Recent Methods and Databases in Vision-based Hand Gesture Recognition: A Review

Pramod Kumar Pisharady and Martin Saerbeck

*Institute of High Performance Computing, A*STAR #16-16 Connexis, 1 Fusionopolis Way, Singapore-138632, Email: pramodkp@mit.edu, Saerbeckm@ihpc.a-star.edu.sg*

Abstract

Successful efforts in hand gesture recognition research within the last two decades paved the path for natural human-computer interaction systems. Unresolved challenges such as reliable identification of gesturing phase, sensitivity to size, shape, and speed variations, and issues due to occlusion keep hand gesture recognition research still very active. We provide a review of vision-based hand gesture recognition algorithms reported in the last 16 years. The methods using RGB and RGB-D cameras are reviewed with quantitative and qualitative comparisons of algorithms. Quantitative comparison of algorithms is done using a set of 13 measures chosen from different attributes of the algorithm and the experimental methodology adopted in algorithm evaluation. We point out the need for considering these measures together with the recognition accuracy of the algorithm to predict its success in real-world applications. The paper also reviews 25 publicly available hand gesture databases and provides the web-links for their download.

Keywords: Gesture recognition, posture recognition, hand pose estimation, sign language recognition, hand gesture dataset, gesture database, survey

1. Introduction

Nonverbal communication, which includes communication through hand gestures, body postures, and facial expressions makes up about two-thirds of all communication among human [1]. Hand gestures are one of the most common category of body language used for communication and interaction. Whilst the rest of the body indicates a more general emotional state, hand gestures can have specific linguistic content in it [2]. Due to the speed and expressiveness in interaction, hand gestures are widely used in sign languages and human-computer interaction systems.

One ongoing goal in human-machine interface design is to enable effective and engaging interaction. For example, vision-based hand gesture recognition (HGR)

systems can enable contactless interaction in sterile environments such as hospital surgery rooms, or simply provide engaging controls for entertainment and gaming applications. However HGR is not as robust as standard keyboard and mouse based interaction. Issues such as sensitivity to size and speed variations, poor performance against complex backgrounds and varying lighting conditions, and the reliable detection of gesturing phase have limited the use of hand gestures as a reliable modality in interface design.

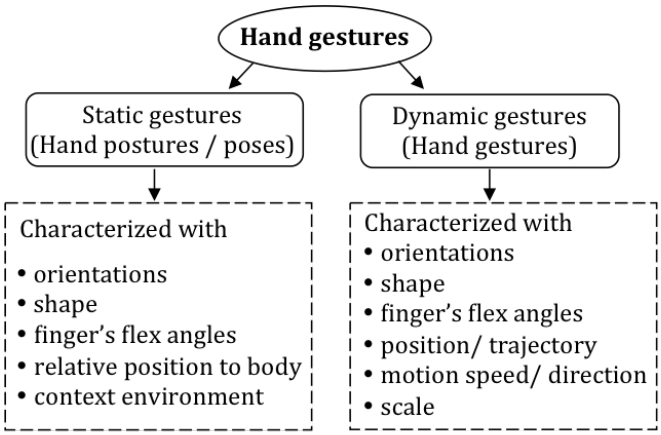

Figure 1: Classification of hand gestures based on temporal nature. Static gestures are time independent whereas dynamic gestures are time dependent.

1.1. Taxonomy of Gestures

There are multiple ways to categorize hand gestures, *1)* based on observable features and *2)* based on the interpretation. In the first category gestures are classified based on temporal relationships, into two types; *static* and *dynamic* gestures (Figure 1). Static hand gestures (*aka* hand postures / hand poses) are those in which the hand position does not change during the gesturing period. Static gestures mainly rely on the shape and flexure angles of the fingers. In dynamic hand gestures, the hand position changes continuously with respect to time. Dynamic gestures generally have three motion phases: preparation, stroke, and retraction [3]. The message in a dynamic gesture is mainly contained in the temporal sequence in the stroke phase. Dynamic gestures rely on the hand trajectories and orientations, in addition to the shape and fingers' flex angles.

In the second category, gestures are classified based on the interpreted meaning. For example emblems, illustrators, regulators, affect displays, and adaptors [4, 5] are the typical classes to describe gestures. Emblems (also labeled as autonomous gestures) are gestures that can be substituted for spoken words (for example, showing *thumbs-up* instead of saying *all right*). Illustrators are gestures

used to illustrate spoken words (for example, giving directions by *pointing*). Regulators support the interaction and communication between speaker and listener (for example, raising hand to manage turn-taking). Affect displays are facial expressions, which when combined with postures reflect the intensity of an emotion (for example, staring at an object and moving the body back reflect the emotion *fear*). Adaptors are gestures used at some point in time for personal convenience, but have turned into a habit (for example, adjusting glasses in a tensed situation).

1.2. Hand Gesture Recognition

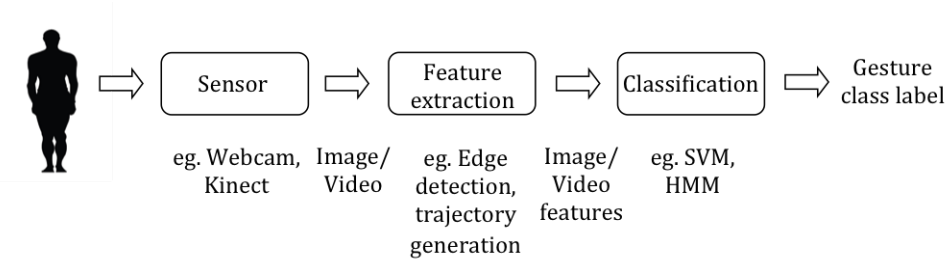

Figure 2: Gesture recognition pipeline.

Figure 2 shows the block diagram of a typical contactless gesture recognition system. The sensor is a camera in vision-based gesture recognition systems. Berman *et al.* [6] reviewed different sensors used in gesture recognition systems and provided a comprehensive analysis of integration of sensors into gesture recognition systems and their impact on the system performance. Based on feature extraction, vision-based gesture recognition systems are broadly divided into two categories, appearance-based methods and three dimensional (3D) hand model-based methods. Appearance-based methods utilize features of training image to model the visual appearance, and compare these parameters with the features of test image. Three-dimensional model-based methods rely on a 3D kinematic model, by estimating the angular and linear parameters of the model.

1.3. Survey and Evaluation of Hand Gesture Recognition Techniques

Our study builds on top of earlier attempts to survey the field of HGR. Mitra *et al.* [7] provided a survey of different gesture recognition methods, covering hand and arm gestures, head and face gestures, and body gestures. The HGR methods investigated in the survey was limited to Hidden Markov Models (HMM), particle filtering and condensation algorithms, and Artificial Neural Networks (ANN). Hand modeling and 3D motion based pose estimation methods are reviewed in [8] (ignoring the gesture classification schemes). An analysis

of sign languages, grammatical processes in sign gestures, and issues relevant to the automatic recognition of sign languages are discussed in [9]. The latest of the above papers ([8]) covered developments till the year 2005. The review concluded that the methods studied are experimental and their use is limited to laboratory environments.

This paper reviews recent works in HGR with a focus on the developments in the last 16 years. Algorithms utilizing conventional RGB cameras (Section 2) as well as the new generation RGB-D cameras (Section 3) are surveyed, making the review unique. The HGR methods are classified and analyzed according to the technique used for gesture classification. We perform a quantitative comparison of HGR algorithms based on different attributes of the algorithm and the experimental methodology followed in algorithm testing. A description of available hand gesture databases (Section 4) and a discussion on hand gesture recognition research (Section 5) are also provided. We hope this survey is timely, given the growing research efforts and expanding market for gestural interactive systems.

2. Conventional Hand Gesture Recognition: RGB Sensor Based Methods

2.1. Recognition of Dynamic Hand Gestures

The techniques used for dynamic HGR can be classified as *a)* HMM [10–23] and other statistical methods [24–31], *b)* ANN [32–34] and other learning based methods [35, 36], *c)* Eigenspace based methods [37, 38], *d)* Curve fitting [39], and *e)* Dynamic programming [40]/ Dynamic time warping [41, 42] (Figure 3).

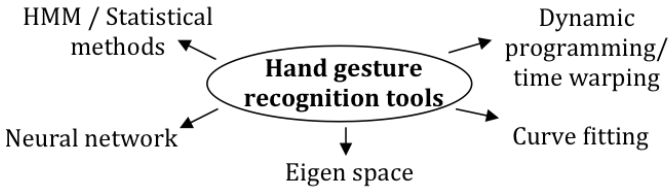

Figure 3: Taxonomy of hand gesture recognition techniques reviewed.

2.1.1. HMM and Other Statistical Methods

HMM is the most widely used HGR technique. HMM is a statistical model in which the system being modeled is assumed to be a Markov process with unknown parameters. HMM represents the statistical behavior of an observable symbol sequence using a network of hidden states with transition and emission probabilities. The HMM can be used for pattern recognition once the hidden parameters are identified using the observable data.

HMM based dynamic hand gesture recognition methods mainly utilize temporal and spatial features of input images. Chen *et al.* [14] utilized Fourier descriptor and optical flow based motion analysis to characterize spatial and temporal features respectively. The algorithm extracts hand shape from complex backgrounds by tracking the hand in realtime. HMM based recognizers identify the best likelihood gesture model for a given pattern. The variations in gesture from a reference pattern reduce the likelihood of the gesture with the model. Lee and Kim [10] introduced an HMM based *threshold model* concept to filter out patterns with less likelihood. Hand movement direction is used to represent the spatio-temporal sequences of gestures. The method reliably detects an end point of a gesture, and finds the start point by backtracking.

Table 1: Hand gesture recognition methods: Features used, classification methods, and reported applications

Work	**Features used**	**Classification method**	**Application**
[10]	direction of hand movement	HMM	browsing commands in $PowerPoint^{(R)}$ presentation
[11]	hand location, angle & velocity	HMM	HCI- recognizing alphanumeric characters & graphic elements
[14]	Fourier descriptors/ optical flow	HMM	Taiwanese sign language
[12]	hand shape & hand motion	HMM	remote robot control
[17]	3D articulation data	accumulative HMM	controlling lights and curtains in smart home
[13]	3D trajectory, hand displacement, color & shape of hand blob	HMM & IOHMM	interact-play, manipulation
[25]	haar-like features	statistical/ syntactic anal.	not specified
[24]	directional features	DBN	controlling media player
[39]	3D motion trajectory	curve fitting	3D bioinformatics data visualization navigation
[37]	hand shape / trajectory	predictive eigen tracker	audio player control
[32]	2D motion field / trajectory	NN	American sign language
[34]	Fourier descriptors (shape of hand blob)	RBF, HMM & RNN	manipulation of objects in windows user interface
[29]	hand motion (motion energy)	FSM	HRI
[42]	3D hand motion features	CDFD & Q-DFFM	Dutch sign language

Descriptions: **HMM**-hidden Markov model, **IOHMM**-input / output hidden markov model, **HCI** -human computer interaction, **DBN**-dynamic Bayesian network, **NN**-neural network, **RBF**-radial basis function, **RNN**-recurrent neural networks, **FSM**- finite state machines, **HRI**-human robot interaction, **CDFD**-combined discriminative feature detectors, **Q-DFFM**-quadratic classification on discriminative features fisher mapping

HMM is based on homogeneous Markov chains as the dynamics of the system

Table 2: Hand gesture recognition methods: Features of the algorithms and experimental methodology adopted in algorithm testing (list at bottom provides description of column titles). Features in column 6 onwards are binary, 1 represents compliance of the work to the feature whereas 0 represents non-compliance.

Work	**Accuracy**	**Class**	**Subj.**	**Samp.**	**UI.**	**Spot**	**BG**	**Noise**	**Scale**	**Light**	**Exten.**	**CV**	**Data**
[10]	93.14	10	8	6.2	0	1	0	0	0	0	0	0	0
[11]	93.25	48	20	5	0	1	0	0	1	1	0	0	0
[14]	93.6	20	20	3	0	0	1	0	1	0	0	0	0
[12]	81.71	5	5	14	0	1	1	1	0	1	0	0	0
[17]	95.42	8	1	60	0	1	0	0	0	0	0	0	0
[13]	75 &98	16 & 7	20 & 7	50 & 10	0	0	0	0	0	0	0	0	1
[25]	87.21	4	1	25	0	1	0	0	1	1	0	0	0
[24]	99.59	10	7	1	0	1	0	0	0	0	0	1	0
[39]	97.9	10	4	2.38	0	0	1	0	0	1	0	0	0
[37]	100	8	1	2	0	0	0	0	1	0	0	0	0
[32]	96.21	40	1	7.6	0	1	1		0	1	0	1	0
[34]	91.9	14	1	21.07	0	1	0	0	1	0	0	0	0
[29]	not reported	5	1	1	0	0	0	0	0	0	0	0	0
[42]	92.3	120	75	15	1	1	0	1	0	0	0	1	0

Descriptions: **Accuracy**-Recognition accuracy of the algorithm in %, **Class**-Number of classes considered, **Subj.**-Number of subjects in the test set, **Samp.**-Number of test samples per class per subject, **UI**-User Independence, is the algorithm tested using different subjects than used for training, **Spot** -Whether algorithm can spot gestures, **BG**-Complex or simple background, 1for complex, **Noise**-Presence of other human in the background, **Scale**-Variation in scale/ size considered or not, **Light**-Variation in lighting considered or not, **Exten.**-Online or offline learning, 1 for online, **CV**-Cross validation or not, **Data**-Public or private dataset, 1 for public

is determined only by time independent transition probabilities. Marcel *et al.* [15] proposed an extension of HMM, namely Input / Output Hidden Markov Model (IOHMM), for HGR. IOHMM is based on a non-homogeneous Markov chain in which emission and transition probabilities depend on the input. The IOHMM learns to map the input sequences, observations, output sequences, and the gesture classes for all the observations using a supervised discriminant learning. Compared to HMMs, IOHMM is a discriminative approach as it directly models posterior probabilities. The study in [15] was limited to 2 classes. Just *et al.* [13] extended the study for the recognition of single and double handed gestures and provided a comparison of HMM and IOHMM. Experiments conducted on larger databases, ranging from 7 to 16 gesture classes, concluded that HMM has better performance than IOHMM for large number of classes.

Hand location, angle and velocity features are combined in [11] to implement an HMM for HGR. Hand is localized by skin-color analysis and tracked by connecting the centroid of moving hand regions. The paper compared the utility of the three features, location, angle, and velocity, and concluded that angular features are most effective, having better discriminative power. Location and velocity features are ranked second and third respectively. A similar HMM implementation utilizing angles of motion along the trajectory of hand centroid is provided in [16].

Ramamoorthy *et al.* developed an HGR system by combining HMM based temporal characterization scheme with a static shape recognition system [12]. They used a Kalman filter based hand contour tracker which provides temporal characteristics of the gesture. Shapes are recognized using contour discriminant based classifier. These symbolic descriptors of the gestures are utilized for training the HMM. The system can reliably recognize dynamic gestures in spite of motion and discrete changes in hand poses. Also the algorithm has the ability to detect the start and end points of gesture sequences.

Dynamic gesture recognition algorithms utilize a backward spotting scheme that first detects the end point of a gesture and then trace back to the start point. Kim *et al.* [17] proposed an alternate method, a forward spotting scheme, that executes gesture segmentation and recognition simultaneously. The start and end points of gestures are detected by zero crossings of differential probability of the signal. A set of 3D articulation based features are extracted by an association mapping technique that correlates the 2D shape data to the 3D articulation data. Gestures are classified by a majority voting using an accumulative HMM.

Davis and Shah [31] decomposed gestures into four distinct phases which occurs in a fixed order, and developed a Finite State Machine (FSM) model for recognition. Temporal signature of hand motion is extracted and hand gesture are modeled using an FSM in [29]. The concept of motion energy is used to es-

timate the dominant motion from an image sequence. Hong *et al.* [30] used 2D positions of the centers of subjects' head and hands to develop the FSM. A dynamic Bayesian network model is proposed in [24] for the recognition of isolated as well as continuous handed gestures. The features utilized are direction codes for hand motion, positional relation between the two hands, and the positional relation between face and hands.

Chen *et al.* [25] proposed a two level approach of statistical and syntactic analysis for the recognition of static and dynamic hand gestures respectively. The first level, statistical analysis, is based on Haar-like features and AdaBoost learning algorithm. The second level, syntactic analysis, is based on a stochastic context-free grammar (SCFG). The Haar-like features effectively describe the hand posture pattern and the AdaBoost algorithm constructs a strong classifier by combining a sequence of weak classifiers. The postures detected by the first level are converted to a sequence of terminal strings according to the grammar, in the second stage.

2.1.2. ANN and other Learning based Methods

Yang *et al.* [32, 33] utilized a time delay neural network (TDNN) to learn the 2D motion trajectories. TDNN is a multi-layer feed-forward network that utilizes shift windows between all layers to represent temporal relationships between events. The classification in TDNN is dynamic as the network sees only a small window of the input motion pattern, and the window slides over the input data while the network makes a series of local decisions. These local decisions are temporally integrated into a global decision at the output layer.

The region based motion algorithms as in [32] outperform intensity-based methods. For example, motion information in areas with little intensity variation is contained in the contours of the associated regions. The motion segmentation algorithm computes correspondences for such regions and finds the best affine transformation that accounts for the change in contour shape. The affine transformation parameters for region at different scales are used to derive a single motion field, which is then segmented to identify moving regions between two frames.

Chan *et al.* [34] proposed a combination of HMM and recurrent neural networks (RNN) which provided better performance compared to HMM or RNN used alone. The shape features used are based on Fourier descriptors, which are the inputs to radial basis function (RBF) network for an initial pose classification. The pose likelihood vector from the RBF network along with the motion information is the input to two independent classifiers, HMM and RNN. Outputs from the classifiers are combined linearly for the prediction of the gesture class.

Shen *et al.* [35] proposed an exemplar-based approach for gesture recognition. Hand gestures are represented using the divergence field of the hand flow mo-

tions. The divergence fields of the optical flow between consecutive image frames are derived and salient regions are detected from the divergence field using a Maximally Stable Extremal Regions (MSER) feature detector. Descriptors are extracted from each detected region to characterize local motion patterns. The database gesture sequences with their descriptors are indexed by a pre-trained hierarchical vocabulary. A new gesture sequence is recognized by matching it against the database.

2.1.3. Eigenspace Based Method

Patwardhan and Roy [37] proposed an eigenspace based framework to model dynamic hand gestures containing both shape and trajectory information. Feature based methods involve a separate time consuming feature detection step which is avoided in this algorithm. The algorithm is invariant to common hand shape deformations: rotation, translation, scale and shear.

2.1.4. Curve Fitting

Shin *et al.* [39] proposed a geometric method using Bezier curves for the trajectory analysis and classification of dynamic gestures. Gestures are recognized by fitting the curve to 3D motion trajectory of hand. The gesture speed is incorporated into the algorithm to enable accurate recognition from trajectories having variations in speed.

2.1.5. Dynamic Programming and Dynamic Time Warping

Kuremoto *et al.* [40] proposed a one-pass dynamic programming based approach for gesture recognition. A biologically motivated feature extraction system based on retina-V1 model proposed by Tohyama and Fukushima [43] estimates the hand motion. Hand gestures are considered as combinations of templates of simple movements. The movements are used to compose a set of 40 templates of gestures.

Dynamic time warping (DTW), an application of dynamic programming, has been widely used in isolated gesture recognition. Andrea Corradini [41] proposed a template based approach with DTW for the time alignment and normalization by computing a temporal transformation between the two signals to be matched. Lichtenauer *et al.* [42] proposed Statistical DTW (SDTW) for time warping and two classifiers, namely combined discriminative feature detectors (CDFDs) and quadratic classification on discriminative features fisher mapping (Q-DFFM), for classification. The classifiers are shown to outperform HMM and SDTW.

A summary and comparison of the features of hand gesture recognition algorithms surveyed in this section are provided in Tables 1 and 2.

2.2. Recognition of Hand Postures

The hand posture recognition methods reviewed are classified as *a)* Supervised learning based methods [35, 44–60], *b)* Unsupervised learning based methods [61], *c)* Graph matching [62–67], and *d)* 3D model based methods [68–72] (Figure 4).

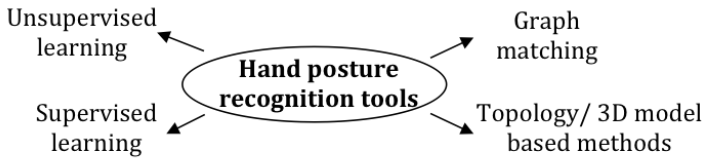

Figure 4: Taxonomy of hand posture recognition techniques reviewed.

2.2.1. Unsupervised Learning

A distributed locally linear embedding (DLLE) algorithm is proposed in [61] for hand posture recognition and dynamic gesture tracking. Locally linearly embedding (LLE) [73] is an unsupervised learning algorithm that attempts to map high-dimensional data to low-dimensional space while preserving the neighborhood relationship. The paper modified LLE to DLLE to discover the inherent properties of the input data, by noticing that some relevant pieces of information are distributed. DLLE extracts the intrinsic structure of data such as neighborhood relationship. The distances between projected data points in the low-dimensional space depend on the similarity of the input images. A probabilistic neural network (PNN) is used to classify different postures based on the distances in the low dimensional space. PNN has good training speed and classification accuracy with negligible retraining time.

2.2.2. Supervised Learning

Supervised learning in LLE algorithm is introduced in [52], for recognizing postures in Chinese sign language (CSL). Supervised LLE (SLLE) makes use of the class label information during the classifier training. Hand is detected using skin color and the intrinsic geometry of hand is used for the recognition.

Zhao *et al.* [51] proposed recursive induction learning based on extended variable-valued logic for hand pose recognition. In inductive learning knowledge is acquired by inducing rules from sets of examples or sets of feature vectors. The paper modified and extended the old concept of Variable-Valued Logic into Extended Variable-valued Logic (EVL) which provided a more powerful representation. A heuristic algorithm namely RIEVL (Rule Induction by Extended Variable-valued Logic) is proposed to learn rules both from examples as well as

rule sets. RIEVL produced more compact rules than other induction algorithms. This capability allows to apply a large feature set to hand poses during training, and to derive a reduced rule set with a subset of the training features during recognition. The algorithm automatically selects the most effective features, which makes it suitable for realtime gesture recognition systems.

Table 3: Hand posture recognition methods: Features used, classification methods, and reported applications

Work	**Features**	**Classification method**	**Application**
[61]	geometric distance	DLLE / PNN	manipulation of objects in windows user interface
[52]	intrinsic geometry of hand	SLLE	Chinese sign language
[51]	multivalued features (centroid, compactness, area of hand)	RIL	gesture commands
[50]	discrete Fourier transform based distance metric	nearest neighbour / maximum likelihood	gesture commands
[44]	shape, texture and color features	SVM	recognition against complex backgrounds
[63]	Gabor jets	EGM	HRI
[69]	joint angles	3D model fitting	not specified
[45]	shape and texture features	Fuzzy-Rough classifier	HRI
[62]	shape features	EGM	not specified
[58]	Gabor features	SVM	recognition under varying illumination
[67]	Histogram of Oriented Gradient	EGM	not specified

Descriptions: **PNN**-probabilistic neural network, **DLLE**-distributed locally linear embedding, **SLLE** -supervised locally linear embedding, **RIL**-recursive induction learning, **SVM**-support vector machines, **EGM**-elastic graph matching, **HRI**-human robot interaction

A common problem of training based methods is their dependence on training data. In order to increase generality and user independence, Licsar and Sziranyi [49, 50] proposed a user-adaptive hand posture recognition system with interactive online training. The system is retrained online for faulty detected postures if the recognition accuracy decreases, realizing fast adaptation to new users. A supervised training method corrects for the unrecognized posture classes, and an unsupervised method continuously runs to follow slight changes in posture styles.

A solution to the complex background problem in hand posture detection and recognition is provided in [44]. The algorithm can handle backgrounds including skin-colored complex backgrounds. The system utilizes a Bayesian model of visual attention to generate a saliency map, and to detect, identify, and segment out the hand region from the complex backgrounds. Feature based visual attention is implemented using a combination of high level (shape, texture) and low

Table 4: Hand posture recognition methods: Features of the algorithms and experimental methodology adopted in algorithm testing (list at bottom provides description of column titles). Features in column 6 onwards are binary, 1 represents compliance of the work to the feature whereas 0 represents non-compliance.

Work	**Accuracy**	**Class**	**Subj.**	**Samp.**	**UI**	**Spot**	**BG**	**Noise**	**Scale**	**Light**	**Exten.**	**CV**	**Data**
[61]	93.2	14	1	20	0	1	0	0	1	1	0	0	0
[52]	90.6	30	1	55	0	1	0	0	0	1	0	0	0
[51]	94.4	20	1	45.4	0	1	0	0	0	0	0	0	0
[50]	98.5	9	4	44.44	1	0	1	0	0	0	1	0	0
[44]	94.36	10	40	5	1	0	1	1	1	0	0	1	0
[63]	85.8	12	19	1.48	1	0	1	0	0	0	0	1	1
[69]	not reported	4	1	1	0	0	0	0	0	0	0	0	0
[45]	98.75	10	19	18	1	0	0	0	1	1	0	1	1
[62]	96.35	10	19	2.52	1	0	0	0	0	0	0	0	1
[58]	96.1	11	10	6	0	0	1	0	1	1	0	0	0
[67]	99.85	10	24	2.7	1	0	1	0	0	0	0	1	1

Descriptions: **Accuracy**-Recognition accuracy of the algorithm in %, **Class**-Number of classes considered, **Subj.**-Number of subjects in the test set, **Samp.**-Number of test samples per class per subject, **UI**-User Independence, is the algorithm tested using different subjects than used for training, **Spot** -Whether algorithm can spot gestures, **BG**-Complex or simple background, 1for complex, **Noise**-Presence of other human in the background, **Scale**-Variation in scale/ size considered or not,**Light**-Variation in lighting considered or not, **Exten.**-Online or offline learning, 1 for online, **CV**-Cross validation or not, **Data**-Public or private dataset, 1 for public

level (color) image features. The segmented hand postures are classified using the shape and texture features, with a support vector machines (SVM) classifier.

Huang *et al.* [58] proposed an algorithm for hand posture recognition under varying illumination and pose conditions. The invariance to lighting conditions is achieved using an adaptive skin color model switching method. Insensitivity to hand pose variations is gained using a Gabor filter based pose angle estimation and correction method. The posture are classified using an SVM classifier.

2.2.3. Graph Algorithms

Starting from the late seventies, graph-based techniques are used as a powerful tool for pattern representation and classification. After the initial enthusiasm, graph algorithms have been practically left unused for a long period of time. This is due to the high computational cost of graph algorithms, which still remains an unresolved problem. However, the use of graphs in computer vision and pattern recognition obtained a growing attention from the research community recently, as the computational cost of the graph-based algorithms is now becoming compatible with the computational power of new generation computers [74].

Elastic graph matching (EGM), a type of graph matching, is a neurally inspired pattern recognition architecture [75]. EGM has the inherent ability to handle geometric distortions, does not require a perfectly segmented input image, and can elegantly represent the variances in object appearance [63].

Image regions are represented by *vertices* in a graph representation. These vertices are related to each other by *edges*, expressing structural relationships between regions. Triesch *et al.* [63–66] utilized the elastic graph matching (EGM) technique to develop a system for person independent hand posture recognition against complex backgrounds. Hand postures are represented by labeled graphs with an underlying two dimensional topology. Attached to the nodes are *jets*, a local image description (image feature) based on Gabor filters. This approach provided scale invariant and user independent recognition, without explicit segmentation of hand region. Different hand postures are represented as attributed graphs and comparisons are made between model graphs (in the database) and data graph (corresponding to the realtime image). The nodes are compared using a similarity function, and the pattern is recognized by calculating the average node similarities.

Bunch graphs [76] are used to model the variability in object appearance. The natural variability in the attributes of corresponding points in several images (of the same object or a class of objects) is captured by labeling each node with a *bunch* of attribute values, extracted from the corresponding points. This method is used by Triesch *et al.* [63, 64] to model complex background in hand posture images. For the matching process, each of the attribute value in the bunch is compared with the local image information in the data graph, and the maximum

of the similarities is taken as the similarity of the bunch graph.

Li and Wachs [67] proposed a hierarchical EGM algorithm for hand gesture recognition. The major improvement to the EGM algorithm is the use of levels of hierarchies assigned to the nodes. The visual features with higher likelihood (to be found on the target image) receive a higher hierarchy level compared to features those are less consistent with the graph model.

2.2.4. Topology / 3D model Based Methods

Three dimensional model fitting is used in [69] for hand pose estimation. The method estimates all joint angles reconstructing the hand pose as a voxel model. Then model fitting is done between the hand model and the voxel model, in the 3D space. The method uses only geometric information of hand model and the voxel model for model fitting and does not need any heuristic or priori information. However the algorithm requires faster implementation for realtime applications.

Yin and Xie [70] introduced a computer vision model of hand, instead of a kinematic model. The algorithm avoids the complexity in estimation of the angular and linear parameters of the kinematic model. They utilized topological features of the hand for 3D hand posture recognition. The edge point of fingers are extracted as points of interest. The hand is segmented from complex backgrounds using a restricted coulomb energy (RCE) neural network based on color segmentation.

A summary and comparison of the features of hand posture recognition algorithms reviewed in this section are provided in Tables 3 and 4.

3. Recent Trends in Hand Gesture Recognition: RGB-D Sensor Based Methods

Depth cameras have been used in computer vision for several years. However the applicability of depth cameras was limited due to its high price and poor quality. The release of low cost color-depth (RGB-D) camera Kinect [77, 78] by Microsoft has created a revolution in gesture recognition by providing high quality depth images, addressing issues like complex backgrounds and illumination variation. The device calculates a three dimensional map of the scene using a combination of RGB and IR camera. Recently Han *et al.* [79] provided a review of how Kinect is useful in addressing the fundamental problems in computer vision. The sensors such as $Microsoft Kinect^{(R)}$ and $ASUS\ Xtion\ PRO\ LIVE^{(R)}$ provide reliable tracking of human body postures in gaming scenarios. Based on the tracking these devices provide features such as the coordinates of a skeletal model, which are utilized for gesture recognition.

The skeletal data from these RGB-D sensors is to be converted to more meaningful and high level features, and algorithms are to be developed for the robust classification of gestures. Recognition of hand gestures is especially challenging due to the complex articulation and relatively smaller area of hand region. In addition, a robust hand gesture recognition algorithm must have invariance with respect to the size and speed of the gesture, and the orientation of gesturer. Rafael *et al.* [80] evaluated the influence of depth information in the gesture recognition process and concluded that use of depth silhouettes increases the recognition accuracy significantly. Dominio *et al.* [81] proposed an algorithm to combine multiple depth-based descriptors for hand gesture recognition.

The RGB-D cameras are mostly used for whole body gesture recognition [78, 82–85], as these cameras provide skeletal tracking. This section of the paper surveys RGB-D camera based HGR algorithms[1] by classifying the related literature into two categories, *a)* Kinect based approaches, and *b)* Other RGB-D sensor based approaches.

Table 5: RGB-D sensor based methods: A comparison

Work	**S/D**	**Sensor**	**Features**	**Classification method**
[86]	D	CSEM Swissranger SR-2	motion primitives	probabilistic edit distance classifier
[87]	S	ToF and RGB camera	Haarlets	NeN
[88]	S	Kinect	hand/finger shape	template matching using FEMD
[89]	D	Kinect	hand area	classifier based on topology
[90]	D	Kinect	Extended-Motion-History-Image	maximum correlation coefficient
[91]	S	Kinect	depth pixel values	randomized classification forests & voting
[92]	D	Kinect	underlying geometry	least squares fitting
[93]	D	Kinect	Euclidean and log-Euclidean distance	NeN
[94]	D	PrimeSense 3-D camera	probabilistic 2D templates from trajectory	MPLCS classifier
[95]	D	Kinect	spatial and motion features	conditional density propagation
[96]	D	Kinect	position, angle, and direction features	probability; pairwise coupling
[97]	D	Kinect	conditional distance	dynamic time warping

S-static, **D**-dynamic, **ToF**-time of flight, **NeN**-nearest neighbor, **FEMD**-finger earth mover's distance, **MPLCS**-most probable longest common subsequence

[1]Many of the articles in this area are published conference proceedings reporting developmental work using Kinect sensor. The current survey is limited to selected relevant research articles.

Table 6: RGB-D sensor based methods: Features of the algorithms and experimental methodology adopted in algorithm testing (list at bottom provides description of column titles). Features in column 6 onwards are binary, 1 represents compliance of the work to the feature whereas 0 represents non-compliance.

Work	**Accuracy**	**Class**	**Subj.**	**Samp.**	**UI**	**Spot**	**BG**	**Noise**	**Scale**	**Light**	**Exten.**	**CV**	**Data**
[86]	92.9	4	10	3	0	1	0	0	0	0	0	0	0
[87]	99.54	6	1	29.17	0	1	1	1	0	0	0	0	0
[88]	93.9	10	10	1	0	1	1	0	1	1	0	0	0
[89]	not reported	9	1	1	0	1	0	0	0	0	0	0	0
[90]	not reported	8-15	multiple (ChaLearn)	multiple (ChaLearn)	1	1	1	1	1	1	0	0	1
[91]	84.3 & 74.3	24 & 9	4 & 5	100 & 10	1	0	0	0	0	1	0	1	1
[92]	91.7	9	2	80	1	0	0	0	0	1	0	1	1
[93]	99.75	8	20	5	1	0	0	0	1	0	0	1	0
[94]	98.7	10	8	5	1	1	0	0	1	0	0	1	0
[95]	95.9	4	4	1	1	1	1	0	0	0	0	1	0
[96]	97.26	10	6	1	1	1	1	0	1	0	0	1	0
[97]	82	179	18	multiple (ChaLearn)	1	0	1	1	1	1	0	1	1

Descriptions: **Accuracy**-Recognition accuracy of the algorithm in %, **Class**-Number of classes considered, **Subj.**-Number of subjects in the test set, **Samp.**-Number of test samples per class per subject, **UI**-User Independence, is the algorithm tested using different subjects than used for training, **Spot** -Whether algorithm can spot gestures, **BG**-Complex or simple background, 1for complex, **Noise**-Presence of other human in the background, **Exten.**-Online or offline learning, 1 for online, **Scale**-Variation in scale/ size considered or not, **Light**-Variation in lighting considered or not, **CV**-Cross validation or not, **Data**-Public or private dataset, 1 for public

3.1. Kinect based Methods

Zhang *et al.* [98] proposed a new higher level descriptor called the Histogram of 3D Facets (H3DF), to explicitly encode the 3D shape information from lower level depth information. Kinect based features are utilized for both dynamic hand gesture recognition [89–93, 95–97, 99–111] and hand posture recognition [88, 112–120].

3.1.1. Recognition of Dynamic Hand Gestures

Wu *et al.* [90] proposed a system to learn gestures from only one learning example per class, namely *One-shot-learning*. Features are extracted based on Extended-Motion-History-Image (Extended-MHI) and the gestures are classified by calculating the maximum correlation coefficient. Motion history images (MHI) [121] are used to represent motions of an object in a video. All frames in a video sequence are projected onto one image across the temporal axis, to capture the temporal information of the motion sequence. The extended-MHI is proposed to improve the performance of MHI by compensating on the non-moving regions and repetitive actions. Multi-view Spectral Embedding (MSE) algorithm is used to fuse the RGB and depth data in a physically meaningful manner. The MSE algorithm discovers the intrinsic relationship between RGB and depth features, improving on the recognition rate of the algorithm.

Lui [92, 99] proposed a gesture recognition algorithm based on a nonlinear regression framework on manifolds. The underlying geometry and a least squares fitting is used to develop the algorithm. The least squares regression is formulated as a composite function, considering geometric properties. Gallo *et al.* [89] proposed a Kinect based gesture recognition system with its application to exploration of medical image data. Various gestures for functions like zooming, animation, region of interest extraction, rotation and translation of medical images are recognized by topological analysis of the hand region. Euclidean distance metric and covariances of a log-Euclidean metric are used as features in [93]. The gestures are classified using nearest neighbor classifier.

A novel one-shot-learning approach for gesture recognition from motion depth images based on template matching is presented in [100]. The method is based on the computation of space-time descriptors from the query video which measures the likeness of a gesture in a lexicon. The classifier is based on correlation coefficient from standard deviation of Fourier transform of the image and the MHI.

An algorithm for detection and recognition of hand gestures by combining DTW with probability estimates is proposed in [102]. The algorithm has robustness against position and orientation of the gesturer and speed of the gesture. Cheng *et al.* [103, 104] proposed DTW based algorithms for 3D hand gesture recognition. A parameterized searching window is introduced in the cost matrix

of traditional DTW approach to detect the beginning and end of specific gestures from an infinite trajectory gesture sequences.

Another algorithm for one-shot learning gesture recognition from RGB-D data is proposed by Wan *et al.* [101]. A new spatio-temporal feature representation called 3D enhanced motion scale-invariant feature transform (3D EMoSIFT) is used. The new feature set is invariant to scale and rotation as it fuses RGB-D data. A sparse coding method namely simulation orthogonal matching pursuit (SOMP) is applied to represent each feature by a linear combination of a small number of codewords.

3.1.2. Recognition of Hand Postures

A novel hand motion capture procedure based on 14-patch hand partition scheme is proposed in [117] for collecting real posture dataset in unconstrained conditions. Liang *et al.* [122] proposed a robust hand parsing scheme to extract a high-level description of the hand from the depth images. The method is robust to complex hand configurations.

Ren *et al.* [88, 112] proposed a hand posture recognition system having robustness against variations in hand orientation, scale, and articulation. A distance metric called Finger-Earth Mover's Distance (FEMD) is proposed for hand dissimilarity measure. The algorithm can recognize hand postures in spite of the variations, as it only matches the fingers (not the whole hand shape). A comparison of FEMD with *shape context* algorithm [123] is provided in [113]. The FEMD based algorithm has better accuracy and computational speed in comparison. In addition [113] presented an application of FEMD based hand posture recognition algorithm for playing Sudoku game.

An algorithm for static and dynamic hand shape classification using randomized decision forests is proposed in [91]. The hand shape is classified using the data from depth sensors. The system performs independent of lighting conditions and it does not need a hand registration step. Class labels are assigned to each pixel on a depth image, and the final class label is determined by voting.

Kirac *et al.* [116] proposed a scheme for extracting the hand skeleton using random regression forests in realtime. The algorithm is robust to self occlusion and low resolution of the depth camera, and can estimate the joint positions even if all of the pixels related to a joint are out of the camera frame. A feature set namely Oriented Radial Distribution is proposed in [118], which can simultaneously localize fingertips and encode hand postures globally.

3.2. Other RGB-D Sensor based Methods

3.2.1. Recognition of Dynamic Hand Gestures

Holte *et al.* [86] utilized an intensity-depth camera (CSEM Swissranger SR-2) to develop a view invariant gesture recognition algorithm. On contrary to

the usual trajectory based approach gestures are recognized based on motion primitives in the 3D data. The primitives are represented in a view invariant manner using harmonic shape context. A probabilistic edit distance classifier is used for classification. The algorithm has orientation invariance, it is trained on data from one viewpoint and tested on data from a different viewpoint.

Probabilistic 2D templates created using hand motion trajectory are used in [94] for the recognition of dynamic gestures. The probabilistic template takes into account different trajectory distortions with different probabilities. A longest common subsequence (LCS) classifier is modified to most probable longest common subsequence (MPLCS) classifier, to measure the similarity between the probabilistic template and the hand gesture sample. Erden *et al.* [124] designed a hand gesture based remote control system which combines infrared sensors with an RGB camera.

3.2.2. Recognition of Hand Postures

Time-of-Flight (ToF) and RGB cameras are combined in [87] to develop a hand detection algorithm based on depth and color. The position of hand is tracked in 3D in spite of its overlap with body parts and other hands in the background. The gestures are recognized using a nearest neighbor search after a dimensionality reduction using Average Neighborhood Margin Maximization (ANMM) [125].

A summary and comparison of the features of hand gesture and posture recognition algorithms surveyed in this section are provided in Tables 5 and 6.

4. Hand Gesture Databases

Researchers from University of Cambridge and Microsoft Research have conducted a study [126] on how to instruct subjects to develop best representative gesture datasets for training machine learning algorithms. They used two measures, *correctness* and *coverage*, to evaluate how good the dataset is in representing real world data from a deployed system. The measure correctness refers to the similarity of subject movements to what the system developer needs them to perform. It depends on the *understanding* by the subject. The measure coverage refers to completeness of the dataset in representing natural and possible variations of associated movement patterns. Coverage is decided by the *freedom* given to the subject. They investigated the most appropriate semiotic modality of instructions and their order to achieve the best correctness and coverage, both for the dataset and the learnt gesture recognition system. The modalities investigated include descriptive text, static image sequence, and video. Video followed by text is selected as the best order of modality to facilitate both understanding and freedom of subjects.

Standard hand gesture databases are necessary for the reliable testing and comparison of hand gesture recognition algorithms. The availability of hand gesture databases was limited till the year 2007 and has been increased recently (Figure 6). This section provides a review of publicly available hand gesture datasets. Table 7 lists hand posture and gesture databases with the web-links for their download. Table 8 describes these datasets with details such as number of classes, subjects, and samples available. The works utilized the datasets are also included to facilitate possible comparative study. A total of 25 datasets are available at the publication time of this review.

4.1. Sebastien-Marcel Hand Posture and Gesture Datasets

The dataset contains three hand posture datasets, the Jochen Triesch Static Hand Posture Database [64], the Jochen Triesch Static Hand Posture Database II [63], and the Sebastien Marcel Static Hand Posture Database [133], and one dynamic hand gesture database, the Sebastien Marcel Dynamic Hand Posture Database [15]. The hand posture datasets have simple as well as complex backgrounds. The dynamic gestures include various commanding signals for *Click*, *Stop-grasp-ok*, *Rotate*, and *No*.

4.2. Cambridge Hand Gesture Dataset

This dataset contains hand posture images. It has sequences of static images corresponding to hand motions, making it suitable for testing dynamic hand gesture recognition algorithms [137]. The data set consists gestures defined by 3 primitive hand shapes (*flat*, *spread*, and *V-shape*) and 3 primitive motions (*leftward*, *rightward*, and *contract*). The target task for this data set is to classify hand shapes and motions at the same time. The dataset has fairly large intra-class variations in spatial and temporal alignment of hand gestures.

4.3. Gesture Dataset by Shen et al.

The database is useful in testing both hand gesture and posture recognition algorithms, as it contains both movement patterns and specific hand shapes [35]. It has 10 classes of dynamic hand gestures (eg. *move right*, *move left*, *rotate up*) performed with 7 different hand poses (eg. *thumb*, *fist*, *all fingers extended*), summing to 70 gesture samples per subject.

4.4. NATOPS Aircraft Handling Signals Database

The database includes 24 body and hand gestures, selected from NATOPS (Naval Air Training and Operating Procedures Standardization) aircraft handling signals [132]. A stereo camera was used to collect the database. The database consists videos with RGB and depth data. It also contains the extracted body and hand feature sets in Matlab and CSV formats.

Table 7: Publicly available hand gesture databases and their sources. See Table 8 for descriptions.

No.	Name, Year	Source
1	ChaLearn gesture* data**, 2011	http://gesture.chalearn.org/data
2	MSRC-12 Kinect gesture* dataset**, 2012	http://research.microsoft.com/en-us/um/cambridge/projects/msrc12/
3	ChaLearn multi-modal gesture data**, 2013	http://sunai.uoc.edu/chalearn/
4	NUS hand posture dataset-II, 2012	http://www.ece.nus.edu.sg/stfpage/elepv/NUS-HandSet/
5	6D motion gesture database*, 2011	http://www.ece.gatech.edu/6DMG/6DMG.html
6	Sebastien Marcel interact play database, 2004	http://www.idiap.ch/resource/interactplay/
7	NATOPS aircraft handling signals database*, 2011	http://groups.csail.mit.edu/mug/natops/
8	Sebastien Marcel hand posture and gesture datasets, 2001	http://www.idiap.ch/resource/gestures/
9	Gesture dataset by Shen *et al.*, 2012	`http://users.eecs.northwestern.edu/~xsh835/GestureDataset.zip`
10	Gesture dataset by Yoon *et al.*, 2001	available on e-mail request to yoonhs@etri.re.kr
11	ChAirGest multi-modal dataset**, 2013	https://project.eia-fr.ch/chairgest/Pages/Download.aspx
12	Sheffield KInect Gesture Dataset**, 2013	http://lshao.staff.shef.ac.uk/data/ SheffieldKinectGesture.htm
13	Keck gesture dataset, 2009	`http://www.umiacs.umd.edu/~zhuolin/Keckgesturedataset.html`
14	NUS hand posture dataset-I, 2010	http://www.ece.nus.edu.sg/stfpage/elepv/NUS-HandSet/
15	Cambridge hand gesture data set, 2007	`http://www.iis.ee.ic.ac.uk/~tkkim/ges_db.htm`
16	Posture dataset by Ren *et al.***, 2011	`http://eeeweba.ntu.edu.sg/computervision/people/home/renzhou/HandGesture.htm`
17	ColorTip dataset**, 2013	`https://imatge.upc.edu/web/res/colortip`
18	NYU Hand Pose Dataset**, 2014	`http://cims.nyu.edu/~tompson/NYU_Hand_Pose_Dataset.htm#overview`
19	General-HANDS dataset**, 2014	http://wildhog.ics.uci.edu:9090
20	VPU Hand Gesture dataset (HGds), 2008	http://www-vpu.eps.uam.es/DS/HGds/
21	Dataset by Kawulok *et al.*, 2014	`http://sun.aei.polsl.pl/~mkawulok/gestures/`
22	ASL Finger Spelling Dataset**, 2011	`http://personal.ee.surrey.ac.uk/Personal/N.Pugeault/index.php?section=FingerSpellingDataset`

*All the gestures in this dataset are not hand gestures. Some are body gestures.
**These are RGB-D sensor based datasets, containing depth/ skeletal information.

Table 8: Description of publicly available hand gesture databases (in same order as in Table 7)

No.	Description	S/D	Works
1	ChaLearn Gesture Challenge, 62,000 samples	D	[82, 90–92, 100, 127]
2	12 classes, 30 subjects, 6,244 samples	D	[126]
3	20 classes, 27 subjects, 13,858 samples	D	[128]
4	10 classes, 40 subjects, 2,750 samples, complex background	S	[44, 129]
5	20 classes, 28 subjects, 5,600 samples	D	[130]
6	16 classes, 22 subjects, 50 samples/ subject	D	[13, 131]
7	24 classes, 20 subjects, 9,600 samples	S & D	[132]
8	Three hand posture datasets, with 10 (gray scale), 12 (color), and 6 (gray scale) classes. One hand gesture dataset with 4 classes	S & D	[63–65, 133]
9	10 classes, 15 subjects, 1,050 samples	S & D	[35]
10	48 classes, 20 subjects, 9,600 samples	D	[11]
11	10 classes, 10 subjects, 1,200 samples recorded with Kinect and inertial motion units	D	[134]
12	10 classes, 6 subjects, 2,160 samples recorded with Kinect and RGB cameras	D	[135]
13	14 classes, 3 subjects, 126 training and 168 testing samples	D	[136]
14	10 classes, 1 subject, 240 samples, color as well as grey scale	S	[45]
15	9 classes, 2 subjects, 900 image sequences, with different illumination conditions	S & D	[137]
16	10 classes, 10 subjects, 1000 samples, color as well as depth maps, cluttered background	S	[88]
17	7 subjects, 9 classes, 7 training sequences of between 600-2000 depth frames	S	[118]
18	2 users, data from 3 Kinects (frontal and 2 sides), 72757 and 8252 frames in training and test sets	S	[138]
19	22 sequences, different view-points, scales, poses, and occlusions	S	-
20	12 classes, 11 subjects, 1 video per gesture (252 frames)	S	[139]
21	32 classes, 18 subjects, gestures from Polish Sign Language and American Sign Language (ASL)	S	[140]
22	24 classes, 9 subjects, 65,000 samples	S	[141]

S-static, **D**-dynamic

4.5. Gesture Dataset by Yoon et al.

This dataset contains 48 class alphabetical gestures (alphanumeric characters & graphic elements) recorded from 20 persons, 10 times each gesture [11]. The dataset contains sequences of x-y coordinates representing unspotted gestures.

4.6. Sebastien Marcel Interact Play Database

The dataset contains 3D trajectories of segmented hand gestures, including the coordinates of head and torso [13, 131]. Each trajectory is stored as a text file in the dataset. The dataset has both single handed (like *stop*, *point left*, *point right*) and two handed (like *swim*, *fly*, *clap*) gestures. Gesture trajectories contain 3D coordinates of center of the head, two hands and the torso.

4.7. Keck Gesture Dataset

The gesture dataset consists of 14 dynamic gestures, which are subsets of military signals (like *turn left*, *go back*, and *speed up*) [136]. The dataset is divided into two, training and testing sets. Training set is captured using a fixed camera with the person viewed against a simple and static background. Testing set is captured from a moving camera, in the presence of background clutter and other moving objects.

4.8. 6D Motion Gesture Database

The 6D Motion Gesture Database (6DMG) provides a comprehensive data of motion gestures, including the position, orientation, acceleration, and angular speed [130]. The data is stored in raw binary form and the dataset comes with sample C++ programs to access and visualize the data.

4.9. ChaLearn Gesture Data

This dataset is created as part of a gesture recognition challenge; the *ChaLearn gesture challenge* [82, 90–92, 100, 127]. The ChaLearn gesture data 2011 consists a total of 62,000 samples. The dataset from 20 subjects is grouped into different batches each with 100 samples. The data is recorded with Kinect camera and consists both RGB and depth videos of dynamic gestures. The dataset also 8,000 samples of translated, scaled and occluded data.

In comparison to other datasets, the gestures in ChaLearn gesture data are useful in wide application domains. It contains nine categories of gestures corresponding to various application domains. The categories are *a)* emblems (*e.g.* Indian Mudras), *b)* illustrators (*e.g.* Italian gestures), *c)* regulators (gesticulations performed to accompany speech), *d)* pantomimes (gestures made to mimic actions), *e)* signs (from sign languages for the deaf), *f)* signals (*e.g.* marshaling signals to guide machinery or vehicle), *g)* body language gestures (*e.g.* scratching head, crossing arms), *h)* actions (*e.g.* drinking or writing), and *i)* dance

postures. Each set of data contains a number of actions presented separately once for training purpose. Combinations of one or more actions in a video sequence are available for testing.

4.10. ChaLearn Multi-modal Gesture Data

In comparison to the ChaLearn gesture data, the testing using ChaLearn multi-modal gesture data [128] is more challenging. The ChaLearn multi-modal gesture data includes recording of continuous sequences, presence of distracter gestures, relatively large number of categories, lengthy gesture sequences, and gestures by a variety of users. Several modalities are provided in the data set, including audio, RGB, depth maps, user masks, and user skeletal model.

4.11. ChAirGest Multi-modal Dataset

This data is acquired using a Kinect camera and 4 inertial motion units attached to the right arm and the neck of subjects. Gestures are started from 3 different resting postures and recorded in 2 different lighting conditions [134].

4.12. Sheffield KInect Gesture (SKIG) Dataset

SKIG dataset [135] has 10 categories of hand gestures, recorded from 6 subjects using RGB and Kinect cameras. The dataset is recorded with 3 different backgrounds (wooden board, white paper, and paper with characters) and 2 illumination conditions (light and dark).

4.13. MSRC-12 Kinect Gesture Dataset

Microsoft Research Cambridge-12 (MSRC-12) is a 12 class dynamic gesture dataset recorded using the skeletal data from Kinect [126]. The dataset consists of sequences of human movements, represented using body-part locations (20 skeletal joints). The data set includes 594 sequences and 719,359 frames.

4.14. NUS Hand Posture Dataset-I

The postures in this dataset are captured with various position and size of the hand within the image frame. Both color and gray-scale versions of the dataset are available. The hand postures in the dataset have less inter-class variation in appearance, which makes the recognition task challenging [45].

4.15. NUS Hand Posture Dataset-II

This complex background hand posture dataset [44] has three subsets; A, B, and C. Subset A has images with complex natural backgrounds and subset B has images with noises like body/ face of the posturer or a group of other human in the background. Subset C consists only background images (to be used as negative images for hand posture detection). The postures have various hand shapes and sizes, and are collected from subjects with various ethnicities. Subset A has 2000 images, B has 750 images, and C has 2000 images.

4.16. Posture dataset by Ren et al.

This is a Kinect posture dataset [88] with 10 classes. It contains both color images and depth maps. The dataset is collected under cluttered backgrounds.

4.17. ColorTip Dataset

The hand gesture annotation in this dataset [118] is done among 9 gesture classes and the dataset has strong intra-class variation. Fingertip annotation is done in the dataset using colored gloves, easing the detection and localization of fingertips.

4.18. NYU Hand Pose Dataset

The NYU hand pose dataset [138] has 72757 and 8252 numbers of frames in the training and test sets respectively. The data is captured using 3 Kinect sensors providing a frontal view and 2 side views. Training set is captured form 1 user and test set is captured from 2 users.

4.19. General-HANDS data-set

This dataset contains a variety of 22 sequences, demonstrating different viewpoints, scales, poses, occlusions, and camera technologies. The dataset is useful to evaluate hand detection and pose estimation algorithms.

4.20. VPU Hand Gesture dataset

This hand gesture dataset [139] contains 12 class data from 11 subjects. Also it contains synthetically generated data and is useful in evaluating hand posture recognition algorithms.

4.21. Dataset by Kawulok et al.

This dataset [140] contains gestures from Polish Sign Language and ASL, and is organized into three series acquired under different conditions. It has up to 32 gesture classes acquired from 18 different subjects.

4.22. ASL Finger Spelling Dataset

ASL Finger Spelling Dataset [141] consists of 24 hand postures, English letters from *a* to *y* except *j*. It contains an *easy* set, Dataset A, captured from 5 subjects without lighting variation, and a *hard* set, Dataset B, captured from 9 subjects with lighting variations.

4.23. Other related hand databases

A few other publicly available databases relevant to hand gesture recognition research are *a)* MSRA Hand Tracking database [142] (`http://research.microsoft.com/en-us/um/people/yichenw/handtracking/index.html`), *b)* American Sign Language Lexicon Video Dataset [143] (`http://www.bu.edu/asllrp/cslgr/`), and *c)* Bosphorus hand databases (`http://bosphorus.ee.boun.edu.tr/hand/Home.aspx`).

5. Discussion

The chart in Figure 5 shows the fast growth in hand gesture recognition research[2] and that in Figure 6 shows the growth in release of hand gesture databases. In spite of these developments there are still unresolved challenges in gesture recognition. This section briefly reviews some of the unresolved issues in the field, provides a comparison of different approaches, and discusses a few future research directions.

5.1. Recognition of Illustrators

The recognition of illustrators is challenging as the meaning of these gestures is depended on the context. The context reference is to be recognized in addition to the recognition of an illustrator gesture. Among various illustrators, the pointing gesture is very useful in applications like mobile robot commanding. Understanding a pointing gesture in 3D involves detection of the gesture, finding the hand position, and identification of the pointed direction. The difficultly in accurate estimation of the pointing direction makes pointing gesture recognition challenging. Heuristic such as the direction at which the subject looks is useful in recognizing a pointing gesture. For example a line joining the center of the eyes with the tip of the index finger can provide an estimate of the pointed direction, which in turn can be utilized to identify the targeted point [144, 145].

The pointing direction estimation using head-hand line is effective when pointing hand extends outwards and lies on the surface of an imaginary hemisphere centered on the shoulder [146, 147]. However this method is not effective in the case of compact pointing gestures in which a person moves only the forearm. Such *small* pointing gestures can be recognized by modeling the kinematic characteristics of forearm and pointing finger [146]. Head orientation can be utilized as a feature to improve the performance of pointing gesture recognizer [148]. In comparison, the direction estimation using head-hand line outperforms

[2]Based on relevant articles covered in this review.

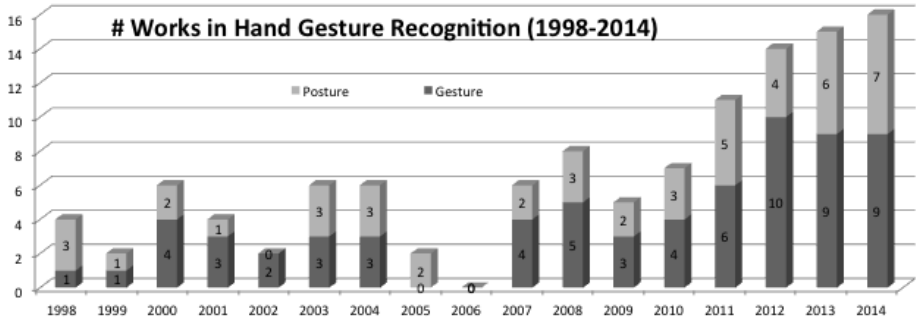

Figure 5: Chart depicting the growing research efforts in hand gesture and posture recognition.

that based on orientation of the forearm, in the case of a normal pointing gesture [148]. Raheja *et al.* [149] proposed an algorithm for hand gesture pointing location detection which is based on locations of head, shoulders and elbows. The method proposed by Pateraki *et al.* [150, 151] combined face pose and head orientation with the hand direction.

5.2. Comparison of Approaches, Features, and Classification Methods

5.2.1. Appearance and Model based Approaches

Appearance-based approaches provide better realtime performance compared to 3D hand model based approaches, as the image feature extraction process is faster. Appearance-based models lead to computationally efficient algorithms that work well under constrained situations, but lack the generality desirable for human computer interaction. Appearance based methods mainly utilizes the 2D shape data of the hand which is dependent on the viewing angle. The use of such methods is limited by the viewing perspectives. A wide class of hand gestures could be covered in 3D hand model based approaches, as the models offer a way for elaborate hand gesture modeling. However 3D models need large image database to cover all the characteristic shapes and its variations under different views. Matching the test image with all the models in the database is time consuming and computationally expensive which limits the usage of 3D models for realtime applications.

5.2.2. Features

Selectivity and invariance are two desired qualities for any image based pattern recognition process. Template based approaches provide good selectivity for shape patterns, lacking invariance. Histogram based approaches have invariance property. However histogram approaches consider the integrated image information, which makes it unsuitable for shape recognition tasks like hand posture

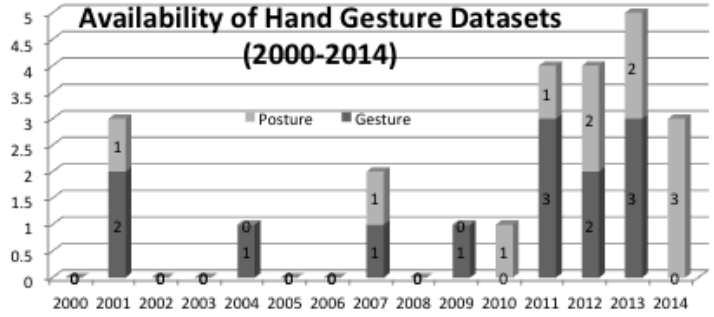

Figure 6: Chart depicting the growth in publicly available hand posture and gesture datasets.

recognition. Shape-texture patterns extracted using biologically inspired approaches [152] provide features having both selectivity and invariance, and are useful in hand posture recognition [44].

Orientation and angular features of gestures provide better invariance compared to positional features. On the other hand positional features are simple and can be extracted with better accuracy. Texture based features have the capability to capture spatial properties better in comparison to that captured by features such as color.

The RGB-D sensors enable extraction of invariant features in spite of complex backgrounds and variations in scale, lighting, and view points. The accurate depth data and position information form these sensors speed-up the extraction of hand models, increasing the utility of model based approaches.

5.2.3. Classification Methods

HMM based methods are effective and are widely used for HGR. However HMM based approaches require a large number of training samples and have the disadvantage of elaborate training procedure. The computational costs of HMM based algorithms increase with the gesture vocabulary. In addition, the performance of HMM based algorithms reduces when there are variations between training and testing conditions. Finding the optimal parameter sets and trajectory spotting for temporal segmentation are other bottlenecks in using HMM.

The design of a TDNN is attractive as its compact structure economizes on the weights, and makes it possible to develop more generic feature detectors. The hierarchy of delays in TDNN optimizes these feature detectors by increasing their scope at each layer. Temporal integration of features at the output layer makes the network shift invariant (insensitivity to exact hand position). The total number of weights in the network is relatively small since only a small

window of the input pattern is fed to TDNN at any instance. This helps to reduce the training time.

Graph based algorithms have the disadvantage of high computational complexity, which leads to its unsuitability for realtime applications. However each node in the graph can be modeled with a bunch of node features, which is useful in addressing issues due to complex backgrounds [63] and size or shape variations.

5.3. Challenges and Future Research Directions

Identification of the gesturing phase is a major challenge in HGR. The presence of unpredictable and ambiguous non-gesture hand motions makes the task challenging. Capability to reject unknown classes is one of the important requirements for an automatic gesture recognizer. The threshold model concept introduced by Lee and Kim [10] is useful for this purpose. The simultaneous gesture segmentation and recognition algorithm proposed by Kim *et al.* [17] utilized a continuous probability estimation of gestures and non-gestures to find the start/ end points. Kang [153] *et al.* proposed a recognition based gesture spotting scheme to filter out unintentional movements. Recently Yin *et al.* [154] used a concatenated HMM to perform gesture spotting in continuous data stream, attaining encouraging experimental results.

The transition movements between adjacent gestures is another related issue in automatic recognition of continuous gestures, especially in applications like sign language recognition. Yang *et al.* [155] addressed the issue of handling movement epenthesis using a dynamic programming based approach. Li *et al.* [156] proposed and compared three methods based on a gesture model for end point localization. The methods investigated are a multi-scale search, dynamic time warping, and dynamic programming. In comparison, the dynamic programming based method outperformed the other two. A nested, level-building based dynamic programming approach is proposed by Sarkar *et al.* [157] to address the uncertainties of sign boundaries in sentences.

Matching an image sequence to a model is a central issue in HGR. Yang *et al.* [158] proposed a minimization algorithm to match groups of image primitives with statistical (HMM) as well as non-statistical (sample-based) models. The algorithm neither needed a perfect segmentation of the scene nor the tracking of features across frames.

The recent trend of *One-shot-learning* [90, 97, 100, 101] in gesture recognition is promising. The one-shot-learning consists of learning a gesture by observing only one instance of that gesture, similar to the learning in human. It has created the opportunity to take-up the challenge of extraction of discriminative features as well as design of competitive classifiers using only one training example per class. Also one-shot-learning facilitates critical comparison between gesture recognition algorithms.

The hand gestures utilized in existing gesture recognition systems are limited to a carefully chosen vocabulary of symbolic gestures (*emblems* and *illustrators*), mainly used for issuing commands. Recognition of gestures from *regulators*, *affect displays*, and *adaptors* (Section 1) are necessary for the natural interaction between humans and machines. Algorithms with better invariance capabilities, having the potential to recognize a wide number of classes without extensive training, are to be developed for making machines with capability to understand human intentions and motion patterns better.

The impact of embodied interactions through gestures on enhancing visual processing and attention is least explored. For example exploring how a waving hand captures human attention will be useful for developing the attentional mechanism of an interactive robot. Another future research direction is the exploration of primate brain areas to develop computational models to imitate the gestural pattern recognition process.

References

[1] K. Hogan, R. Stubbs, Can't get Through 8 Barriers to Communication, Pelican Publishing Company, Grenta, LA, 2003.

[2] D. K. Spencer, M. M. Sarah, R. Sabrina, Gesture gives a hand to language and learning: Perspectives from cognitive neuroscience, developmental psychology and education, Language and Linguistics Compass 2 (4) (2008) 569–588.

[3] A. Kendon, Current issues in the study of gesture, The Biological Foundation of Gestures: Motor and Semiotic Aspects (1986) 23–47.

[4] L. L. B. Malandro, L. A., A. B. Deborah, Nonverbal Communication, 2nd ed., Addison-Wesley, MA, 1989.

[5] A. Kendon, Gesture and Speech: How They Interact. In Wiemann, John M. and Harrison, Randall P., Nonverbal Interaction., Sage Publications, Beverly Hills, 1983.

[6] S. Berman, H. Stern, Sensors for gesture recognition systems, Ieee Transactions on Systems Man and Cybernetics Part C-Applications and Reviews 42 (3) (2012) 277–290.

[7] S. Mitra, T. Acharya, Gesture recognition : A survey, IEEE Transactions on Systems, Man, and Cybernetics-Part C: Application and Reviews 37 (3) (2007) 311–324.

[8] A. Erol, G. Bebis, M. Nicolescu, R. D. Boyle, X. Twombly, Vision-based hand pose estimation: A review, Computer Vision and Image Understanding 108 (2007) 52–73.

[9] S. C. W. Ong, S. Ranganath, Automatic sign language analysis: A survey and the future beyond lexical meaning, IEEE TPAMI 27 (6) (2005) 873–891.

[10] K. H. Lee, J. H. Kim, An hmm based threshold model approach for gesture recognition, IEEE TPAMI 21 (10) (1999) 961–973.

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

[12] A. Ramamoorthy, N. Vaswani, S. Chaudhury, S. Banerjee, Recognition of dynamic hand gestures, Pattern Recognition 36 (2003) 2069–2081.

[13] A. Just, S. Marcel, A comparative study of two state-of-the-art sequence processing techniques for hand gesture recognition, Computer Vision and Image Understanding 113 (4) (2009) 532–543.

[14] F. S. Chen, C. M. Fu, C. L. Huang, Hand gesture recognition using a real-time tracking method and hidden markov models, Image and Vision Computing 21 (2003) 745–758.

[15] S. Marcel, O. Bernier, J. E. Viallet, D. Collobert, Hand gesture recognition using input/output hidden markov models, in: IEEE FG 2000, 2000, pp. 456–461.

[16] N. Liu, B. C. Lovell, P. J. Kootsookos, Evaluation of hmm training algorithms for letter hand gesture recognition, in: 3rd IEEE International Symposium on Signal Processing and Information Technology, Darmstadt, GERMANY, 2003, pp. 648–651.

[17] D. Kim, J. Song, D. Kim, Simultaneous gesture segmentation and recognition based on forward spotting accumulative hmms, Pattern Recognition 40 (11) (2007) 3012–3026.

[18] W. H. A. Wang, C. L. Tung, Dynamic hand gesture recognition using hierarchical dynamic bayesian networks through low-level image processing, in: 7th International Conference on Machine Learning and Cybernetics, Kunming, China, 2008, pp. 3247–3253.

[19] C. L. Huang, M. S. Wu, S. H. Jeng, Gesture recognition using the multi-pdm method and hidden markov model, Image and Vision Computing 18 (11) (2000) 865–879.

[20] J. Beh, D. K. Han, R. Durasiwami, H. Ko, Hidden markov model on a unit hypersphere space for gesture trajectory recognition, Pattern Recognition Letters 36 (2014) 144–153.

[21] J. Beh, D. Han, H. Ko, Rule-based trajectory segmentation for modeling hand motion trajectory, Pattern Recognition 47 (4) (2014) 1586–1601.

[22] J. H. Lee, T. Delbruck, et al., Real-time gesture interface based on event-driven processing from stereo silicon retinas, Ieee Transactions on Neural Networks and Learning Systems 25 (12) (2014) 2250–2263.

[23] S. Theodorakis, V. Pitsikalis, P. Maragos, Dynamic-static unsupervised sequentiality, statistical subunits and lexicon for sign language recognition, Image and Vision Computing 32 (8) (2014) 533–549.

[24] H. I. Suk, B. K. Sin, S. W. Lee, Hand gesture recognition based on dynamic bayesian network framework, Pattern Recognition 43 (9) (2010) 3059–3072.

[25] Q. Chen, N. D. Georganas, E. M. Petriu, Hand gesture recognition using haar-like features and a stochastic context-free grammar, IEEE Transactions on Instrumentation and Measurement 57 (8) (2008) 1562–1571.

[26] G. Caridakis, K. Karpouzis, A. Drosopoulos, S. Kollias, Somm: Self organizing markov map for gesture recognition, Pattern Recognition Letters 31 (1) (2010) 52–59.

[27] M. Abid, E. Petriu, E. Amjadian, Dynamic sign language recognition for smart home interactive application using stochastic linear formal grammar, IEEE Transactions on Instrumentation and Measurement 64 (3) (2014) 596–605.

[28] W. W. Kong, S. Ranganath, Towards subject independent continuous sign language recognition: A segment and merge approach, Pattern Recognition 47 (3) (2014) 1294–1308.

[29] M. Yeasin, S. Chaudhuri, Visual understanding of dynamic hand gestures, Pattern Recognition 33 (11) (2000) 1805–1817.

[30] P. Hong, M. Turk, T. S. Huang, Gesture modeling and recognition using finite state machines, in: IEEE FG 2000, 2000, p. 410415.

[31] J. Davis, M. Shah, Recognizing hand gestures, in: Proceedings of the European Conference on Computer Vision, 1994, p. 331340.

[32] M. H. Yang, N. Ahuja, M. Tabb, Extraction of 2d motion trajectories and its application to hand gesture recognition, IEEE TPAMI 24 (8) (2002) 1061–1074.

[33] M. H. Yang, N. Ahuja, Extraction and classification of visual motion patterns for hand gesture recognition, in: IEEE CVPR 1998, Santa Barbara, CA, USA, 1998, pp. 892–897.

[34] C. W. Ng, S. Ranganath, Real-time gesture recognition system and application, Image and Vision Computing 20 (2002) 993–1007.

[35] X. H. Shen, G. Hua, L. Williams, Y. Wu, Dynamic hand gesture recognition: An exemplar-based approach from motion divergence fields, Image and Vision Computing 30 (3) (2012) 227–235.

[36] J. Cheng, C. Xie, W. Bian, D. C. Tao, Feature fusion for 3d hand gesture recognition by learning a shared hidden space, Pattern Recognition Letters 33 (4) (2012) 476–484.

[37] K. S. Patwardhan, S. D. Roy, Hand gesture modelling and recognition involving changing shapes and trajectories, using a predictive eigentracker, Pattern Recognition Letters 28 (2007) 329–334.

[38] K. Daniel, M. John, M. Charles, A person independent system for recognition of hand postures used in sign language, Pattern Recognition Letters 31 (2010) 1359–1368.

[39] M. C. Shin, L. V. Tsap, D. B. Goldgof, Gesture recognition using bezier curves for visualization navigation from registered 3-d data, Pattern Recognition 37 (5) (2004) 1011–1024.

[40] T. Kuremoto, Y. Kinoshita, L. Feng, S. Watanabe, K. Kobayashi, M. Obayashi, A gesture recognition system with retina-v1 model and one-pass dynamic programming, Neurocomputing 116 (2012) 291–300.

[41] A. Corradini, Dynamic time warping for off-line recognition of a small gesture vocabulary, in: IEEE ICCVW, 2001, p. 8289.

[42] J. F. Lichtenauer, E. A. Hendriks, M. J. T. Reinders, Sign language recognition by combining statistical dtw and independent classification, IEEE TPAMI 30 (11).

[43] K. Tohyama, K. Fukushima, Neural network model for extracting optic flow, Neural Networks 18 (5-6) (2005) 549556.

[44] P. K. Pisharady, P. Vadakkepat, A. P. Loh, Attention based detection and recognition of hand postures against complex backgrounds, International Journal of Computer Vision 101 (03) (2013) 403–419.

[45] P. K. Pisharady, P. Vadakkepat, A. P. Loh, Hand posture and face recognition using a fuzzy-rough approach, International Journal of Humanoid Robotics 07 (03) (2010) 331–356.

[46] P. K. Pisharady, P. Vadakkepat, A. P. Loh, Fuzzy-rough discriminative feature selection and classification algorithm, with application to microarray and image datasets, Applied Soft Computing 11 (04) (2011) 3429–3440.

[47] P. K. Pisharady, Q. S. H. Stephanie, P. Vadakkepat, A. P. Loh, Hand posture recognition using neuro-biologically inspired features, in: International Conference on Computational Intelligence, Robotics and Autonomous Systems (CIRAS) 2010, Bangalore, 2010.

[48] J. Alon, V. Athitsos, Q. Yuan, S. Sclaroff, A unified framework for gesture recognition and spatiotemporal gesture segmentation, IEEE TPAMI 31 (09) (2009) 1685–1699.

[49] A. Licsar, T. Sziranyi, Dynamic training of hand gesture recognition system, in: J. Kittler, M. Petrou, M. Nixon (Eds.), ICPR 2004, Cambridge, England, 2004, pp. 971–974.

[50] A. Licsar, T. Sziranyi, User-adaptive hand gesture recognition system with interactive training, Image and Vision Computing 23 (2005) 1102–1114.

[51] M. Zhao, F. K. H. Quek, X. Wu, Rievl: Recursive induction learning in hand gesture recognition, IEEE TPAMI 20 (11) (1998) 1174–1185.

[52] X. Teng, B. Wu, W. Yu, C. Liu, A hand gesture recognition system based on local linear embedding, Journal of Visual Languages & Computing 16 (2005) 442–454.

[53] M. Hasanuzzamana, T. Zhanga, V. Ampornaramveth, H. Gotoda, Y. Shirai, H. Ueno, Adaptive visual gesture recognition for human-robot interaction using a knowledge-based software platform, Robotics and Autonomous Systems 55 (8) (2007) 643–657.

[54] Y. Wu, T. S. Huang, View-independent recognition of hand postures, in: IEEE CVPR 2000, Vol. 2, 2000, pp. 88–94.

[55] O. Eng-Jon, R. Bowden, A boosted classifier tree for hand shape detection, in: IEEE FG 2004, 2004, pp. 889–894.

[56] E. Stergiopoulou, N. Papamarkos, Hand gesture recognition using a neural network shape fitting technique, Engineering Applications of Artificial Intelligence 22 (8) (2009) 1141–1158.

[57] P. Premaratne, S. Ajaz, M. Premaratne, Hand gesture tracking and recognition system using lucaskanade algorithms for control of consumer electronics, Neurocomputing 116 (2013) 242–249.

[58] D. Y. Huang, W. C. Hu, S. H. Chang, Gabor filter-based hand-pose angle estimation for hand gesture recognition under varying illumination, Expert Systems with Applications 38 (5) (2011) 6031–6042.

[59] P. K. Pisharady, P. Vadakkepat, A. P. Loh, Computational Intelligence in Multi-Feature Visual Pattern Recognition - Hand Posture and Face Recognition using Biologically Inspired Approaches, Springer, Singapore, 2014.

[60] P. K. Pisharady, Computational intelligence techniques in visual pattern recognition, PhD Thesis, National University of Singapore.

[61] S. S. Ge, Y. Yang, T. H. Lee, Hand gesture recognition and tracking based on distributed locally linear embedding, Image and Vision Computing 26 (2008) 1607–1620.

[62] P. K. Pisharady, P. Vadakkepat, A. P. Loh, Graph matching based hand posture recognition using neuro-biologically inspired features, in: International Conference on Control, Automation, Robotics and Vision (ICARCV) 2010, Singapore, 2010.

[63] J. Triesch, C. Malsburg, A system for person-independent hand posture recognition against complex backgrounds, IEEE TPAMI 23 (12) (2001) 1449–1453.

[64] J. Triesch, C. Malsburg, Robust classification of hand postures against complex backgrounds, in: IEEE FG 1996, Killington, VT, USA, 1996, pp. 170–175.

[65] J. Triesch, C. Malsburg, A gesture interface for human-robot-interaction, in: IEEE FG 1998, Nara, Japan, 1998, pp. 546–551.

[66] J. Triesch, C. Eckes, Object recognition with multiple feature types, in: ICANN'98, 8th International Conference on Artificial Neural Networks, Skovde, Swedan, 1998.

[67] Y.-T. Li, J. P. Wachs, Hegm: A hierarchical elastic graph matching for hand gesture recognition, Pattern Recognition 47 (1) (2014) 80–88.

[68] V. Athitsos, S. Sclaroff, Estimating 3d hand pose from a cluttered image, in: IEEE CVPR 2003, Vol. 2, 2003, pp. 432–9.

[69] E. Ueda, Y. Matsumoto, M. Imai, T. Ogasawara, A hand-pose estimation for vision-based human interfaces, IEEE Transactions on Industrial Electronics 50 (4) (2003) 676–684.

[70] X. Yin, M. Xie, Estimation of the fundamental matrix from uncalibrated stereo hand images for 3d hand gesture recognition, Pattern Recognition 36 (2003) 567–584.

[71] J. Lee, T. Kunii, Model-based analysis of hand posture, IEEE Comput. Graph. Appl. 15 (5) (1995) 77–86.

[72] A. El-Sawah, N. D. Georganas, E. Petriu, A prototype for 3-d hand tracking and posture estimation, IEEE Transactions on Instrumentation and Measurement 57 (8) (2008) 1627–1636.

[73] S. T. Roweis, L. K. Saul, Nonlinear dimensionality reduction by locally linear embedding, Science 290 (5500) (2000) 2323–2326.

[74] D. Conte, P. Foggia, C. Sansone, M. Vento, Thirty years of graph matching in pattern recognition, International Journal of Pattern Recognition and Artificial Intelligence 18 (3) (2004) 265–298.

[75] M. Lades, J. C. Vorbruggen, J. Buhmann, J. Lange, C. Malsburg, R. P. Wurtz, W. Konen, Distortion invariant object recognition in the dynamic link architecture, IEEE Transactions on Computers 42 (3) (1993) 300–311.

[76] L. Wiskott, J. M. Fellous, N. Kruger, C. Malsburg, Face recognition by elastic bunch graph matching, IEEE TPAMI 19 (7) (1997) 775–779.

[77] Z. Zhang, Microsoft kinect sensor and its effect, IEEE Multi Media 19 (2) (2012) 04–10.

[78] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, A. Blake, Real-time human pose recognition in parts from single depth images, in: IEEE CVPR 2011, Colorado Springs, 2011.

[79] J. Han, L. Shao, D. Xu, J. Shotton, Enhanced computer vision with microsoft kinect sensor: A review, Ieee Transactions on Cybernetics 43 (5) (2013) 1318–1334.

[80] R. Munoz-Salinas, R. Medina-Carnicer, F. J. Madrid-Cuevas, A. Carmona-Poyato, Depth silhouettes for gesture recognition, Pattern Recognition Letters 29 (3) (2008) 319–329.

[81] F. Dominio, M. Donadeo, P. Zanuttigh, Combining multiple depth-based descriptors for hand gesture recognition, Pattern Recognition Letters 50 (2014) 101–111.

[82] M. R. Malgireddy, I. Inwogu, V. Govindaraju, A temporal bayesian model for classifying, detecting and localizing activities in video sequences, in: IEEE CVPRW 2012, 2012, pp. 43–48.

[83] J. Sung, C. Ponce, B. Selman, A. Saxena, Human activity detection from rgbd images, in: AAAI workshop on Pattern, Activity and Intent Recognition (PAIR), 2011.

[84] J. Sung, C. Ponce, B. Selman, A. Saxena, Unstructured human activity detection from rgbd images, in: IEEE ICRA, 2012.

[85] P. K. Pisharady, M. Saerbeck, Kinect based body posture detection and recognition system, in: International Conference on Graphic and Image Processing (ICGIP), 2012.

[86] M. B. Holte, T. Moeslund, P. Fihl, Fusion of range and intensity information for view invariant gesture recognition, in: IEEE CVPRW 2008, 2008, pp. 1–7.

[87] M. Van den Bergh, D. Carton, R. De Nijs, N. Mitsou, C. Landsiedel, K. Kuehnlenz, D. Wollherr, L. Van Gool, M. Buss, Real-time 3d hand gesture interaction with a robot for understanding directions from humans, in: IEEE International Symposium on Robot and Human Interactive Communication (IEEE RO-MAN), 2011.

[88] R. Zhou, Y. Junsong, Z. Zhengyou, Robust hand gesture recognition based on finger-earth movers distance with a commodity depth camera, in: In Proceedings of ACM Multimeida, 2011.

[89] L. Gallo, A. P. Placitell, M. Ciampi, Controller-free exploration of medical image data: experiencing the kinect, International Symposium on Computer-Based Medical Systems (CBMS).

[90] W. Di, Z. Fan, S. Ling, One shot learning gesture recognition from rgbd images, in: IEEE CVPRW 2012, 2012.

[91] C. Keskin, F. Kirac, Y. Kara, L. Akarun, Randomized decision forests for static and dynamic hand shape classification, in: IEEE CVPRW 2012, 2012, pp. 31–46.

[92] Y. M. Lui, A least squares regression framework on manifolds and its application to gesture recognition, in: IEEE CVPRW, 2012, 2012, pp. 13–18.

[93] K. J. Lai, K., P. Ishwar, A gesture-driven computer interface using kinect, in: IEEE SSIAI, 2012, pp. 185–188.

[94] D. Frolova, H. Stern, S. Berman, Most probable longest common subsequence for recognition of gesture character input, IEEE Transactions on Cybernetics 43 (3) (2013) 871–880.

[95] H. Jiang, B. S. Duerstock, J. P. Wachs, A machine vision-based gestural interface for people with upper extremity physical impairments, Ieee Transactions on Systems Man Cybernetics-Systems 44 (5) (2014) 630–641.

[96] P. K. Pisharady, M. Saerbeck, A robust gesture detection and recognition algorithm for domestic robot interactions, in: International Conference on Control, Automation, Robotics and Vision (ICARCV) 2014, Singapore, 2014.

[97] R. Krishnan, S. Sarkar, Conditional distance based matching for one-shot gesture recognition, Pattern Recognition 48 (4) (2015) 1302–1314.

[98] C. Zhang, T. Yingli, Histogram of 3d facets: A depth descriptor for human action and hand gesture recognition, Computer Vision and Image Understanding In press.

[99] L. Y. M., Human gesture recognition on product manifolds, Journal of Machine Learning Research 13 (2012) 3297–3321.

[100] U. Mahbub, H. Imtiaz, T. Roy, M. Rahman, M. Ahad, A template matching approach of one-shot-learning gesture recognition, Pattern Recognition Letters 34 (2013) 1780–1788.

[101] W. Jun, R. Qiuqi, L. Wei, S. D., One-shot learning gesture recognition from rgb-d data using bag of features, Journal of Machine Learning Research 14 (2013) 2549–2582.

[102] P. K. Pisharady, M. Saerbeck, Robust gesture detection and recognition using dynamic time warping and multi-class probability estimates, in: IEEE Symposium on Computational Intelligence for Multimedia, Signal and Vision Processing (CIMSIVP), 2013.

[103] H. Cheng, Z. Dai, Z. Liu, Image-to-class dynamic time warping for 3d hand gesture recognition, in: IEEE ICME 2013, 2013.

[104] H. Cheng, J. Luo, X. Chen, A windowed dynamic time warping approach for 3d continuous hand gesture recognition, in: IEEE ICME 2014, 2014.

[105] E. Ohn-Bar, M. Trivedi, Hand gesture recognition in real time for automotive interfaces: A multimodal vision-based approach and evaluations, Ieee Transactions on Intelligent Transportation Systems 15 (6) 2368–2377.

[106] M. G. Jacob, J. P. Wachs, Context-based hand gesture recognition for the operating room, Pattern Recognition Letters 36 (2014) 196–203.

[107] O. Mendels, H. Stern, S. Berman, User identification for home entertainment based on free-air hand motion signatures, Ieee Transactions on Systems, Man, and Cybernetics: Systems 44 (11) 1461–1473.

[108] S.-Z. Li, B. Yu, W. Wu, S.-Z. Su, R.-R. Ji, Feature learning based on sae-pca network for human gesture recognition in rgbd images, Neurocomputing 151 (2015) 565–573.

[109] Y. Ming, Hand fine-motion recognition based on 3d mesh mosift feature descriptor, Neurocomputing 151 (2015) 574–582.

[110] R. Schramm, C. R. Jung, E. R. Miranda, Dynamic time warping for music conducting gestures evaluation, Ieee Transactions on Multimedia 17 (2) (2015) 243–255.

[111] F. A. Kondori, S. Yousefi, J.-P. Kouma, L. Liu, H. Li, Direct hand pose estimation for immersive gestural interaction, Pattern Recognition Letters In press.

[112] Z. Ren, J. Yuan, J. Meng, Z. Zhang, Robust part-based hand gesture recognition using kinect sensor, Ieee Transactions on Multimedia 15 (5) (2013) 1110–1120.

[113] R. Zhou, M. Jingjing, Y. Junsong, Depth camera based hand gesture recognition and its applications in human computer interaction, in: International Conference on Information, Communications and Signal Processing (ICICS), 2011.

[114] Y. Li, Hand gesture recognition using kinect, in: 3rd International Conference onSoftware Engineering and Service Science (ICSESS), 2012.

[115] D. Paul, A. Vassilis, K. Dimitrios, S. P., Hand shape and 3d pose estimation using depth data from a single cluttered frame, Advances in Visual Computing 7431 (2012) 148–158.

[116] F. Kirac, Y. E. Kara, L. Akarun, Hierarchically constrained 3d hand pose estimation using regression forests from single frame depth data, Pattern Recognition Letters 50 (2014) 91–100.

[117] Y. Yao, Y. Fu, Contour model based hand-gesture recognition using kinect sensor, Ieee Transactions on Circuits and Systems for Video Technology 24 (11) (2014) 1935–1944.

[118] X. Suau, M. Alcoverro, A. Lopez-Mendez, J. Ruiz-Hidalgo, J. R. Casas, Real-time fingertip localization conditioned on hand gesture classification, Image and Vision Computing 32 (8) (2014) 522–532.

[119] F. Kirac, Y. E. Kara, L. Akarun, Hierarchically constrained 3d hand pose estimation using regression forests from single frame depth data, Pattern Recognition Letters 50 (2014) 91–100.

[120] C. Wang, Z. Liu, S.-C. Chan, Superpixel-based hand gesture recognition with kinect depth camera, Ieee Transactions on Multimedia 17 (1) (2015) 29–39.

[121] A. Bobick, J. Davis, The recognition of human movement using temporal templates, IEEE TPAMI 23 (3).

[122] H. Liang, J. Yuan, D. Thalmann, Parsing the hand in depth images, Ieee Transactions on Multimedia 16 (5) (2014) 1241–1253.

[123] S. Belongie, J. Malik, J. Puzicha, Shape matching and object recognition using shape contexts, IEEE TPAMI 24 (3) (2002) 509–522.

[124] F. Erden, A. E. Cetin, Hand gesture based remote control system using infrared sensors and a camera, Ieee Transactions on Consumer Electronics 60 (4) (2014) 675–680.

[125] F. Wang, C. Zhang, Feature extraction by maximizing the average neighborhood margin, in: IEEE CVPR 2007, 2007, pp. 1–8.

[126] F. Simon, M. M. Helena, K. Pushmeet, N. Sebastian, Instructing people for training gestural interactive systems, in: International Conference on Human Factors in Computing Systems, CHI, ACM, 2012, pp. 1737–1746.

[127] I. Guyon, V. Athitsos, P. Jangyodsuk, B. Hamner, H. Escalante, Chalearn gesture challenge: Design and first results, in: IEEE CVPRW 2012, 2012, pp. 1–6.

[128] S. Escalera, J. Gonzlez, X. Bar, M. Reyes, O. Lopes, I. Guyon, V. Athistos, H. Escalante, Multi-modal gesture recognition challenge 2013: Dataset and results, in: 15th ACM International Conference on Multimodal Interaction (ICMI), Sydney, Australia, 2013.

[129] Y. Chuang, L. Chen, G. Chen, Saliency-guided improvement for hand posture detection and recognition, Neurocomputing 133 (2014) 404–415.

[130] M. Chen, G. AlRegib, B. H. Juang, 6dmg: A new 6d motion gesture database, in: IEEE CVPRW, 2011.

[131] A. Just, O. Bernier, S. Marcel, Hmm and iohmm for the recognition of mono- and bi-manual 3d hand gestures, in: British Machine Vision Conference (BMVC), 2004.

[132] S. Yale, D. David, D. Randall, Tracking body and hands for gesture recognition: Natops aircraft handling signals database, in: IEEE FG 2011, Santa Barbara, CA, 2011, pp. 500–506.

[133] S. Marcel, Hand posture recognition in a body-face centered space, in: Proceedings of the Conference on Human Factors in Computer Systems (CHI), 1999.

[134] S. Ruffieux, D. Lalanne, E. Mugellini, Chairgest: A challenge for multimodal mid-air gesture recognition for close hci, in: 15th ACM on International Conference on Multimodal Interaction (ICMI), 2013.

[135] L. Liu, L. Shao, Learning discriminative representations from rgb-d video data, in: International Joint Conference on Artificial Intelligence (IJCAI), 2013.

[136] J. Zhuolin, L. S. Davis, Recognizing actions by shape-motion prototype trees, in: IEEE ICCV, 2009, 2009, pp. 444–451.

[137] T.-K. Kim, S.-F. Wong, R. Cipolla, Tensor canonical correlation analysis for action classification, in: IEEE CVPR 2007, 2007, pp. 1–8.

[138] J. Tompson, M. Stein, Y. Lecun, K. Perlin, Real-time continuous pose recovery of human hands using convolutional networks, ACM Transactions on Graphics 33.

[139] E. Kollorz, J. Penne, J. Hornegger, A. Barke, Gesture recognition with a time-of-flight camera, Int. J. Intell. Syst. Technol. Appl. 5 334–343.

[140] M. Kawulok, J. Kawulok, J. Nalepa, Spatial-based skin detection using discriminative skin-presence features, Pattern Recognition Letters 41 (2014) 3–13.

[141] N. Pugeault, R. Bowden, Spelling it out: Real-time asl fingerspelling recognition, in: ICCV 2012, 2011.

[142] Q. Chen, S. Xiao, W. Yichen, T. Xiaoou, S. Jian, Realtime and robust hand tracking from depth, in: IEEE CVPR 2014, 2014.

[143] C. Neidle, A. Thangali, S. Sclaroff, Challenges in development of the american sign language lexicon video dataset corpus, in: 5th Workshop on the Representation and Processing of Sign Languages: Interactions between Corpus and Lexicon, LREC 2012, 2012.

[144] R. Kehl, L. Van Gool, Real-time pointing gesture recognition for an immersive environment, in: IEEE FG 2004, Seoul, Korea, 2004, p. 577582.

[145] N. Jojic, B. Brumitt, B. Meyers, S. Harris, T. Huang, Detection and estimation of pointing gestures in dense disparity maps, in: IEEE FG 2000, Grenoble, France, 2000, p. 10001007.

[146] C. B. Park, S. W. Lee, Real-time 3d pointing gesture recognition for mobile robots with cascade hmm and particle filter, Image and Vision Computing 29 (1) (2011) 51–63.

[147] C. B. Park, M. C. Roh, S. W. Lee, Ieee, Real-Time 3D Pointing Gesture Recognition in Mobile Space, IEEE International Conference on Automatic Face and Gesture Recognition, FG 2008, 2008.

[148] K. Nickel, R. Stiefelhagen, Visual recognition of pointing gestures for human-robot interaction, Image and Vision Computing 25 (12) (2007) 1875–1884.

[149] J. L. Raheja, A. Chaudhary, S. Maheshwari, Hand gesture pointing location detection, Optik 125 (3) (2014) 993–996.

[150] M. Pateraki, H. Baltzakis, P. Trahanias, Visual estimation of pointed targets for robot guidance via fusion of face pose and hand orientation, Computer Vision and Image Understanding 120 (2014) 1–13.

[151] M. Pateraki, H. Baltzakis, P. Trahanias, Ieee, Visual estimation of pointed targets for robot guidance via fusion of face pose and hand orientation, 2011 ICCVW.

[152] T. Serre, L. Wolf, S. Bileschi, M. Riesenhuber, T. Poggio, Robust object recognition with cortex-like mechanisms, IEEE TPAMI 29 (3) (2007) 411–426.

[153] H. Kang, W. L. Chang, K. C. Jung, Recognition-based gesture spotting in video games, Pattern Recognition Letters 25 (15) (2004) 1701–1714.

[154] Y. Yin, R. Davis, Gesture spotting and recognition using salience detection and concatenated hidden markov models, in: 15th ACM on International conference on multimodal interaction, ICMI 2013, 2013, pp. 489–494.

[155] R. D. Yang, S. Sarkar, B. Loeding, Handling movement epenthesis and hand segmentation ambiguities in continuous sign language recognition using nested dynamic programming, IEEE TPAMI 32 (3) (2010) 462–477.

[156] L. Hong, G. Michael, Model-based segmentation and recognition of dynamic gestures in continuous video streams, Pattern Recognition 44 (8).

[157] S. Sarkar, B. Loeding, R. Yang, S. Nayak, A. Parashar, Segmentation-robust representations, matching, and modeling for sign language, in: IEEE CVPRW, 2011, pp. 13–19.

[158] R. D. Yang, S. Sarkar, Coupled grouping and matching for sign and gesture recognition, Computer Vision and Image Understanding 113 (6) (2009) 663–681.