze interactions, activities, context understanding and detecting objects as shown in [Fig.1].

Received: 6 June 2017
Accepted: 18 June 2017
Published: 15 July 2017

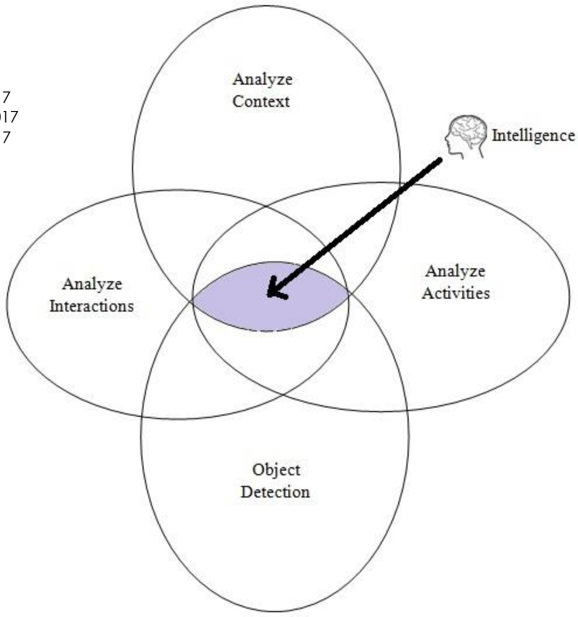

Fig.1: Subsystems to acquire intelligence.

***Corresponding Author**
Email: balasundaram2682@gmail.com
Tel.: +91 9841288826

It can be seen that complete intelligence is acquired when all the four subsystems go hand in hand. Developing an intelligent surveillance system is application specific and requires multiple templates that represent the object's behavior. These predefined templates are in other words termed as priori. However, not all characteristics of the object can be determined using the priori information and proper training data for the system plays a vital role in determining the accuracy of intelligence acquired.

RECENT WORK DONE TOWARDS VISION BASED GESTURE RECOGNITION

The spectrum of vision based applications include analysis of complex scientific data which is used across multiple domains such as medical science, military applications, surveillance, health care, educational developments etc,. [Table 1] shows some of the recent work performed towards vision based gesture recognition based on face and hand gestures.

Table 1: Recent work related to vision based gesture recognition

Reference number	Work done	Feature extractor	Classifier used	Database
[2]	Human motion classification using 2D stick-model matching regression coefficients	Polynomial Regression Data Fit	Bayes, Lazy, Function, Meta, Misc, Rules and Trees classifiers	Own Database
[3]	An Architecture for Personality-Based, Nonverbal Behavior in Affective Virtual Humanoid Character	3D Hybrid model which considers non-verbal behavior	Rule Based	Own Database
[4]	Gaze patterns during scene processing in typical adults and adults with autism spectrum disorders	Statistical methods	Rule Based	Medical Database
[5]	Walking behavior change detector for a "smart" walker	Gait Analysis	Binary ANN cascade classifiers	Own Database
[6]	Framework for Traffic Personnel Gesture Recognition	Cumulative Block Intensity Vector (CBIV) of n-frame cumulative difference	Support Vector Machine (SVM), Decision Tree and Random Forests (RF)	Own Database
[7]	Gait-Based Emotion Detection of Children with Autism Spectrum Disorders: A Preliminary Investigation	Gait Analysis	Rule Based	Own Database
[8]	Flexible human action recognition in depth video sequences using masked joint trajectories	3D Hybrid models	Support Vector Machine (SVM), Hidden Markov Model (HMV)	Own Database
[9]	Detecting & interpreting self-manipulating hand movements for student's affect prediction	Sobel-operated local binary pattern (SLBP)	Three-layered Bayesian network (BN)	Own Database
[10]	Framework of Single-Frame Face Superresolution Across Head Pose, Facial Expression, and Illumination Variations	face hallucination, Face superresolution,	Pattern Matching	CAS-PEAL-R1 Face Database and CMU PIE database
[11]	Real-World and Rapid Face Recognition Toward Pose and xpression Variations via Feature Library Matrix	3D face reconstruction, feature library matrix	Support Vector Machines	Face Recognition Technology (FERET), Carnegie Mellon University- Pose, Illumination, and Expression (CMU-PIE), and Labeled Faces in the Wild (LFW) face databases
[12]	Recognition of Face Expressions using Local Principal Texture Pattern	LPTP, LBP	Support Vector Machines	CK and JAFFE
[13]	Hand Body Language Gesture Recognition Based on Signals From Specialized Glove and Machine Learning Algorithms	Glove Based	Probabilistic neural network, Support Vector Machine, and k-nearest neighbors algorithm	Real time
[14]	Nonparametric Feature Matching Based Conditional Random Fields for Gesture Recognition from Multi-Modal Video	Probabilistic Model	Structured Support Vector Machines, Naïve Bayes	MSRC-12 Dataset
[15]	Direction Estimation for Pedestrian Monitoring System in Smart Cities: An HMM Based Approach	2D camera model	Hidden Markov Model	CASIA Dataset A, CASIA Dataset B and NITR Conscious Walk Dataset (Own Dataset)
[16]	Vision-based position computation from in-vehicle video log images for road sign inventory	3D based Analytical Model of Road Sign Positioning (AM-RSP) using first order approximation	Mean and deviation of 3D position errors of road sign are computed to present both the position measurements and	

IIOAB JOURNAL
ISSN: 0976-3104

			uncertainties.	

TAXONOMY OF GESTURES

The taxonomy of gestures can be drawn based on various considerations such human computer interaction based gesturing, gesture mappings, physical characteristics etc. [Fig.2] shows the taxonomy based on Human Computer Interaction based gesture classification.

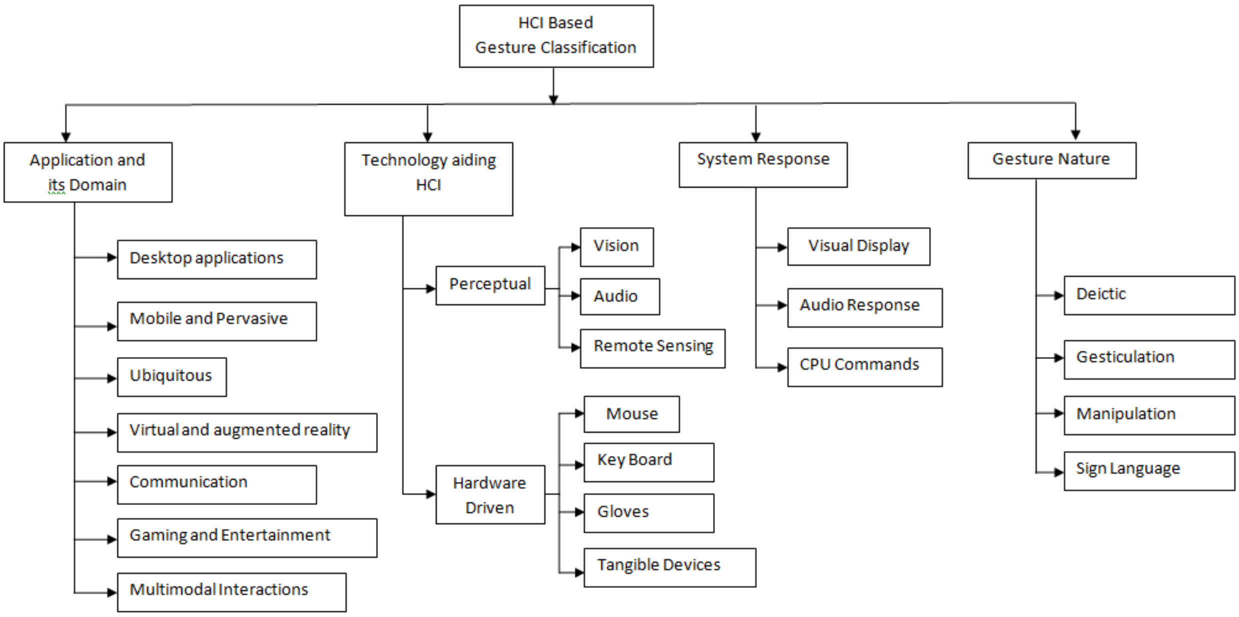

Fig. 2: Gesture taxonomy based on human computer interaction.

Gesture taxonomy can also be classified based on the dimensions under consideration. These include various dimensional aspects which map the gestures to the nature of activity performed, whether the form is static or dynamic, whether the contextual parameters are taken into consideration and whether the data available is discrete or continuous in nature and so on. [Fig.3] shows the classification of gestures based on the dimensional aspects.

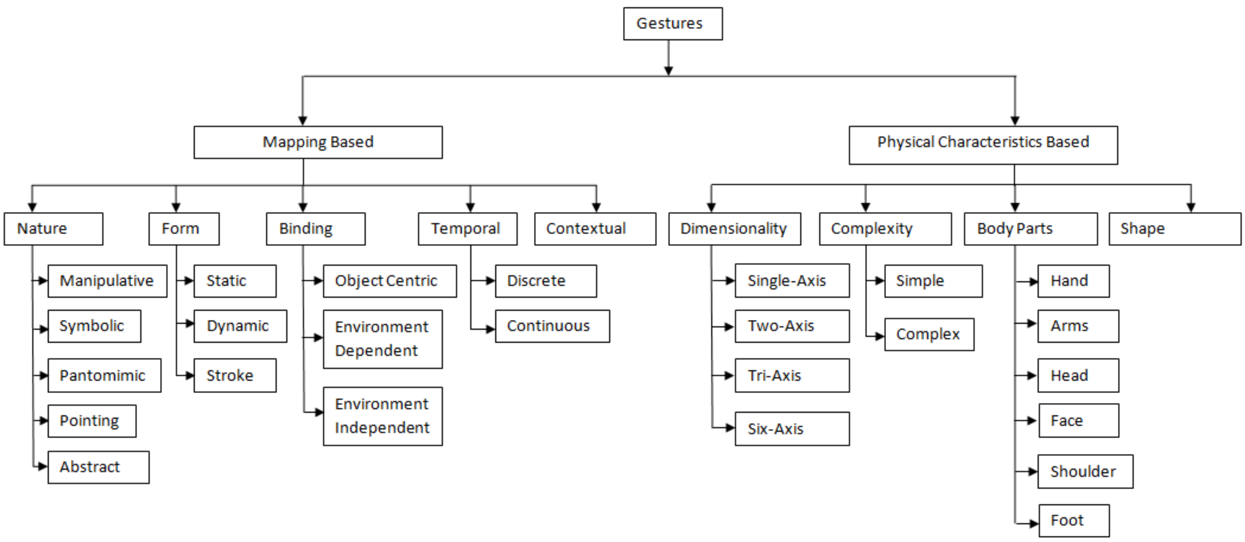

Fig. 3: Gesture taxonomy based on dimensionalities.

FACE GESTURE RECOGNITION TECHNIQUES

Face gesture recognition requires the face to be detected first and then the detected face is tracked. This is followed by extracting and selecting the salient features that will be required for determining the facial

IIOAB JOURNAL
ISSN: 0976-3104

gestures.[Fig. 4] shows the sequential flow of these steps along with the algorithms used across each step.

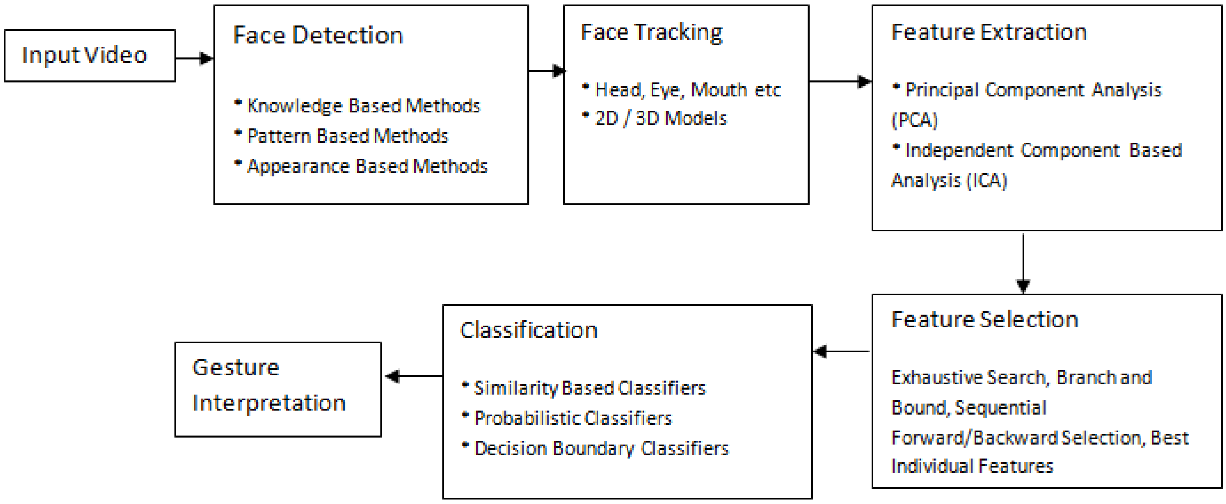

Fig. 4: Phases involved in facial gesture interpretation along with algorithms.

Based on the requirement and the necessity of the application, the respective algorithms are chosen across each phase. [Table 2] lists the various algorithms used across each phase of gesture interpretation along with its salient features.

Table 2: Algorithms used across each phases of gesture interpretation

Module	Algorithm	Sub-types if any	Significance
Face Detection	Knowledge Based Methods		Rule based Ruled-based methods that encode our knowledge of human faces
	Feature Invariant Methods		Algorithms that try to find invariant features of a face despite it's angle or position.
	Template Matching Methods		These algorithms compare input images with stored patterns of faces or features.
	Appearance Based Methods (SVM, Naïve Bayes, Nueral Network Based etc.,)		A template matching method whose pattern database is learnt from a set of training images
Face Tracking	Individual Feature Tracking (Head, eye, mouth etc,.)		Certain features tracked specifically
	2D / 3D Models		2D models track face and output is shown in 2D image space. In 3D models pose and orientation variations are also considered
Feature Extraction	Principal Component Analysis (PCA)		Eigenvector-based, linear map
	Kernel PCA		Eigenvector-based , non-linear map, uses kernel methods
	Weighted PCA		PCA using weighted coefficients
	Linear Discriminant Analysis (LDA)		Eigenvector-based, supervised linear map
	Semi-supervised Discriminant Analysis (SDA)		Semi-supervised adaptation of LDA
	Independent Component Analysis (ICA)		Linear map, separates non-Gaussian distributed features
	Neural Network based methods		Diverse neural networks using PCA, etc
	Multidimensional Scaling (MDS)		Nonlinear map, sample size limited, noise sensitive.
	Self-organizing map (SOM)		Nonlinear, based on a grid of neurons in the feature space
	Active Shape Models (ASM)		Statistical method, searches boundaries
	Active Appearance Models		Evolution of ASM, uses shape and

	(AAM)		texture
	Gabor wavelet transforms		Biologically motivated, linear filter
	Discrete Cosine Transform (DCT)		Linear function, Fourier-related transform, usually used 2D-DCT
	MMSD, SMSD		Methods using maximum scatter difference criterion.
Feature Selection	Exhaustive search		Evaluate all possible subsets of features. Optimal, but too complex.
	Branch and bound		Can be optimal. Complexity of max O(2^n).
	Best individual features		Evaluate and select features individually. Simple but noit very effective.
	Sequential Forward Selection (SFS)		Evaluate growing feature sets (starts with best feature). Retained features cannot be discarded.
	Sequential Backward Selection (SBS)		Evaluate shrinking feature sets (starts with all the features). Deleted features cannot be revaluated.
Classifiers	Similarity Based Classifiers	Template Matching	Assign sample to most similar template. Templates must be normalized
		Nearest Mean	Assign pattern to nearest class mean
		K-NN	Like 1-NN, but assign to the majority of k nearest patterns.
		Vector Quantization Methods	Assign pattern to nearest centroid. There are various learning methods
		Self Organizing Maps	Assign pattern to nearest node, then update nodes pulling them closer to input pattern
	Probabilistic Classifiers	Bayesian	Assign pattern to the class with the highest estimated posterior probability.
		Logistic Classifier	Predicts probability using logistic curve method
	Decision Boundary Classifiers	Fisher Linear Discriminant (FLD)	Linear classifier. Can use MSE optimization
		Binary Decision Tree	Nodes are features. Can use FLD. Could need pruning
		Perceptron	Iterative optimization of a classifier (e.g. FLD)
		Radial Basis Network	Optimization of a Multi-layer perceptron. One layer at least uses Gaussian transfer functions.
		Support Vector Machines	Maximizes margin between two classes.

HAND GESTURE RECOGNITION TECHNIQUES

Hand gesture recognition techniques involve detecting the hand movements, tracking the same and then recognize the respective gestures. As suggested by Rautaray S.S. et.al.,[17] [Fig. 5] shows the taxonomy for various hand gesture recognizing techniques.

IIOAB JOURNAL ISSN: 0976-3104

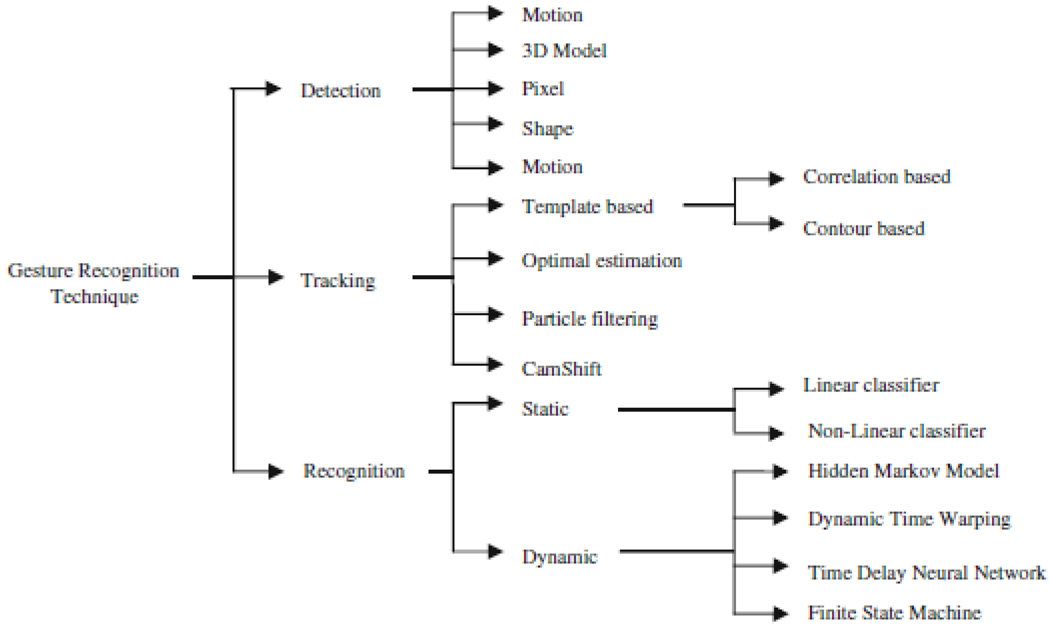

Fig. 5: Vision based hand gesture recognizing techniques.

TAXONOMY OF MACHINE LEARNING TECHNIQUES FOR GESTURE RECOGNITION

Machine learning involves studying how to automatically learn to make accurate predictions based on the past observations. [Fig.6] shows the taxonomy of machine learning techniques.

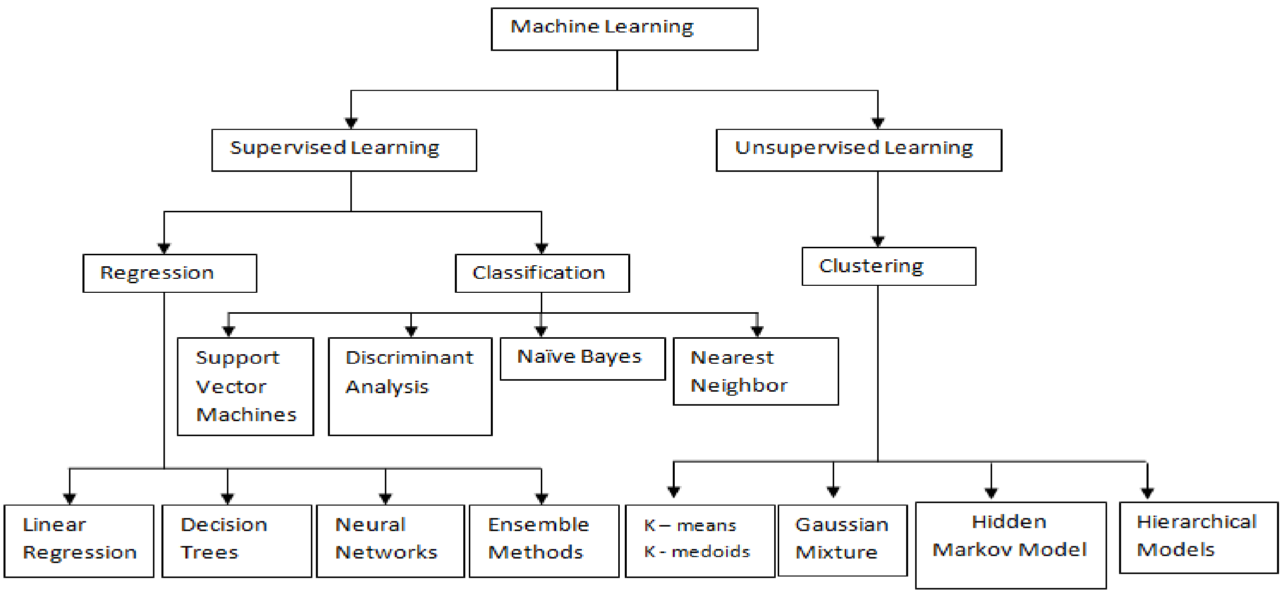

Fig. 6: Machine learning techniques for gesture recognition

In general machine learning techniques can be broadly classified into Supervised and Unsupervised learning. Supervised learning involves developing a predictive model based on both input given to the system and output obtained. Unsupervised learning involves discovering an internal representation from the input data alone. These machine learning techniques are selected according to the nature of the input and the system requirements. [Table 3] lists the merits and demerits of various machine learning techniques.

Table 3: Merits and demerits of different machine learning techniques

Machine learning technique	Advantages	Disadvantages
1R (1 Rule)	Simple Consistent Result Handles Missing Data	Applicable for simple applications only Susceptible to over fitting

Naïve Bayes	Missing values neglected	More assumptions Slow results due to redundant attributes
Bayes Network classifier	Less Assumptions	Complex when compared with Naïve Bayes
Statistical Distributions	Easily Understandable Works well with continuous numeric attributes	Mismatch exists between assumed and actual distributions
Perception	Simple and can handle multiple functions	Can handle linear functions only
Back Propagation	Handles non-linear functions	Difficulty in explaining classification rules Requires large input for effective learning / training
Divide and Conquer	Provides visual representation for rules	An attribute with few values can reduce the result accuracy Replicated sub-tree problem occurs
Genetic Algorithms	Handles both linear and non-linear data Seeks global maximization	Difficult to depict exact structure to user Can sometimes be slow to arrive at a good result
Instance Based	Adding new instances increase learning Simple to implement	Determining appropriate weights for attributes is difficult Time consuming for large data sets

ARCHITECTURE OF GESTURE BASED CONTEXT AWARE INTELLIGENT SYSTEM (GCAIS)

Based on the knowledge acquired by understanding the taxonomies of the various means for studying face and hand gestures, an architecture for gesture based context aware intelligent system is proposed as shown in [Fig 7]. The input to this gesture-based context aware intelligent system may be a video sequence or a live video stream. From the input video, the face and hand portions are detected by the system. This is followed by a series of pre-processing activities that include normalization, thresholding, image resizing, cropping and noise removal. After this, the refined contents are analyzed. The facial reactions are analyzed and the gesture is studied. Similarly the hand movements are analyzed and the hand gestures are identified. Using these gestures, the computational intelligence is acquired by applying various options like feature matching, machine learning techniques, pattern recognition concepts etc.

While computing intelligence, the analyzed gestures are compared with the trained data related to gestures which is available in the database. Similarly, the context in video is compared against the contextual patterns available in the database. The relevancy of gestures is mapped to the context to study the contextual behavior. The computed intelligence is subjected to various measuring algorithms which takes in order to consideration various uncertainties, calibration and gauging factors. Based on all these aspects the outcome is predicted or a decision making is performed. It can be seen from the above architecture that acquiring computational intelligence through various machine learning algorithms plays a significant role in determining the prediction or decision making.

RESEARCH CHALLENGES TOWARDS GESTURE RECOGNITION

According to T.Fang et al. [18], the major challenges faced in Facial Emotion Recognition are Database Challenges, Algorithm Improvement and availability of standardized protocols. In addition, as stated by Zeng et al. [19] another important challenge for the system is to be context aware while making gesture interpretations. This is again an unexplored area in Gesture recognition. Also, the limitations in the availability of hardware components like sensors make it difficult to acquire proper data to prepare well trained datasets [20]. Another key challenge to gesture interpretation is that a same expression may be displayed at different intensity levels [21][22]. It is highly essential for a Gesture based intelligence system to address these challenges while interpreting these gestures.

IIOAB JOURNAL
ISSN: 0976-3104

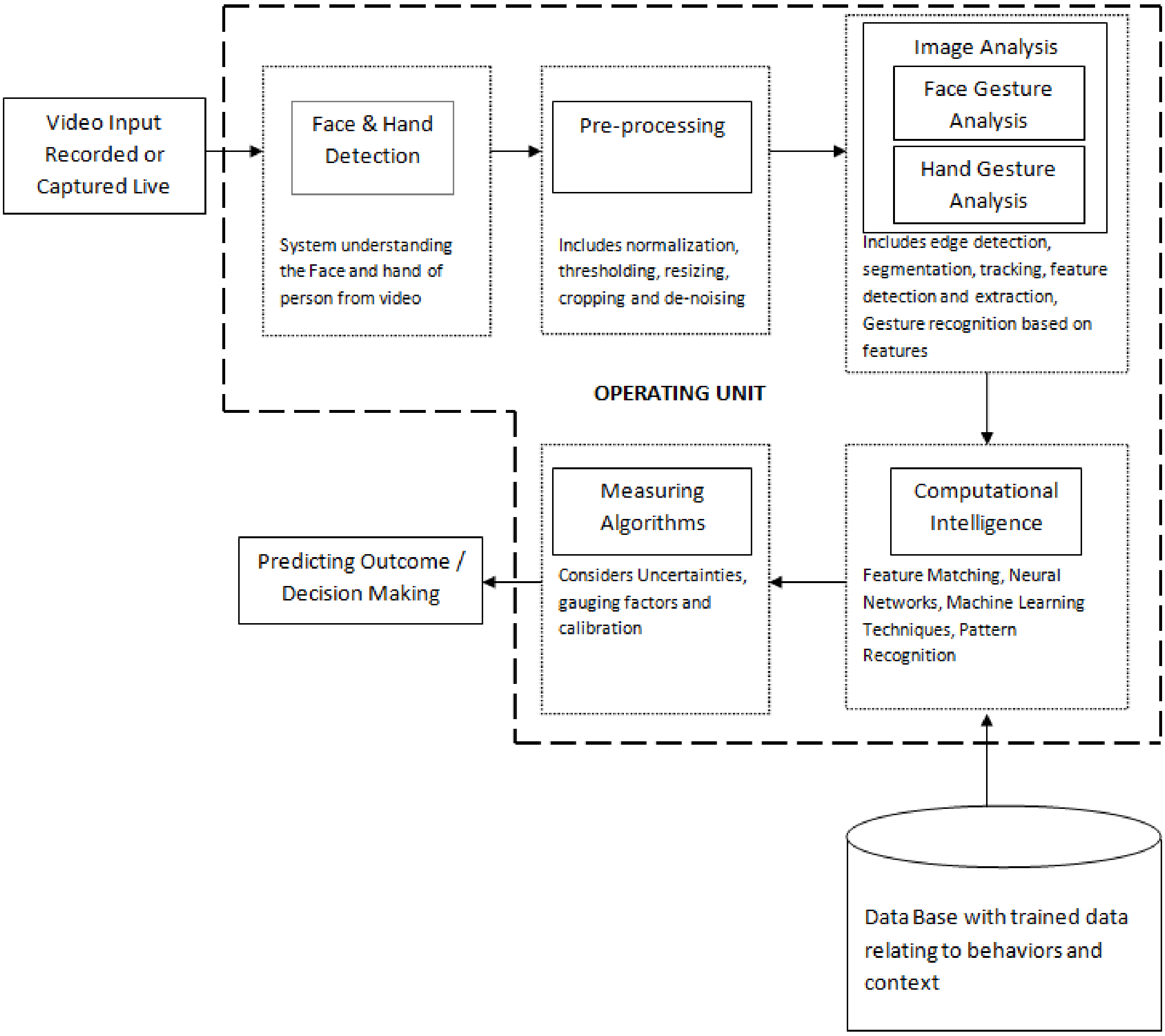

Fig. 7: Architecture of GCAIS

CONCLUSION AND FUTURE WORK

A comprehensive survey of the various techniques, methods and algorithms involved in computer vision based face and hand gesture recognition was carried out and the respective taxonomies were studied. This included study of various techniques involved in face and hand gesture recognition. Also, architecture for a context aware gesture based intelligence system was proposed. Future work will involve implementing the GCAIS architecture using desirable techniques and quantize the results using suitable metrics.

CONFLICT OF INTEREST

None

ACKNOWLEDGEMENTS

None

FINANCIAL DISCLOSURE

None

REFERENCES

[1] P VK. Borges, N Conci,A. Cavallaro. [2013] Video-Based Human Behavior Understanding: A Survey. IEEE Transactions on Circuits and Systems for Video Technology. 23(11):1993-2008

[2] CK Chan, WP Loh, I Abd Rahim. [2016] Human motion classification using 2D stick-model matching regression coefficients. Applied Mathematics and Computation. 283(20):70-89
[3] Maryam Saberi, Ulysses Bernardet, Steve DiPaola. [2014] An Architecture for Personality-based, Nonverbal Behavior in Affective Virtual Humanoid Character. Procedia Computer Science. 41(1):204-211
[4] Sarah N. Rigby, Brenda M. Stoesz, Lorna S. Jakobson.[2016] Gaze patterns during scene processing in typical adults and adults with autism spectrum disorders. Research in Autism Spectrum Disorders. 25:24-36
[5] Viviana Weiss, Guido Bologna, Séverine Cloix, David Hasler, Thierry Pun. [2014] Walking Behavior Change Detector for a Smart Walker. Procedia Computer Science. 39:43-50
[6] R. Sathya, M. Kalaiselvi Geetha. [2015] Framework for Traffic Personnel Gesture Recognition, Procedia Computer Science. 46:1700-1707
[7] Nursuriati Jamil, Nor Haniza Mohd Khir, Marina Ismail, Fariza Hanis Abdul Razak. [2015] Gait-Based Emotion Detection of Children with Autism Spectrum Disorders: A Preliminary Investigation. Procedia Computer Science.76: 342-348
[8] Tejero-de-Pablos A, Nakashima Y, Yokoya N. et al. [2016] Flexible human action recognition in depth video sequences using masked joint trajectories. J Image Video Proc.
[9] Hussain A, Abbasi AR Afzulpurkar N. [2012] Detecting & interpreting self-manipulating hand movements for student's affect prediction. Hum. Cent. Comput. Inf. Sci.
[10] X Ma, H Song, X Qian. [2015] Robust Framework of Single-Frame Face Superresolution Across Head Pose, Facial Expression, and Illumination Variations. IEEE Transactions on Human-Machine Systems. 45(2):238-250
[11] A Moeini, H Moeini. [2015] Real-World and Rapid Face Recognition Toward Pose and Expression Variations via Feature Library Matrix. IEEE Transactions on Information Forensics and Security. 10(5):969-984
[12] A Ramirez Rivera, JA Rojas Castillo and O Chae. [2012] Recognition of face expressions using Local Principal Texture Pattern. IEEE International Conference on Image Processing
[13] P Plawiak, T. Sosnicki, M. Niedzwiecki, Z. Tabor and K. Rzecki. [2016] Hand Body Language Gesture Recognition Based on Signals From Specialized Glove and Machine Learning Algorithms. IEEE Transactions on Industrial Informatics. 12(3):1104-1113
[14] J. Y. Chang. [2016] Nonparametric Feature Matching Based Conditional Random Fields for Gesture Recognition from Multi-Modal Video. IEEE Transactions on Pattern Analysis and Machine Intelligence. 38(8):1612-1625
[15] R. Raman, P. K. Sa, B. Majhi and S. Bakshi. [2016] Direction Estimation for Pedestrian Monitoring System in Smart Cities: An HMM Based Approach. IEEE Access. 4:5788-5808
[16] Z. Hu and N. Li. [2016] Vision-based position computation from in-vehicle video log images for road sign inventory. IET Intelligent Transport System. 10(6):414-420
[17] Rautaray, S.S. & Agrawal.[2015] Vision based hand gesture recognition for human computer interaction: a survey. Artif Intell Review. 43(1):1-54
[18] T Fang, X Zhao, O Ocegueda, SK Shah, IA Kakadiaris. [2011] 3D facial expression recognition: A perspective on promises and challenges. Face and Gestures.
[19] Z Zeng, M Pantic, G Roisman, T Huang. [2007] A survey of affect recognition methods: Audio, visual, and spontaneous expressions. IEEE Transactions on Pattern Analysis and Machine Intelligence. 31(1):39–58
[20] L. Yin, X Wei, Y Sun, J Wang, M Rosato. [2006] A 3D facial expression database for facial behavior research. Proc. 7th International Conference on Automatic Face and Gesture Recognition
[21] M Pantic, L Rothkrantz. [2000] Automatic analysis of facial expressions: The state of the art. IEEE Transactions on pattern Analysis and Machine Intelligence. 22(12):1424–1445
[22] Y Wang, X Huang, C.-S. Lee, S Zhang, Z. Li D. Samaras D. Metaxas A. Elgammal, P Huang. [2004] High resolution acquisition, learning and transfer of dynamic 3-D facial expressions. Computer Graphics Forum. 23(3):677–686