An Integrated HCI Framework for Interpreting Meaningful Expressions

Omer Rashid and Ayoub Al-Hamadi
Institute of Electronics, Signal Processing and Communications (IESK),
Otto von Guericke University Magdeburg, Germany

Abstract: **Integration of different modalities aims to increase the robustness of system and improves its performance in an unconstrained environment. With this motivation, the proposed research is focused on integrating two modalities such as hand gesture and posture in a unified framework where the criterion is measured by employing Particle Filter system. The proposed framework has two main modules: 1) gesture and posture recognition system, and 2) Particle filter based integration of these systems by incorporating Context Free Grammar rules to generate interpretation and inferences. In the recognition part, gesture classification is performed by HMM using hand orientation as a feature. However, the posture signs are firstly categorized into groups by fingertip detection for ASL finger spelling and then classified by SVM with statistical and geometrical features. The integration is performed by mapping the classification outcome on Particle Filter system approximating the probability density function that acts as contribution-weight at decision level. As the computed contribution-weights are independent of classification outcome, so it allows us to resolve classification ambiguities. Based on the contribution-weights, we have obtained a sequence of gesture and posture signs, which are used to infer the "meaningful expressions" by employing CFG for our test scenario. Experiments are conducted on 500 different combinations of restaurant orders with an overall 98.3% inference accuracy whereas classification accuracy of gesture and posture recognition is approximately 98.6% which proves the significance of proposed approach.**

Keywords: Posture Recognition, Gesture Recognition, Integration, Context Free Grammar, Particle Filter

I. Introduction

A lot of research has been done to provide the naturalness and intuitiveness while interacting with the computers. This naturalness and intuitiveness is achieved through HCI which bridges the gap between humans and computers by exploiting different modalities. These modalities are divided according to their input in three main categories namely visual, audio and sensor-based modality. First, the visual modality deals with solving the issues using human responses while taking the observations from a visual input device (i.e. camera). The main research applications in this modality includes gaze detection, face and facial expression, and gesture and posture recognition. Second, the audio-based modality takes the audio signals as the input observations. Speech and speaker recognition, musical and emotion recognition are the application areas in this modality. The third type deals with the sensor-based modality and includes the pen-based interaction, mouse, keyboard and pressure sensors etc. However, the selection of these modalities is a crucial choice and it should provide the desirable level of applicability in terms of naturalness and intuitiveness for HCI systems.

In recent years, HCI envisions to investigate virtual interaction approaches with machines through virtual reality, virtual interfaces and haptic interfaces. The aim is to allow humans to interact virtually and feels the interaction by involving hand or body movement in space which is not possible with traditional 2D devices. This technological development is a big achievement of HCI but with the compromise of ease and convenience. To overcome these limitations, the researchers exploit computer vision algorithms with HCI to devise new interaction methods which open new research avenues in the field of HCI. With this motivation, in this paper, we propose an integrated vision based HCI system using visual modality with hand gesture and posture recognition to support the natural and intuitive interaction [25].

The main contents of this paper are described as follows: overview of literature is presented in Section II which is followed by the main contributions of the paper in Section III. Afterwards, we begin by explaining the overview of proposed approach in Section IV. Section V presents the preprocessing phase where as feature extraction and classification for gesture and posture recognition is presented in Section VI. The integration module is described in Section VII which includes the particle filter system. Section VIII presents the interpretation and inference module along with Context Free Grammar (CFG) inference rules. The experimental results are demonstrated in Section IX whereas the concluding remarks are sketched in section X.

II. Related Work

Vision based approaches for hand gesture and posture-based interaction[1] is an emerging field in which the observations (i.e. inputs) are acquired using camera and different vision algorithms are exploited to process the image streams. We can divide the vision based approaches into two main types [9]. The first type uses the gloves with markers (i.e. specified color glove with marker is used for hand posture and fingertip detection) but it does not reflect naturalness and convenience. The second type offers complete naturalness as it takes the observations directly from bare hand of the user for recognition. In this research, we are considering the later type. The bare hand approaches have some difficulties

[1]Posture is defined as a static sign/expression or a hand pose. In contrast, the gesture is a sequential combination of different instances narrating a particular message (i.e. hand waving).

in the processing and recognition due to skin color detection, objects segmentation and complex background.

In the literature, two different approaches have been adopted to devise methods for vision based hand recognition namely model based approaches and appearance based approaches. In the model based approaches, 3D model of the hand is constructed which contains 3D hand kinematics with certain Degrees Of Freedom (DOF). Moreover, the hand parameters are extracted from this model and are matched with already observed images or image features. However, the focus of our research is based on appearance based approaches in which the image features (i.e. contours, edges, image moments, eigenvectors, fingertip etc) are extracted and compared with the observed features set. There are two major issues which are to be considered in appearance based approaches namely feature selection and dataset training. Feature selection is an essential step where unique feature representation is indispensable to ensure robustness under real-time processing (i.e. invariance to translation, rotation and scaling) for the classification. Dataset training process ensures that the samples should be sufficient for optimized learning of the classifier. Afterwards, classification is performed on the selected features based on the learned parameters. The capability of operating in real-time environment motivates many researchers to explore the applicability of appearance based approaches to recognize gesture and posture for HCI. In the following, we have reviewed each of these in detail within the scope of this work.

A. *Gesture Recognition*

In gesture recognition, Yoon et al. [29] developed a hand gesture system by combining the location, angle and velocity for the recognition. Liu et al. [23] developed a system to recognize 26 alphabets by using different HMM topologies. In the similar context, Hunter et al. [15] used HMM for recognition where Zernike moments are used as image features for hand gesture sequences. With a different motivation, Chen et al. [7] presented a system for gesture recognition in which hands are recognized using Haar-like features. Moreover, in this approach, the training algorithm is used based on AdaBoost which selects different Haar-like features for the classification. Bretzner et al. [4] presented an approach for static backgrounds to recognize the hand gestures using multi-scale color features. In a hierarchical model, the shape and color cues are combined with image features at different levels. Further, particle filter is used for the tracking and recognition of hand state. Similarly, Elmezain et al. [10] presents a gesture spotting and recognition framework for the isolated and key gestures to classify numbers from 0 to 9 using HMM.

B. *Posture Recognition*

Among many applications of posture recognition, sign language recognition is one of the application domains for HCI systems to interpret hand posture as sign. Mainly, the sign language recognition is categorized into three main groups namely finger spelling, word level sign and non-manual features [3]. Lamar and Bhuiyant [20] proposes a technique for feature extraction of the hand postures based on Principal Component Analysis (PCA) and perform the analysis on American Sign Language (ASL) and Japanese Sign Language (JSL) using multilayer Neural network classifier. However, the drawback of the approach is the use of colored gloves and testing the approach on static background. Isaacs and Foo [16] proposed an approach using two layer feed-forward neural network to recognize fingerspelling ASL alphabets. Similarly, Handouyahia et al. [12] proposed a recognition system based on shape description using size functions for International Sign Language (ISL). Neural Network is used to train alphabets from the features computed for sign languages. However, the computed features in their proposed approach are not rotation invariant.

Malassiotis and Strintzis [24] used the Elliptic Fourier Descriptor for 3D hand posture recognition. In their system, orientation and silhouettes of the hand are used to recognize 3D hand postures. Similarly, Licsar and Sziranyi [21] used Fourier coefficients from modified Fourier descriptor approach to model hand shapes for hand gestures recognition. Altun et al. [1] proposed a method to increase the effect of fingers in fingerspelling Turkish Sign Language (TSL). In their method, hand shapes with strong edges are extracted and matched against the template.

C. *Approaches using Data Fusion*

Integration of different systems is used to enhance the performance and results in better recognition of subjects under observation. In this context, integration of different modalities is used to improve the recognition (i.e. identification of a human by combining face and voice traits [5]) in the field of biometrics. Moreover, in the multi-modal biometric systems, fusion takes place at different levels which includes sample level, feature level, match score level and decision level fusion [27]. Chang et al. [6] proposed a face recognition system in which the fusion of 2D and 3D information of the face images is performed to improve the recognition. Particular for the hand recognition, Kumar et al. [19] performed fusion at feature level and match score level to combine the palm prints and hand geometrical features. Similarly, Wu et al. [28] proposed a multi-model system to combine the gait recognition with face recognition system to recognize the humans. It is observed that the main motivation of exploiting different modalities is to achieve better performance and to cop the limitations of uni-modal approach.

III. Contributions of the Paper

The main contributions of the paper are elaborated in following aspects:

1. In gesture and posture framework, we have extracted invariant feature vectors which results in robust recognition along with lower number of training samples [26]. Besides, we have categorized the posture signs according to detected fingertips by curvature analysis to avoid the mis-classifications among posture signs which results in the improvement of posture classification rate.

2. According to our knowledge, integration of gesture and posture recognition system is not addressed yet. We have designed an effective interaction-interface for HCI by integrating gesture and posture recognition system

IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 5, No 3, September 2012
ISSN (Online): 1694-0814
www.IJCSI.org

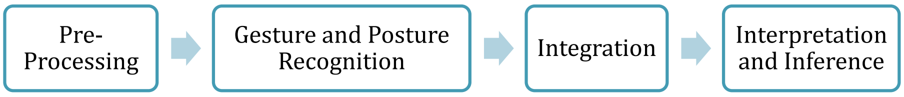

Figure. 1: presents the process flow of the proposed framework.

where the integration criteria is computed by particle filter system at decision level. These contribution-weights sets a criteria for the combination of extracted symbols which are then mapped on CFG rules and results in the interpretation of new meaningful expressions based on the developed lexicon database.

3. The use of particle filter in gesture and posture recognition helps in resolving the ambiguities occurred due to low classification rate (i.e. posture signs can be classified even when the classification percentage is 20%).

IV. The Proposed Framework

The proposed framework is staged in several modules (see Fig. 1) to infer the meaningful expressions by integrating gesture and posture recognition modalities in which the integration criterion is based on particle filter. First, the preprocessing step is carried out in which the data is acquired by Bumblebee2 camera and the objects of interest (i.e. hands and face) are extracted using color and depth information. Second, gesture and posture feature vectors are computed by exploiting different properties of hand (i.e. statistical and geometrical) for the categorization and classification. Further, HMM is employed to recognize the gesture symbols from alphabets and numbers whereas for the posture signs, fingertip categorization is performed to group finger-spelling ASL signs and then posture signs are classified by SVM. For the integration of gesture and posture modalities, a particle filter system is proposed to define integration-criteria (i.e. by computing the contribution-weights). Afterwards, the interpretation is performed by processing CFG production rules which results in the inference of meaningful expressions. In the following, detailed description of each phase is presented along with experimental results.

V. Pre-Processing

Our pre-processing step consists of two modules namely data gathering and skin color segmentation. These modules are presented as follows:

A. Data Gathering

In the suggested approach, Bumblebee2 camera is used for capturing the input sequence which gives us stereo information as 2D image and depth sequences. The input 2D images are firstly transformed from RGB color space to YC_bC_r color space for hand and face segmentation because skin color lies in a small region of chrominance components in YC_bC_r color space. In this manner, the effect of brightness variation is reduced by ignoring the luminance channel. Further, the depth images are exploited to select the region of interest for segmentation of hands and face where the depth lies in range from 30 cm to 200 cm (i.e. in our experiments).

B. Skin Color Segmentation

In the proposed approach, the skin color pixels are detected from 2D images and are then modeled by normal Gaussian distribution characterized by mean and variance as shown in Fig. 2(b). Normal Gaussian distribution probability for an observation $\mathbf{x}$ is calculated as:

$$\mathcal{P}(\mathbf{x}) = \frac{1}{2\pi\sqrt{|\Sigma|}} e^{-\frac{1}{2}\left((\mathbf{x}-\mu)^T\Sigma^{-1}(\mathbf{x}-\mu)\right)} \tag{1}$$

where μ and Σ represents mean vector and covariance matrix respectively of training data. The computed probability $\mathcal{P}(\mathbf{x})$ derived from the above model categorizes the pixels as a skin pixel or not (shown in Fig. 2(c)). After that, skin color pixels are binarized and contours are extracted by computing chain code representation for detection of hands and face. In the initialization phase, we have used Haar detector to detect face in the image. Moreover, left and right hands are marked according to their spatial positions in the image.

VI. Gesture and Posture Recognition

In this section, components of the proposed gesture and posture recognition are presented which consist of two main modules namely feature extraction and classification.

A. Feature Extraction

Feature extraction is an essential step for the recognition process but selection of good features is always challenging. Hand gesture and posture features are described as follows.

1) Gesture Features

In the proposed approach, hand orientation is used as a feature and is determined between two consecutive centroid points when drawing gesture path. The equation used to compute the orientation θ_t is:

$$\theta_t = arctan\left(\frac{y_{t+1}-y_t}{x_{t+1}-x_t}\right); t = 1, 2, ..., T-1 \tag{2}$$

where T represents length of gesture path, x_t and y_t are centroid points at frame t. The computed angle θ_t is quantized in range from 1 to 18 by dividing it by 20 degrees. These quantized values give us discrete vector which is used in HMM to classify gesture symbols. The feature set $\mathcal{G}_{fv}$ is denoted as:

$$\mathcal{G}_{fv} = (\theta_t)^T \tag{3}$$

2) Posture Features

In the proposed approach, posture features computation comprises of two steps. The first step presents the categorization

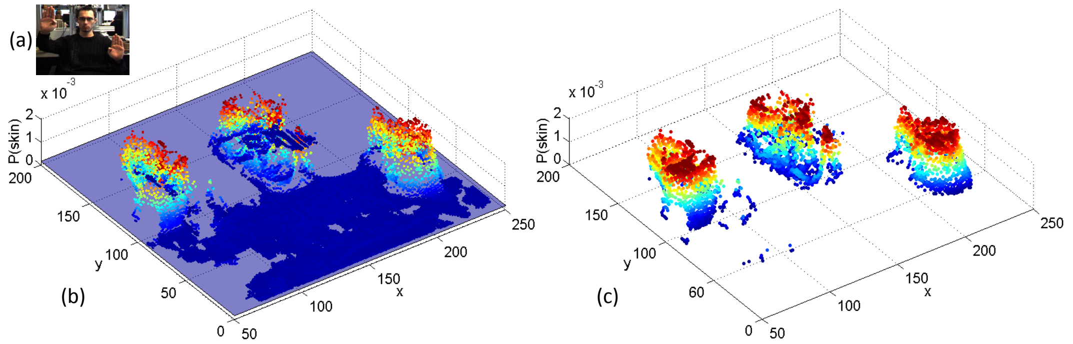

Figure. 2: (a) input image (b) shows the skin color pixel probabilities of the depth region (i.e. 30cm to 200cm) using normal Gaussian distribution (c) shows the detected skin color pixels for hand contour extraction and binarization.

of posture signs based on detected fingertips (i.e. by curvature ($\mathcal{C}$)). The statistical and geometrical features are presented in the second step which are then given to the classification module for the recognition. The posture feature set is described as follows:

$$\mathcal{P}_{fv} = \{\mathcal{C}, \mathcal{P}_{(stat,geo)}\} \tag{4}$$

Fingertip Detection for Categorization: The categorization of ASL alphabets is performed based on the number of detected fingertips from the hand's contour. The main idea in categorizing ASL alphabet signs is to reduce the misclassifications observed among these symbols. In the proposed approach, we have formed four groups for ASL alphabets according to number of detected fingertips (i.e. Group I with no fingertip, Group II with one fingertip, Group III with two fingertips and Group IV with three fingertips).

Given the contour of detected hand, curvature is estimated by considering the neighbor contour points to detect the fingertip [18]. Mathematically, curvature gives the ratio of length (i.e. sum of distances that a curve has) and displacement measures the distance from the first to last point if curve covers a straight line. Curvature $\mathcal{C}(k)$ is computed from the following equation:

$$\mathcal{C}(k) = \left\| \left(P_{k-n/2} - P_{k+n/2}\right) \right\|^{-1} \sum_{i=(k-n/2)}^{i=(k+n/2)} \left\|(P_i - P_{i+1})\right\| \tag{5}$$

where k is the contour point of object at which curvature is estimated, n is total number of pixels used for curvature estimation, and P_i and $P_{(i+1)}$ are the objects contour points. The main idea of fingertip categorization is to find high curvature values from contour points which results in detection of peaks from hands contour and provides clue about the fingertip. The number of contour points are adaptively determined (i.e. after conducting empirical experiments) by exploiting the depth information to find the distance between object and camera.

In Fig. 3(a and b), experimental results show the images and contour pixels of left hand with curvature values as peaks on z-axis. The contour points with value greater than $\sqrt{2}$ are selected as the candidates for fingertip (see. peaks and valleys in Fig. 3). In Fig. 3(a), we have extracted two clusters named as C_1 and C_2 with values above threshold and the maximum value from these individual clusters are selected using maximum local extreme value. The resulted points/-peaks are marked as a fingertip. However, it is observed that both the peaks in the hand's contour (i.e. $\cap$) and valleys (i.e. $\cup$) can be inferred as a fingertip. Therefore, the next step is to remove valleys from being detected as a fingertip. For this purpose, selected contour points are taken and their distances are computed from the center point of the hand. Further, the normalization is done and these points are scaled ranging from 0 to 1. We pick the points whose values are greater than 0.5 for fingertip detection. In this way, fingertips (FT) are successfully detected for the categorization of ASL alphabet signs. Table. A presents the confusion matrix of detected fingertips among the groups.

Statistical and Geometrical Features: After the categorization, statistical (i.e. Hu-Moment $\mathcal{P}_{hu}$) and geometrical properties (i.e. $\mathcal{P}_{geo}$) of the hand posture signs are computed. It is denoted as:

$$\mathcal{P}_{(stat,geo)} = \{\mathcal{P}_{hu}, \mathcal{P}_{geo}\} \tag{6}$$

Statistical Feature Vector: In Eq. 6, $\mathcal{P}_{hu}$ represents the Hu-Moments [14] which are derived from basic moments, and describes the properties of objects shape statistically (i.e. area, mean, variance, covariance and skewness etc). Hu [14] derived a set of seven moments which are orientation, translation and scale invariant. Equations of Hu-Moments are defined as:

$$\phi_1 = \eta_{20} + \eta_{02} \tag{7}$$

$$\phi_2 = (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2 \tag{8}$$

$$\phi_3 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2 \tag{9}$$

$$\phi_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2 \tag{10}$$

$$\phi_5 = (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] + (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \tag{11}$$

$$\phi_6 = (\eta_{20} - \eta_{02})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03}) \tag{12}$$

$$\phi_7 = (3\eta_{12} - \eta_{03})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] + (3\eta_{12} - \eta_{30})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \tag{13}$$

From the above equations, the seven moments (i.e. $\phi_1, \ldots, \phi_7$) are derived from second and third order moments

IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 5, No 3, September 2012
ISSN (Online): 1694-0814
www.IJCSI.org

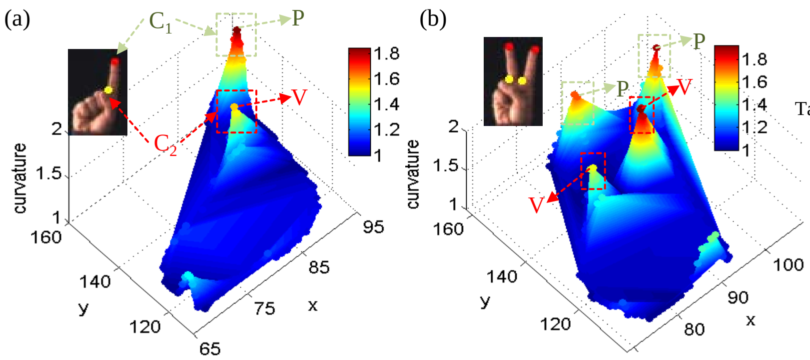

Table A: Confusion Matrix of Fingertips (i.e. all)

FT/Gr.	1	2	3	4
0	99.8	0.2	0.0	0.0
1	1.2	96.8	2.0	0.0
2	0.0	0.0	95.1	4.9
3	0.0	0.0	0.9	99.1

Figure. 3: (a) shows the posture signs "D" with contour pixels and curvature values. The clusters (C_1 and C_2) are presented with threshold above $\sqrt{2}$ and maximum local extreme values are selected from these clusters. These selected values are the interest points for fingertip detection which are then normalized to identify the peaks and valleys in the hand. The peak values represent the fingertip of the hand in the image. (b) presents the posture sign "V" (i.e. the cluster formation and normalization steps are same as in (a)). (c) Table A presents the confusion matrix of fingertips for categorization.

whereas zero and first order moments are not used in this process. The first six Hu-Moments are invariant to reflection [8] however the seventh moment changes the sign. Our statistical FV contains these seven moments and are presented in the following set:

$$\mathcal{P}_{hu} = (\phi_1, \phi_2, \phi_3, \phi_4, \phi_5, \phi_6, \phi_7)^T \tag{14}$$

Geometrical Feature Vector: In Eq. 6, the second feature set is geometrical FV ($\mathcal{P}_{geo}$) which is used to describe ASL sign shape. We have exploited the properties of circularity ($Cir = \frac{Perimeter^2}{4\pi \times Area}$) and rectangularity ($Rect = \frac{Area}{l \times w}$) as the measures of the shape that defines how much object's shape is closer to circle and rectangle respectively. We show that these geometric attributes are capable of providing good representations for ASL signs as the shapes varies significantly for each corresponding ASL sign. Geometrical FV is stated as:

$$\mathcal{P}_{geo} = (Cir, Rect)^T \tag{15}$$

Both the statistical and geometrical FV set are combined together to form a feature vector set and is defined as:

$$\mathcal{P}_{(stat,geo)} = \{\phi_1, \phi_2, \phi_3, \phi_4, \phi_5, \phi_6, \phi_7, Cir, Rect\}^T \tag{16}$$

Normalization: The normalization is pertinent for posture feature vectors to keep them in a particular range for the classification module. Our normalized feature vector is defined as:

$$\mathcal{NP}^i_{(stat,geo)} = \frac{(\mathcal{P}^i_{(stat,geo)} - c^i_{min})}{(c^i_{max} - c^i_{min})}\,; \tag{17}$$

$$c^i_{min} = \mu^i - 2\sigma^i\,,\; c^i_{max} = \mu^i + 2\sigma^i \tag{18}$$

$\mathcal{NP}^i_{(stat,geo)}$ is the normalized feature vectors for posture recognition where i refer to respective feature. c^i_{max} and c^i_{min} are the respective maximum and minimum values used for normalization of these features.c^i_{max} and c^i_{min} are selected from FV of training data and it defines the upper and lower limit for the normalization of the respective feature vector. μ^i and σ^i are the mean and standard deviations of the training data respectively. The normalization is done because SVM needs the normalized data for the training and testing of posture signs.

Table 1: Confusion Mat: 1-Detected FT

Sign	A	B	D	I	H/U
A	99.8	0.0	0.0	0.0	0.2
B	0.0	98.2	1.0	0.0	0.8
D	0.0	0.0	98.7	1.3	0.0
I	0.6	0.0	0.8	98.6	0.0
H/U	0.0	3.1	0.0	0.2	96.7

Table 2: Confusion Mat: 2-Detected FT

Sign	C	L	P	Q	V	Y
C	98.7	0.2	0.0	0.7	0.0	0.4
L	0.4	98.5	0.0	0.7	0.0	0.4
P	0.0	0.0	98.7	1.3	0.0	0.0
Q	0.0	0.0	3.8	96.2	0.0	0.0
V	0.2	0.0	0.0	0.0	99.3	0.5
Y	0.0	0.0	0.0	0.0	0.7	99.3

B. *Classification*

Two classification approaches are employed for gesture and posture recognition and are described as follows:

1) *Gesture Classification*

In the classification of gesture signs (i.e. alphabets and numbers), Baum-Welch (BW) algorithm is used to train the parameters of HMM [11] by the discrete vector θ_t. We have used Left-Right banded model with 9 states for hand motion recognition of gesture path. Classification of hand gesture path is done by selecting the maximal observation probability of the gesture model by the Viterbi algorithm. In our case, the maximal gesture model is the classified symbol and has the largest observation probability among all the alphabets (i.e. A-Z) and numbers (i.e. 0-9).

2) *Posture Classification*

In the classification of posture signs, a set of thirteen ASL alphabets and ten ASL numbers are recognized using Support Vector Machines (SVM). SVM [22] is a supervised learning technique for optimal modeling of data. We have used

Table 3: Confusion Matrix of ASL Numbers

Nr.	0	1	2	3	4	5	6	7	8	9
0	99.8	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1	0.3	99.4	0.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2	0.0	0.0	98.3	0.4	0.0	0.0	1.3	0.0	0.0	0.0
3	0.0	0.0	0.4	98.2	0.9	0.0	0.5	0.0	0.0	0.0
4	0.0	0.0	0.0	0.2	98.2	1.6	0.0	0.0	0.0	0.0
5	0.0	0.0	0.0	0.0	2.4	97.6	0.0	0.0	0.0	0.0
6	0.0	0.0	0.8	0.6	0.0	0.0	98.6	0.0	0.0	0.0
7	0.0	0.0	0.0	0.0	0.0	0.0	0.7	98.3	0.6	0.4
8	0.0	0.0	0.0	0.4	0.0	0.0	0.2	0.4	98.4	0.6
9	0.0	0.0	0.0	0.5	0.0	0.0	0.0	0.4	0.5	98.6

normalized statistical and geometrical features vectors (i.e. $\mathcal{NP}^i_{fv}$) to train and classify signs using Radial Basis Function (RBF). RBF is a Gaussian kernel which works robustly with given number of features and provides optimum results when compared to other kernels. In the experiments, categorization is performed to make groups based on number of detected fingertips (see. confusion matrix in Table A). Confusion matrix (CM) in Table 1 and Table 2 presents classification probabilities of Group 2 (i.e. $\mathcal{A}, \mathcal{B}, \mathcal{D}, \mathcal{I}, \mathcal{H}/\mathcal{U}$) and Group 3 (i.e. $\mathcal{C}, \mathcal{L}, \mathcal{P}, \mathcal{Q}, \mathcal{V}, \mathcal{Y}$) respectively. Group 1 (i.e. $\mathcal{A}, \mathcal{B}$) has no mis-classifications and Group 3 has only one alphabet (i.e. $\mathcal{W}$). Table 3 presents confusion matrix of ASL numbers.

VII. Integration

In this research, the basic idea of the integration is the parallel interpretation of multiple signs driven from different modalities to infer meaningful expressions[2]. To achieve this objective, we have combined the gesture and posture modalities through particle filter system that allows us to infer new symbols at any instance of time. Integration $\mathcal{I}$ of gesture and posture systems fused at decision level is formulated as:

$$\mathcal{I} = \langle \alpha_{gstr} \times \mathcal{R}_{hmm} \rangle \bigcap \langle \alpha_{pstr} \times \mathcal{R}_{svm} \rangle \tag{19}$$

where $\mathcal{R}_{hmm}$ and $\mathcal{R}_{svm}$ are the classification outcome of gesture and posture system. α_{gstr} and α_{pstr} are the contribution-weights corresponding to gesture and posture recognition which defines the integration criteria. In the following, we present our algorithm to compute contribution-weights based on particle filter.

A. Particle Filter System

Condensation algorithm [17]) is a form of Bayesian estimation which provides a way to compute the a-posteriori probability (i.e. contribution-weight) and allows an effective integration process due to recursive state estimation process. A system of particle filters (i.e. comprising of two separate particle filters (i.e. for gesture and posture)) is implemented and for which the idea, algorithm and its processes are explained as follows.

The key idea of particle filter (PF) is to approximate the probability density function using a collection of random samples with associated weights from classification outcome. In our particle filter system, we maintain the classification outcomes of gesture and posture recognition as state vectors which are represented by $x_{gstr} = [\mathcal{R}_{hmm}]$ and $x_{pstr} = [\mathcal{R}_{svm}]$ respectively.

Initialization. In the initialization phase, the particles are generated which begins by obtaining a set of initialization observations as shown in Fig. 4(a),(b). The parameters for each particle is then generated by sampling from Gaussian distribution describing the classification outcomes. In the proposed approach, the measurements (i.e. for gesture and posture) at each time instance t is described as $z_m = \{z_{gstr}, z_{pstr}\}$ where z_{gstr} and z_{pstr} are the measurements of gesture and posture modalities. A set of particles in vector $\mathcal{S}(n)$ is represented as follows:

$$\mathcal{S}(n) = \{s_k^{(gstr)}, s_k^{(pstr)}\} \tag{20}$$

A set of N random points (i.e. 100) called particles x_k^n with weights w_k^n denotes the initial distribution of particles at time k for both gesture and posture systems[3]. These particles are denoted as:

$$s_k^{(gstr|pstr)} = \{x_k^n, w_k^n\}_N^{n=1} \tag{21}$$

We propose a generic framework for probability distribution in the particle filter as follows:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(\mathbf{x}-(z_m)^2)}{(2\sigma^2)}} \tag{22}$$

where σ is the standard deviation for the gesture or posture modality. The above distribution is sampled for each new particle using a random index from the cumulative probability for each value of **x**.

Selection. In the selection step, factored sampling is used to select the particles based on their weights and is achieved by selecting a random index from the cumulative weight of particles. Consequently, the particles having weights closer to the peaks in the probability distribution are sampled many times whereas the samples with low values in probability distributions are discarded. For the new observations, we only

[2]The classified gesture and posture signs are used to fetch the specific information stored corresponding to the symbol in the designed data-dictionary for restaurant lexicon scenarios. This fetched information is treated by CFG inference rules which results in the formulation of "meaningful expressions"

[3]The same notation is used for both the particle filters (i.e. gesture and posture), except when stated otherwise.

keep the samples with weights more than the average weights where as the samples falling below the average weights are discarded and are replaced with new samples using the initialization distribution.

Prediction. The prediction step reflects the underlying temporal behavioral model where the particles are moved to generate the hypothetical state (x_k) at time k is based on the previous state (x_{k-1}). This step involves sampling from the state transition and is formulated as follows:

$$p^{(n)}\left(x_k|z_{k-1}\right) = p^{(n)}\left(x_k|x_{k-1}\right)p^{(n)}\left(x_{k-1}|z_{k-1}\right) \tag{23}$$

where $p\left(x_k|z_{k-1}\right)$ is a-priori probability, $p\left(x_{k-1}|z_{k-1}\right)$ is the previous a-posteriori probability and $p\left(x_k|x_{k-1}\right)$ is the state transition model.

Updation. In the updation step, we compute the contribution-weights (a-posteriori probability) through the computed a-priori probability $p(x_k|z_{k-1})$ and the likelihood $p(z_k|x_k)$ by incorporating the new measurement z_k extracted from the classification outcomes of gesture and posture recognition systems (i.e. where the propagation process follows the same process 22). The contribution-weights $\alpha_{gstr|pstr}$ or a-posteriori probability $p(x_k|z_k)$ for gesture and posture system is computed as follows:

$$\alpha_{gstr|pstr} = p\left(x_k|z_k\right) = \frac{\sum_{n=1}^{N} p^{(n)}\left(z_k|x_k\right)p^{(n)}\left(x_k|z_{k-1}\right)}{\sum_{n=1}^{N} p^{(n)}\left(z_k|x_k\right)} \tag{24}$$

Using N values of $p(z_k|x_k)$, we have obtained a probability distribution for the state space at time instance k. Each sample is assigned a weight according to its particular position in state space relative to observational density. In this way, we obtain the contribution-weights which defines the integration-criteria for the fusion of these system. The same procedure is followed for each frame as presented in Fig. 4(d-f). The final integration is carried out when contribution-weights of gesture α_{gstr} and posture α_{pstr} signs satisfy the threshold ($T = 70\%$) at any time frame.

$$(\alpha_{gstr}|\alpha_{pstr}) \geq T \tag{25}$$

After obtaining the contribution-weights, we have used AND/OR combination for gesture and posture recognition signs. Integration $\mathcal{I}$ is formulated as:

$$\mathcal{I} = \langle\alpha_{gstr} * \mathcal{R}_{hmm}\rangle \bigcap \langle\alpha_{pstr} * \mathcal{R}_{svm}\rangle \tag{26}$$

$$\mathcal{I} = \langle\alpha^{i}_{gstr} * \mathcal{R}^{i}_{hmm}; i = 1\cdots m\rangle \bigcap \langle\alpha^{j}_{pstr} * \mathcal{R}^{j}_{svm}; j = 1\cdots n\rangle \tag{27}$$

In the suggested approach, integration of gesture and posture module interprets and infers when multiple posture symbols (i.e. described above as n) are combined with one gesture symbol (i.e. m). So, in the next section, we present our interpretation and inference module for the fusion of gesture and posture modalities to generate meaningful expressions.

VIII. Interpretation and Inferences

Having the contribution-weights, our interpretation module presents the possible outcomes for the integration of gesture and posture modalities. In the inference module, the developed lexicon is presented along with CFG inference rules. These modules are demonstrated as follows:

A. Interpretation

In this module, the different interpretations derived for the integration of gesture and posture recognition system are stated which includes:

1. $\langle Gesture \Rightarrow Detected\rangle$; $\langle Posture \Rightarrow Detected\rangle$; $\langle Integration \Rightarrow Yes\rangle$ *Description:* The ideal case of integration in which both gesture and posture systems recognize the symbol at any time frame.

2. $\langle Gesture \Rightarrow NotDetected\rangle$; $\langle Posture \Rightarrow Detected\rangle$; $\langle Integration \Rightarrow No\rangle$ *Description :* Gesture system does not classify any symbol because HMM is not activated when gesture drawing process starts. In contrast, the posture system classifies the sign with the contribution-weights α_{pstr} above the threshold.

3. $\langle Gesture \Rightarrow SemiDetected\rangle$; $\langle Posture \Rightarrow Detected\rangle$; $\langle Integration \Rightarrow Yes/No\rangle$ *Description:* There can be some predictions about gesture symbols dependent upon the inference from HMM states. In this case, gesture symbol is still incomplete and it gives a clue about user's intention while drawing the gesture symbol. Intentions are predicted only when contribution-weight α_{gstr} of gesture sign pass the threshold criterion.

4. $\langle Gesture \Rightarrow NotDetected\rangle$; $\langle Posture \Rightarrow NotDetected\rangle$; $\langle Integration \Rightarrow No\rangle$ *Description:* No match has occurred from gesture and posture systems.

B. Inference

For building the inferences, the recognition from both the gesture and posture modalities are important. So, in the gesture recognition, the HMM classifier and its states model recognizes the alphabets and numbers after processing some frames. In contrast, posture recognition system recognizes the symbol at every frame because a single frame is sufficient to recognize ASL symbols in finger spelling domain (i.e. Exceptions are "J" and"Z" and are not considered here). In the proposed approach, we treat the integration as a problem of

Table 4: Lexicon of Symbols

Symbols ⇒ Fruits	Symbols ⇒ Fruits
A⇒Apple, Apricot	N⇒Nectarine
B⇒Blueberry, Banana	O⇒Orange, Oval Kumquat
C⇒Cherry, Cantaloupe	P⇒Pear, Peach
D⇒Date, Dewberry	Q⇒Quince
E⇒Elderberry, Eggfruit	R⇒Raspberry, Rambutan
F⇒Fig, Farkleberry	S⇒Star Fruit, Strawberry
G⇒Grapes, Gooseberry	T⇒Tangerine, Tart Cherry
H⇒Honeymelon, Hackberry	U⇒Ugli Fruit, Uniq Fruit
I⇒Imbe	V⇒Voavanga
J⇒Jackfruit, Jambolan	W⇒Watermelon, Wolfberry
K⇒Kaffir Lime, Kiwi	X⇒Xigua
L⇒Lemon, Lychee	Y⇒Yunnan Hackberry
M⇒Mango, Melon	Z⇒Zinfandel Grapes

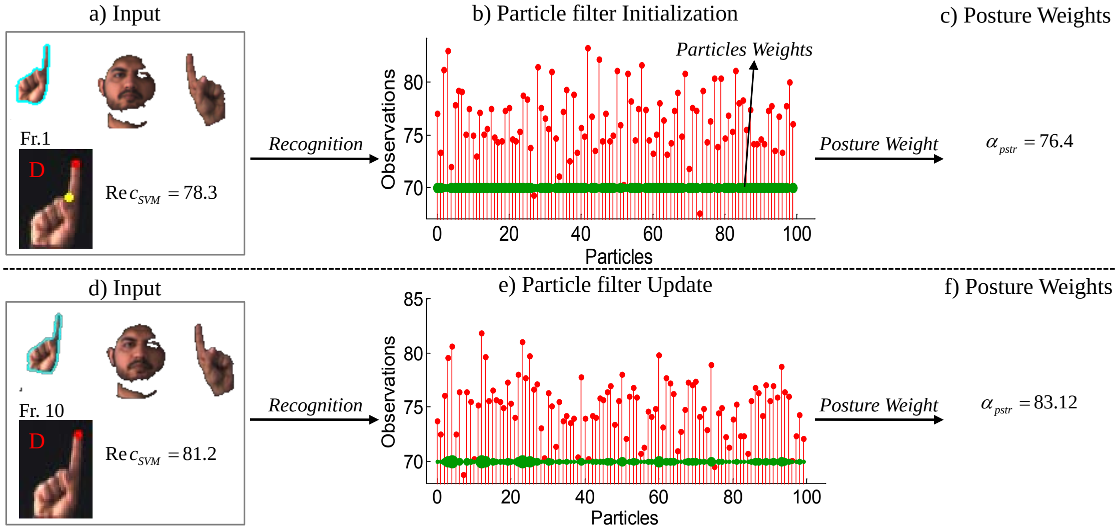

Figure. 4: shows the particle filter process for the frames 1 and 10. a) presents the classification of ASL sign "D" with recognition outcome "78.3". b) indicates the particle filter weights initialization based on classification outcome. c) computes the maximum weight from the corresponding particles and measures the posture contribution-weight. d) presents the classification of ASL sign "D" with recognition rate "81.2" for frame 10. e) indicates the particle filter updation based on classification outcome. f) computes the maximum weight from the corresponding particles and measures the posture contribution-weights.

regular language where the gesture and posture modalities are governed by Context Free Grammar (CFG [13]) inferencing rules. Before describing the context specific inference rules which we have employed in this research, it is essential to first present the proposed structure of the language. The CFG grammar is defined in a quadruple (i.e. 4-tuple) as follows:

$$Grammar = \langle \mathcal{V}, \mathcal{T}, \mathcal{S}, \mathcal{R} \rangle \tag{28}$$

where $\mathcal{V}$ is the set of objects and contains non-terminals as well as terminals symbols, $\mathcal{T}$ is the set of terminals, $\mathcal{S}$ is start symbol and it is a subset of $\mathcal{V}$ (i.e. $\mathcal{S} \in \mathcal{V}$), and $\mathcal{R}$ is the set of production rules. The inference of our CFG rules (see. Context Free Grammar Inference Rules) is presented as follows:

$$\langle Posture_{Alphabet} \rangle \Rightarrow \textit{The set of recognized posture alphabet symbols;} \tag{29}$$

$$\langle Gesture_{Alphabet} \rangle \Rightarrow \textit{The set of recognized gesture alphabet symbols;} \tag{30}$$

$$\langle Posture_{Digit} \rangle \Rightarrow \textit{The set of recognized numbers.} \tag{31}$$

Different symbols can be devised in integration process depending upon the lexicon as shown in Table 4. The inferences of "meaningful expressions" derived from CFG rules is as follows:

$$S \rightarrow \langle Posture_{Alphabet} \rangle \langle Gesture_{Alphabet} \rangle \langle Posture_{Digit} \rangle$$

where the derivation is referred as the posture sign $\langle Posture_{Alphabet} \rangle$ followed by $\langle Posture_{Digit} \rangle$ whereas $\langle Gesture_{Alphabet} \rangle$ yields to one outcome during the inference process.

IX. Experimental Results

In the proposed approach, the experimental setup involves the tasks of data acquisition, gesture and posture classification and particle filter based integration process and is then linked to CFG inference rules to generate "meaningful expressions" by the mechanism of interpretations and inferences. We have demonstrated the applicability of proposed approach on our real-time example scenario and show the description of meaningful expressions generated from the integration of these systems. Unlike other domains of computer vision such as image recognition where a variety of datasets are available to continue the improvement in the suggested solution. A quite few datasets of ASL are available such as [2] which are designed for specific applications with non-flexible assumption. A potential factor is the domain of research fields (i.e. both HCI and computer vision) which is very much context sensitive and application oriented.

Defs. Rules 1 Context Free Grammar (CFG)*

$Definitions\ and\ Rules:$

$$\begin{aligned}
\mathcal{V} = \{&\mathcal{S}, Posture_{Alphabet}, X, Gesture_{Alphabet}, Alphabet, \\
&Posture_{Digit}, Y, Digit, 0_p|1_p, ..., 9_p, a_g|b_g, ..., z_g, \\
&a_p|b_p, ..., z_p\}
\end{aligned}$$

$$\mathcal{T} = \{0_p|1_p, ..., 9_p, a_g|b_g, ..., z_g, a_p|b_p, ..., z_p\}$$

$$\begin{aligned}
\mathcal{S} &\rightarrow \langle Posture_{Alphabet} \rangle \langle X \rangle \\
Posture_{Alphabet} &\rightarrow \langle Alphabet \rangle \langle Posture_{Alphabet} \rangle \mid \\
&\quad\ \langle Alphabet \rangle \\
X &\rightarrow \langle Posture_{Digit} \rangle \langle Y \rangle \\
Posture_{Digit} &\rightarrow \langle Digit \rangle \langle Posture_{Digit} \rangle \mid \langle Digit \rangle \\
Y &\rightarrow \langle Gesture_{Alphabet} \rangle \langle Posture_{Digit} \rangle \\
Digit &\rightarrow 0_p|1_p|2_p, ..., 9_p \\
Alphabet &\rightarrow a_p|b_p|c_p, ..., z_p \\
Gesture_{Alphabet} &\rightarrow a_g|b_g|c_g, ..., z_g
\end{aligned}$$

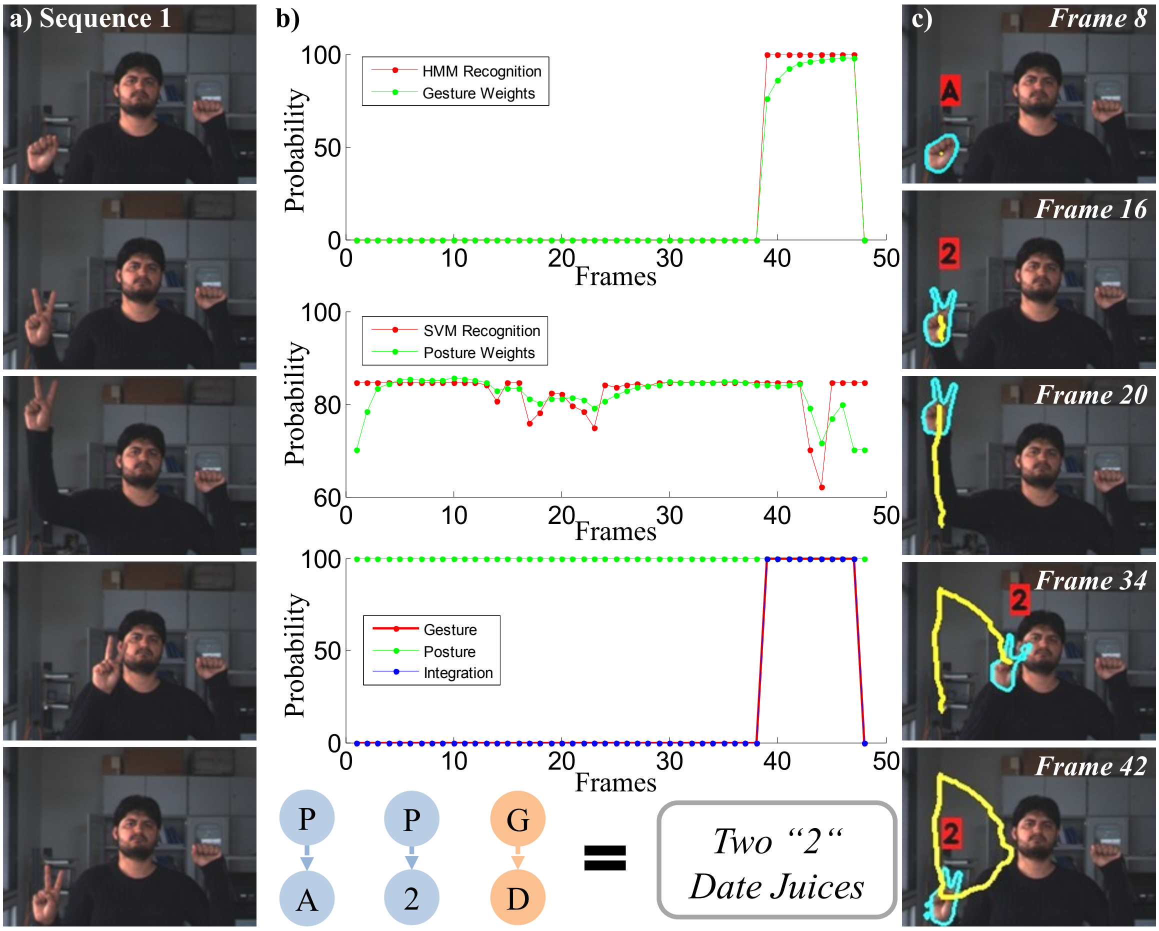

Figure. 5: presents the meaningful expression "Two Date Juices" which results from recognized gesture symbol "D" and classified posture alphabet "A" and number "2". (a) presents the images from sequence 1. (b) The first two graphs in this figure presents the recognition of gesture and posture systems along with contribution-weights whereas the third graph presents gesture and posture results with their integration. (c) presents the images having gesture and posture recognition.

Based on aforementioned information, we have developed our own dataset in the laboratory which comprises of eight actors performing the gesture and posture signs where the image sequences are captured by Bumblebee2 stereo camera with 240*320 pixels image resolution. These developed signs are used for the training and testing purposes in which the gesture signs contain 30 video observations for each gestures sign whereas 3600 image observation are used for posture signs. Out of these video and image observations, 20 videos and 2400 image observations are used as training observations for gesture and posture respectively. It is worth to mention that, we have tested our examples case independent to classification process. So, the applicability can be extended by designing the lexicon and CFG rules according to the scenario under observation.

The proposed concept of integration is tested on a real-time example scenario, for instance we have designed restaurant lexicon which reflects the functionality of food and drink order placement at counter. For this purpose, we have studied type of food and drink item in a menu (e.g. name of fruit, drinks, fast food, etc.). In this work, we have chosen 45 different fruits for this choice as shown in Table. 4 and make different (i.e. currently our system supports 500 combinations) choices for the menu-order by combining recognized gestures and postures signs. For instance, an order is placed by integrating the first and second/third alphabet of the fruit name from gesture and ASL posture where the quantity of the desired fruit item is inferred by concatenating with another posture sign number as shown in Fig. 5 and Table. 4.

Fig. 5 and Fig. 6 presents an interpretation based on fusion of gesture and posture recognition system. In Fig. 5, the posture system firstly recognizes the alphabet "A". However, gesture recognition system did not recognize any symbol during the initial frames due to HMM nature. The next posture symbol recognized is "2" which indicates the quantity of order. From frames 38 to 48, gesture recognition system computes the probability of possible signs which the user can draw depending on HMM states and most likely candidates for the gesture recognition. Moreover, it selects the highest probability element and mark it the "best" element for recognition. At frame 48, the gesture ends and the recognized symbol is "D", thus completing the order (i.e. $\langle Rec_{pstr} = \text{"}A\text{"}\rangle, \langle Rec_{pstr} = \text{"}2\text{"}\rangle, \langle Rec_{gstr} = \text{"}D\text{"}\rangle$). The first two graphs in Fig. 5(b) presents the classification and weight-contribution results of gesture and posture recognition for whole sequence. The recognition of gesture and posture system after applying the threshold is presented in third graph along with the integration of these systems. In this sequence, the recognized gesture elements for the first order is $\langle Date = \text{"}D\text{"}\rangle$, $\langle Posture_{Alphabet} = \text{"}A\text{"}\rangle$ which means $Date$ and from the posture recognized sign, it is "2". It means $\langle Two\ Date\ Juices \rangle$. In Fig. 6, the interpretation starts in which the user draws the posture symbol "L". The next posture sign recognized is the number which describes the quantity as "2" and finally the gesture symbol which has been recognized is the symbol "B" (i.e. $\langle Rec_{pstr} = \text{"}L\text{"} \rangle, \langle Rec_{pstr} = \text{"2"} \rangle, \langle Rec_{gstr} = \text{"}B\text{"} \rangle$). Gesture and posture recognition works optimally and recognizes the signs correctly (See Graphs in Fig. 6). The second order from the gesture and posture recognition is $\langle Two\ Blueberry\ Juices \rangle$. By changing the lexicon, the proposed approach can be used in other scenarios.

In the graph of Fig. 6, the classification outcome of SVM in frames $15-18$ is 65%, frames 30 is 55% and different recognitions from frames $20 - 26$ which we have termed as ambiguous classification outcome. So, we have addressed the ambiguous behavior for our integration process through particle filter where the current state t is linked with its previous state at frame $t-1$ and therefore, the previous weights in particle filter affects on the current classification outcome. So, under the ambiguous behavior of SVM, we can handle and control the computation of contribution weights with particle filter. We argument that when the recognition result iteself is considered as contribution-weights, the process of integration suffers due to this ambiguous behavior.

We have tested our proposed approach on the restaurant lexicon database with the overall 98.3% inference accuracy. It is observed that the classification inaccuracies do not affect the performance due to particle filter based weight computation technique. One of the potential reasons is, the particle filter works on the principle of prediction and updation mechanism, therefore, the inference of meaningful expression is achieved successfully.

X. Conclusion

In this paper, we have proposed a framework for recognition of gesture signs , ASL posture signs and their integration, interpretation and inferences. In gesture and posture recognition systems, features are extracted which are invariant to translation, orientation and scaling. Besides, fingertips are detected for ASL alphabets and used as a measure to categorize thus avoids the mis-classifications among posture signs. SVM is applied for recognition of ASL signs whereas HMM is used for the classification of gesture symbols. Moreover, a novel approach is proposed for the integration of gesture and posture recognition in which contribution-weights are computed using Particle filter system by incorporating CFG rules. The proposed approach is tested on restaurant lexicon which successfully integrates both systems and enables to interpret multiple inferences at the same instance of time. Future research is focused on words recognition for gesture and posture systems along with their integration.

Acknowledgments

This work is supported by Transregional Collaborative Research Center SFB/TRR 62 "Companion-Technology for Cognitive Technical Systems" funded by German Foundation.

References

[1] O. Altun and S. Albayrak. Increasing the effect of fingers in fingerspelling hand shapes by thick edge detection and correlation with penalization. In *Pacific-Rim Symp. on Image and Video Tech.*, pages 1133–1141, 2006.

[2] V. Athitsos, C. Neidle, S. Sclaroff, J. Nash, A. Stefan, Y. Q., and A. Thangali. The asl lexicon video dataset. In *CVPR, Workshop on Human Communicative Behaviour Analysis*, 2008.

[3] R. Bowden, A. Zisserman, T. Kadir, and M. Brady. Vision based interpretation of natural sign languages. In *in Proc. 3rd ICCV*, pages 391–401, 2003.

[4] L. Bretzner, I. Laptev, and T. Lindeberg. Hand gesture recognition using multi-scale colour features, hierarchical models and particle filtering. In *Automatic Face and Gesture Recognition*, pages 423–428, 2002.

[5] R. Brunelli and D. Falavigna. Person identification using multiple cues. *IEEE Trans. on PAMI*, 17:955–966, 1995.

[6] K. Chang, K. W. Bowyer, and P. J. Flynn. Face recognition using 2d and 3d facial data. In *ACM Workshop on Multimodal User Authentication*, pages 25–32, 2003.

[7] Q. Chen, N. Georganas, and E. Petriu. Real-time vision based hand gesture recognition using haar-like features. In *Instrumentation and Measurement Technology Conference*, 2007.

[8] J. Davis and G. Bradski. Real-time motion template gradients using intel cvlib. In *IEEE ICCV Workshop on Framerate Vision*, 1999.

[9] D. Dimov, M. Alexander, and N. Zlateva. Cbir approach to the recognition of a sign language alphabet. In *International Conference on Computer Systems and Technologies CompSysTech'07*, pages 1–9, 2007.

[10] M. Elmezain, A. Al-Hamadi, and B. Michaelis. A novel system for automatic hand gesture spotting and recognition. *WSCG*, pages 89–96, 2009.

[11] M. Elmezain, A. Al-Hamadi, O. Rashid, and B. Michaelis. Posture and Gesture Recognition for Human-Computer Interaction. *In-Tech, "Advanced Technologies",Book Chapter*, pages 415–440, 2009.

[12] M. Handouyahia, D. Ziou, and S. Wang. Sign language recognition using moment-based size functions. In *Int. Conf. of Vision Interface*, pages 210–216, 1999.

[13] J. E. Hopcroft, R. Motwani, and J. D. Ullman. *Introduction to Automata Theory, Languages, and Computation*. Pearson Addison-Wesley, 2. edition, 2001.

[14] M. Hu. Visual pattern recognition by moment invariants. *IRE Trans. on Information Theory*, 8(2):179–187, 1962.

[15] E. Hunter, J. Schlenzig, and R. Jain. Posture estimation in reduced-model gesture input systems. In *International Workshop on Automatic Face-and Gesture-Recognition*, pages 290–295, 1995.

[16] J. Isaacs and S. Foo. Hand pose estimation for american sign language recognition. In *Thirty-Sixth Southeastern Symposium on IEEE System Theory*, pages 132–136, 2004.

[17] M. Isard and A. Blake. Condensation - conditional density propagation for visual tracking. *Int. Jour. of Computer Vision*, 29:5–28, 1998.

[18] D. H. Kim and M. J. Kim. A curvature estimation for pen input segmentation in sketch-based modeling. *Computer-Aided Design*, 38(3):238 – 248, 2006.

[19] A. Kumar, D. Wong, H. Shen, and A. Jain. Personal verification using palmprint and hand geometry biometric. In *4th Int. Conf. on Audio and Video-based Biometric Person Authentication*, pages 668–678, 2003.

[20] M. Lamar and M. Bhuiyant. Hand alphabet recognition using morphological pca and neural networks. In *International Joint Conference on Neural Networks*, pages 2839–2844, 1999.

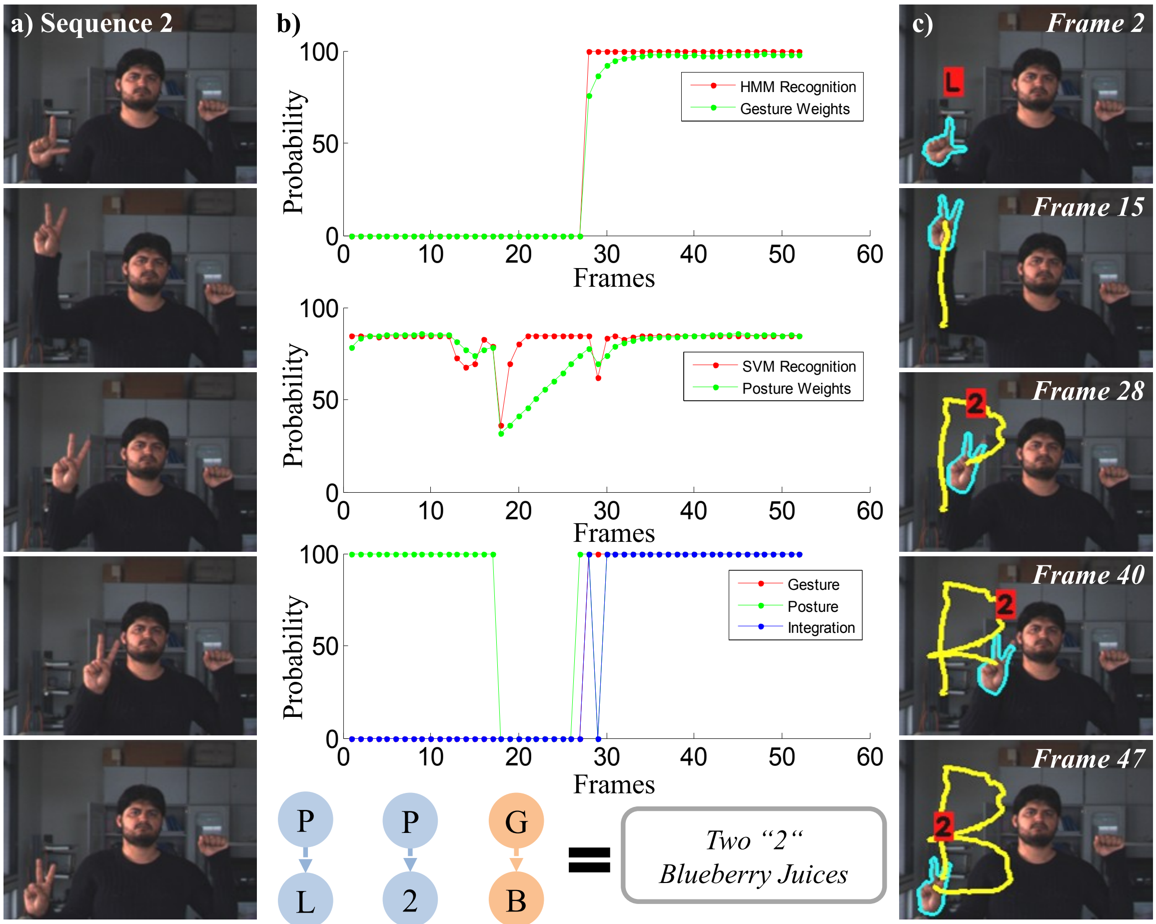

Figure. 6: presents the meaningful expression "Two Blueberries Juices" which results from recognized gesture symbol "B" and classified posture alphabet "L" and number "2". (a) presents the images from sequence 2. (b) The first two graphs in this figure presents the recognition of gesture and posture systems along with contribution-weights whereas the third graph presents gesture and posture results with their integration. (c) presents the images having gesture and posture recognition.

[21] A. Licsar and T. Sziranyi. Supervised training based hand gesture recognition system. In *ICPR*, pages 999–1002, 2002.

[22] C. Lin and R. Weng. Simple probabilistic predictions for support vector regression. Technical report, National Taiwan University, 2004.

[23] N. Liu, B. Lovel, and P. Kootsookos. Evaluation of hmm training algorithms for letter hand gesture recognition. In *IEEE Int. Sym. on SPIT*, pages 648–651, 2003.

[24] S. Malassiotis and M. Strintzis. Real-time hand posture recognition using range data. *Image and Vision Computing*, 26(7):1027–1037, 2008.

[25] O. Rashid, A. Al-Hamadi, and B. Michaelis. A framework for the integration of gesture and posture recognition using hmm and svm. In *IEEE International Conference on Intelligent Computing and Intelligent Systems*, pages 572–577, 2009.

[26] O. Rashid, A. Al-Hamadi, and B. Michaelis. Utilizing invariant descriptors for finger spelling american sign language using svm. In *ISVC*, 2010.

[27] A. Ross and A. Jain. Multimodal biometrics: An overview. In *12th Signal Processing Conference*, pages 1221–1224, 2004.

[28] Q. Wu, L. Wang, X. Geng, M. Li, and X. He. Dynamic biometrics fusion at feature level for video-based human recognition. pages 152–157, 2007.

[29] H. Yoon, J. Soh, Y. Bae, and H. Yang. Hand gesture recognition using combined features of location, angle and velocity. *Pattern Recognition*, 34:1491–1501, 2001.

A. Biography

Omer Rashid is a PhD student in Otto-von-Guericke University, Magdeburg, Germany. He received his Bachelors degree in Computer Engineering from University of Engineering and Technology, Lahore, Pakistan. Further, he has completed his Master degree from Otto-von-Guericke University, Magdeburg, Germany. His current research is focused on human computer interaction, image processing and pattern recognition.

Ayoub K. Al-Hamadi was born in Yemen in 1970. He received his Masters Degree in Electrical Engineering in 1997 and his PhD in Technical Computer Science at the Otto von Guericke University, Magdeburg, Germany in 2001. Since 2002 he has been Junior-Research-Group-Leader at the Institute for Electronics, Signal Processing and Communications at University of Magdeburg. In 2008, he became Professor of Neuro-Information Technology. His research work concentrates on the field of image processing, pattern recognition and artificial neural networks. Professor Al-Hamadi is the author of more than 170 articles in peer-reviewed international journals and conferences.