Vision Based Hand Gesture Recognition: A Review

G. Simion, V. Gui, and M. Otesteanu

***Abstract*—** This paper reviews some major trends and the recent evolution in the field of hand gesture recognition. While providing a non-exhaustive inventory of the huge amount of past research in the field, the paper reviews in more detail part based approaches, particularly those embedded in the compositional framework, an emerging dominant trend in computer vision. Several traditional and new applications are also discussed in the paper.

***Keywords*—** hand gestures, human computer interaction, model based approach, recognition, view based approach.

I. Introduction

PEOPLE perform various gestures in their daily lives. It is in our nature to use gestures in order to improve the communication between us. Try to imagine speaking with a person who makes no gesture. It is very difficult to understand if your message is clear for him or her, if he or she agrees with your saying, in other words it is very hard to guess what type of reaction your message produces. Between all kind of gestures that we perform, hand gestures play an important role. Hand gestures can help us say more in less time. In these days, computers have become an important part in our lives, so why not use hand gesture in order to communicate with them.

The direct use of the hand as an input device is an attractive method for providing natural Human–Computer Interaction. Two approaches are commonly used to interpret gestures for Human Computer Interaction.

Methods Which Use Data Gloves: Since now, the only technology that satisfies the advanced requirements of hand-based input for HCI is glove-based sensing This method employs sensors (mechanical or optical) attached to a glove that transducers' finger flexions into electrical signals for determining the hand posture. Several drawbacks make this technology not so popular: first of all interaction with the computer-controlled environment loses naturalness and easiness, the user is forced to carry a load of cables, which are connected to the computer and it also requires calibration and setup procedures.

Methods which are Vision Based: Computer vision based techniques have the potential to provide more natural and non-contact solutions, there are non invasive and are based on the way human beings perceive information about their surroundings. Although it is difficult to design a vision based interface for generic usage, yet it is feasible to design such an interface for a controlled environment but has no lake of challenges including accuracy and processing speed.

This paper is organized as follows: In section 2 we provide a survey on vision based hand gesture recognition. In section 3 we present two different gesture taxonomies. In section 4 various applications areas for gesture recognition are presented, while in section 5 we give the conclusions.

II. Vision based hand gesture recognition

The approaches to Vision based hand gesture recognition can be divided into two categories: 3 D hand model based approaches and appearance based approaches [1].

A. Model based approach

Model-based approaches generate model hypotheses and evaluate them on the available visual observations. Essentially, this is performed by formulating an optimization problem whose objective function measures the discrepancy between the visual cues that are expected due to a model hypothesis and the actual ones. The employed optimization method must be able to evaluate the objective function at arbitrary points in the multidimensional model parameters space. By large, the approach consists of searching for the kinematic parameters that brings the 2D projection of a 3D model of hand into correspondence with an edge-based image of a hand.

The model of the hand can be more or less elaborated.

A 3D model with 27 degrees of freedom (DOF), shown in Fig. 1 a, was introduced in [2] and has been used in many studies afterwards.

This paper was supported by the project "Develop and support multidisciplinary postdoctoral programs in primordial technical areas of national strategy of the research - development - innovation" 4D-POSTDOC, contract nr. POSDRU /89/1.5/S/52603, project co-funded from European Social Fund through Sectorial Operational Program Human Resources 2007-2013

G. Simion is with the Department of Communication, Politehnica University of Timisoara, Bd. Vasile Parvan, No. 2, 300223, Timisoara, Romania (phone: (+40)-0256-403291; fax: (+40)-0256-403295; e-mail: Georgiana.simion@etc.upt.ro).

V. Gui is with the Department of Communication, Politehnica University of Timisoara, Bd. Vasile Parvan, No. 2, 300223, Timisoara, Romania (e-mail: vasile.gui@etc.upt.ro).

M. Otesteanu is with the Department of Communication, Politehnica University of Timisoara, Bd. Vasile Parvan, No. 2, 300223, Timisoara, Romania (e-mail: marius.otesteanu@etc.upt.ro).

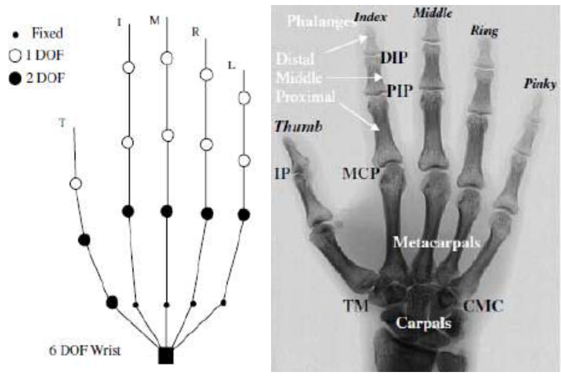

Fig.1 Skeletal hand model: (a) Hand anatomy, (b) the kinematic model according to [7]

The CMC joints, shown in Fig. 1b, are assumed to be fixed, which quite unrealistically models the palm as a rigid body. The fingers are modeled as planar serial kinematic chains attached to the palm at anchor points located at MCP joints.

Over the years the kinematic model was improved by adding extra twist motion to MCP joints [12], [3], by introducing one flexion/extension DOF to CMC joints [4] or by using a spherical joint for TM [5].

Rehg and Kanade [6] proposed one of the earliest model based approaches to the problem of bare hand tracking. They used a 3D model with 27 DOF for their system called DigitEyes. Their work describes a model based hand tracking system, which can recover the state of a 27 DOF hand model from ordinary gray scale images at speeds of up to 10 Hz. The hand tracking problem is posed as an inverse problem: given an image frame (edge map) find the underlying parameters of the model. The inverse mapping is nonlinear due to the trigonometric functions modeling the joint movements and the perspective image projection. A key observation is that the resulting image changes smoothly as the parameters are changed. Therefore, this problem is a promising candidate for assuming local linearity. Several iterative methods that assume local linearity exist for solving non-linear equations (e.g. Newton’s method). Upon finding the solution for a frame, the parameters found are used as initial values for the next frame and the fitting procedure is repeated. The approach can be thought of as a series of hypotheses and tests, where a hypothesis of model parameters at each step is generated in the direction of the parameter space (from the previous hypothesis) achieving the greatest decrease in miss-correspondence. These model parameters are then tested against the image. This approach has several disadvantages that have kept it from real-world use. First, at each frame the initial parameters have to be close to the solution, otherwise the approach is liable to find a suboptimal solution (i.e. local minima). Secondly, the fitting process is also sensitive to noise (e.g. lens aberrations, sensor noise) in the imaging process. Finally, the approach cannot handle the inevitable self-occlusion of the hand.

Heap et al.[8] proposed a deformable 3D hand model and modeled the entire surface of the hand by a surface mash constructed via PCA from training examples. Real-time tracking is achieved by finding the closest possibly deformed model matching the image. Such a representation requires further processing to extract useful higher-level information, such as pointing direction; however, it was shown to be very effective to reliably locate and track the hand in images. The method however, is not able to handle the occlusion problem and is not scale and rotation invariant.

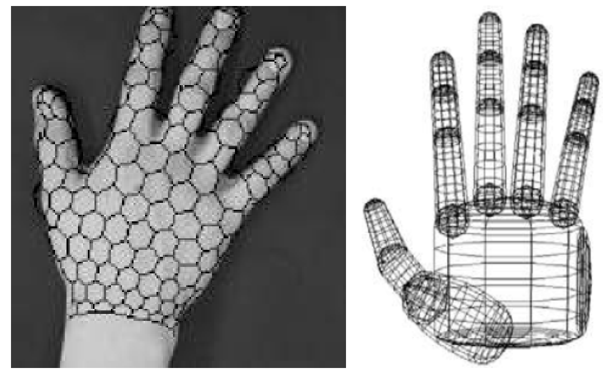

Fig.2 a) Hand tracking using 3D Point Distribution Model from [8] and b) Quadrics-based hand model from [9]

Stenger et al. used quadrics as shape primitives in [9]. The use of quadrics to build the 3D model yields a practical and elegant method for generating the contours of the model, which are then compared with the image data. The pose of the hand model is estimated with an Unscented Kalman filter (UKF), which minimizes the geometric error between the profiles and edges extracted from the images. The use of the UKF permits higher frame rates than more sophisticated estimation methods such as particle filtering, whilst providing higher accuracy than the extended Kalman filter. In [10] the authors have reformulated the problem within a Bayesian (probabilistic) framework. Bayesian approaches allow for the pooling of multiple sources of information (e.g. system dynamics, prior observations) to arrive at both an optimal estimate of the parameters and of a probability distribution of the parameter space, to guide future search for parameters. In contrast with the Kalman filter approaches, Bayesian approaches allow nonlinear system formulations and non-Gaussian (multi-modal) uncertainty modeling (e.g. errors caused by occlusions) at the expense of losing a closed-form solution of the uncertainty.

In [11], a model-based visual hand posture tracking algorithm is proposed to guide a dexterous robot hand. The approach adopts a 3D model-based framework with full-DOF kinematic and an effective measurement method based on chamfer distance for both silhouette and edges. Both a color-based approach using an adaptive skin-color model and a motion-based approach are used to locate the hand region in images. A genetic algorithm (GA) is integrated to traditional particle filter (PF) as a solution of high-dimensional and multi-modal tracking. In each tracking step, the particles updated from the PF are treated as the initial population of the GA, and the following selection, crossover, and mutation operators will draw the particles to optimized solutions. Compared with exponentially increasing number of particles in the traditional PF approach, the overhead added by GA operations is much less. Experimental results show a significant improvement of tracking performance compared with traditional PFs.

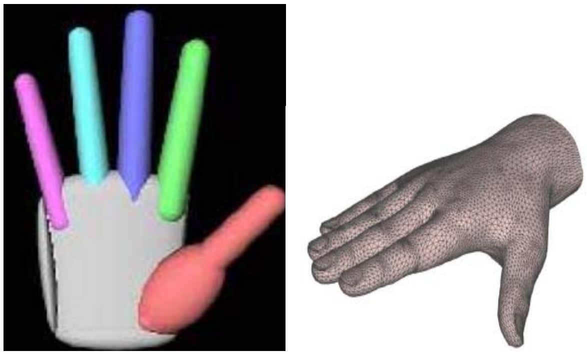

Fig.3 a) The 3D model presented in [11],b) The 3D model as a polygonal surface presented in [12]

In [13,14], an approach to the recovery of geometric and photometric pose parameters of a 3D model with 28 DOF from monocular image sequences is presented.

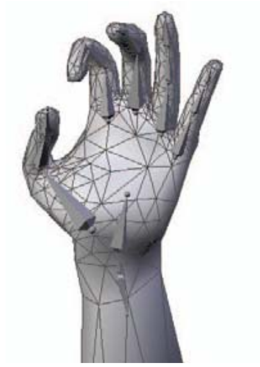

Fig.4 The deformed hand triangulated surface from [14].

The 3D hand pose, the hand texture and the illuminant are dynamically estimated through minimization of an objective function. Derived from an inverse problem formulation, the objective function enables explicit use of texture temporal continuity and shading information, while handling important self-occlusions and time-varying illumination. The minimization is done efficiently using a quasi-Newton method, for which a rigorous derivation of the objective function gradient was proposed. The deformed hand triangulated surface from [14] is presented in Fig.4.

In [15] truncated quadrics are used to build a 3D hand model where the DOF for each joint correspond to the DOF of a real hand. Quadratic chamfer distance function is used to compute the edge likelihood and the silhouette likelihood is performed by a Bayesian classifier and online adaptation of skin color probabilities. Particle filtering is used to track the hand by predicting the next state of 3D hand model.

Recent works [16], [17] use the depth information provided by the camera. One of the first who worked with range data for hand gesture recognition was Malassiotis [18], the 3D information was acquired following a structured light approach with a frame rate of twelve image pairs per second.

Bray [12] applied a similar method by using also structured light to generate depth maps in combination with a skin color model to eliminate the background. Their model consists of a polygonal skin, driven by an underlying skeleton. A new pose is computed by linearly blending the motions that each skin vertex would undergo when rigidly coupled to a subset of the skeleton joints. Several Stochastic Meta-Descent trackers (SMD) are combined as 'particles' within a particle filter framework, increasing the chance to find the global optimum. After propagating the particles, SMD is performed and the resulting new particle set is included, such that the original Bayesian distribution is not altered. The SPF tracker's performance is demonstrated on 3D hand tracking, using a structured light sensor sampling 720 × 576 pixels at 12.5 frames per second, and a Sunfire 1.2GHz for processing. To illustrate the advantage of using multiple SMD particles, a comparison between a single SMD tracker and 'smart particle filter'(SPF) with 4 additional SMD particles is done. For fast motion, a single particle fails to match the target and gets stuck in a local minimum. In contrast, the SPF tracks the target correctly over time with only $N = 4$, i.e. 8 particles.

In [16] a PMD-Sensor with a resolution of 160 × 120 pixels and a viewing angle of 40 was used. The goal of their system is to recognize 12 different static hand gestures. Segmentation of the hand and arm was done via distance values. In their experiences the distance between the hand and the sensor was between 70 cm and 110 cm.

In [17] a model-based method for efficient full DOF hand model initialization and tracking using data acquired by a Kinect sensor is proposed. The combination of a careful modeling of the problem, a powerful optimization method, the exploitation of modern GPUs and, the quality of the data provided by the Kinect sensor, results in a robust and efficient method for tracking the full pose of a hand in complex articulation. The experimental results demonstrate that accurate and robust 3D hand tracking is achievable at 15Hz.

Perrin et al. [19] worked with a target-oriented laser beam and two mirrors to perform 3D tracking. The authors determine distance measurements by using the absolute value of the reflected light, but the system cannot be used in real time.

A recent work is presented by Breuer [20]. They describe a system based on Swissranger SR2 TOF camera which provides depth information. To reconstruct the hand, a principle component analysis and a special hand model are used. In this work, they reconstruct the first seven degrees of freedom of the hand with a frame rate of about 3Hz.

In [21] the authors formulate an optimization problem that takes into account the constraints stems from the properties of the natural world, i.e., the fact that the hand and the object cannot share the same physical space. Thus, the 3D shape and pose of the object provides important information on the articulation of the hand and vice versa.

The 3D hand models are articulated deformable objects with many degrees of freedom; a very large image database is required to cover all the characteristic shapes under different views. Another common problem with model based approaches is the problem of feature extraction and lack of capability to deal with singularities that arise from ambiguous views.

B. Appearance based approaches

Appearance-based models are derived directly from the information contained in the images and have traditionally been used for gesture recognition. No explicit model of the hand is needed, which means that no internal degrees of freedom have to be specifically modeled.

When only the appearance of the hand in the video frames is known, differentiating between gestures is not as straight forward as with the model based approach. The gesture recognition will therefore typically involve some sort of statistical classifier based on a set of features that represent the hand. In many gesture applications all that is required is a mapping between input video and gesture. Therefore, many have argued that the full reconstruction of the hand is not essential for gesture recognition. Instead many approaches have used the extraction of low-level image measurements that are fairly robust to noise and can be extracted quickly. Low-level features that have been proposed in the literature include: the centroid of the hand region [22], principle axes defining an elliptical bounding region of the hand, and the optical flow/affine flow [23] of the hand region in a scene.

Another approach is to look for skin colored regions in the image. This is a very popular method [24], [25], [26], [27],[28],[29],[30] but has some drawbacks. First, skin color detection is very sensitive to lighting conditions. While practicable and efficient methods exist for skin color detection under controlled (and known) illumination, the problem of learning a flexible skin model and adapting it over time is challenging.

Lindberg [31] used scale-space color features to recognize hand gestures. Multi scale features can be found in an image at different scales. Therefore, the hand can be described as one bigger blob feature for the palm, having smaller blob features representing the finger tips which are connected by some rigid features. Furthermore, it was proposed to perform the feature extraction directly in the color space, as this allows the combination of probabilistic skin colors directly in the extraction phase. The advantage of directly working on a color image lies in the better distinction of hand and background regions, but the authors showed real time application only with no other skin colored objects present in the scene.

Another approach is to use the eigenspace. Given a set of images, eigenspace approaches construct a small set of basis images that characterize the majority of the variation in the training set and can be used to approximate any of the training images. To reconstruct an image in the training set, a linear combination of the basis vectors (images) is taken, where the coefficients of the basis vectors are the result of projecting the image to be reconstructed on to the respective basis vectors. In [32], an approach for tracking hands by an eigenspace approach is presented. The authors provide three major improvements to the original eigenspace approach formulation, namely, a large invariance to occlusions, some invariance to differences in background from the input images and the training images and the ability to handle both small and large affine transformations (i.e. scale and rotation) of the input image with respect to the training images. The authors demonstrate their approach with the ability to track four hand gestures using 25 basis images.

In the last years is noticeable a new trend: more and more approaches use invariant local features [33], [34], [35], [36], [37], [38], [39], [40].

In [33], the Adaboost learning algorithm with SIFT features is used. The Scale Invariant Feature Transform (SIFT) introduced by Lowe [41] consists of a histogram representing gradient orientation and magnitude information within a small image patch. SIFT is a rotation and scale invariant feature and is robust to some variations of illuminations, viewpoints and noise. The accuracy of multi-class hand posture recognition is improved by the sharing feature concept. However, different features such as contrast context histogram need to be studied and applied to accomplish hand posture recognition in real time. In order to prove the power of their algorithm, three hand posture classes, ”palm”, ”fist” and ”six”, are trained and recognized. For testing, 275 images were collected using the onboard Logitech QuickCam Pro 5000 with a resolution of 320x240. Using the feature sharing concept the recognition rate increases to 95.6% from 90.9%.

In [34] Bag-of-Words representation (BoW) and SIFT features are used. In a typical BoW representation, “interesting” local patches are first identified from an image, either by densely sampling, or by an interest point detector. These local patches, represented by vectors in a high dimensional space, are often referred to as the key points. The bag-of-words methods main idea is to quantize each extracted key point into one of the visual words, and then represent each image by a histogram of visual words.

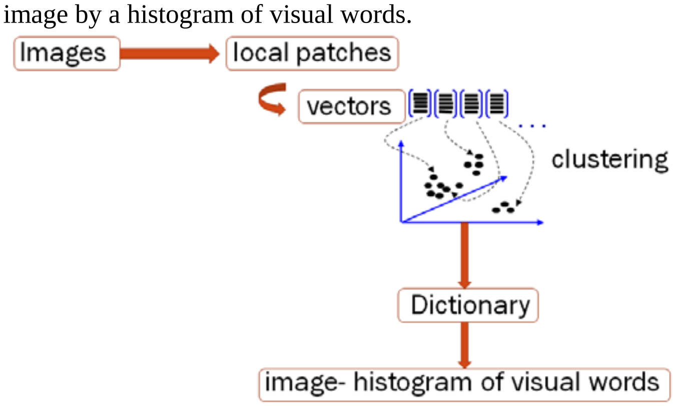

Fig.5 BoW representation

A clustering algorithm is generally used to generate the visual words dictionary. In [34] K-means algorithm has been used for clustering. A multi-class SVM was used to train the classifier model. In the testing stage, the keypoints were extracted from every image captured from the webcam and fed into the cluster model to map them with one (Bag-of-words) vector, which is finally fed into the multi-class SVM training classifier model to recognize the hand gesture.

In [35] the ARPD descriptor (Appearance and Relative Position Descriptor) is proposed. This descriptor includes color histogram, relative-position information, and SURF [42]. The process of constructing ARPD includes two steps: extracting SURF keypoints and color histogram from images, and computing relative-position information of every keypoint within images, the relative-position information is also included as part of ARPD. The ARPD was used in the BoW representation.

The BoW was used to detect and recognize hand posture based on a sliding-window framework. To meet real-time request, several approaches were proposed to speed up hand posture recognition process. The CAMESHIFT algorithm is used to track the hand motion and a strategy based on histogram is used to reinitialize the tracking process.

The comparison between improved BoW and standard Bow is presented in Fig.6 .

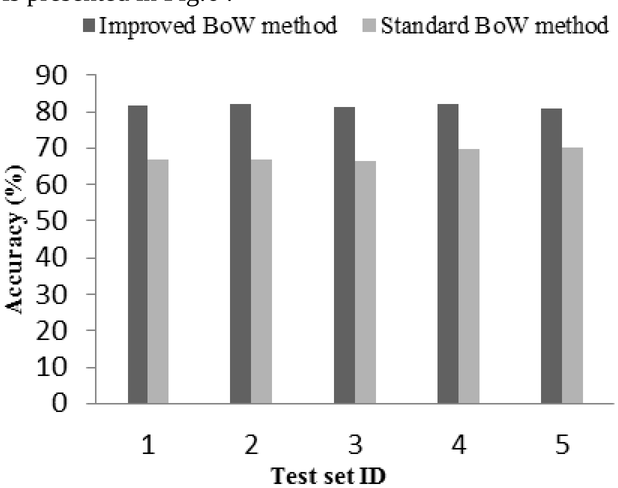

Fig.6 The comparison between improved BoW and standard Bow according to [35]

In [36], compositional techniques are used for hand posture recognition. A hand posture representation is based on compositions of parts: descriptors are grouped according to the perceptual laws of grouping [43] obtain a set of possible candidate compositions. These groups are a sparse representation of the hand posture based on overlapping subregions.

The detected part descriptors are represented as probability distributions over a codebook which is obtained in the learning phase. A composition is a mixture of the part distributions. From all candidate compositions, relevant compositions must be selected. There are two types of relevant compositions: those compositions that occur frequently in all categories and also those which are specific for a category. The category posterior of compositions is learned in the training phase, and it is a measure of relevance. The entropy of the category posterior helps to discriminate between categories. A cost function is obtained by combining the priors of the prototypes and the entropy. The process of recognition is based on bag of composition method, where a discriminative function is defined.

In [37], Maximally Stable Extremal Region (MSER) detector and color likelihood maps are used for hand tracking. Such a combination allows performing repeated figure/ground segmentation in every frame in an efficient manner.

The MSER detector is one of the best interest region detectors in computer vision [44]. MSER detection is mostly applied to single gray scale images, but the method can be easily extended for analysis of color images by defining a suitable ordering relationship on the color pixels. In general the MSER detector finds bright connected regions which have consequently darker values along their boundaries. The set of MSERs is closed under continuous geometric transformations and is invariant to affine intensity changes. Furthermore MSERs are detected at all scales. Therefore, due to these properties MSER detection is suited for segmentation purposes. The proposed method was implemented in C++ and runs with 25 frames per second on 320 × 240 video sequences.

In [38], [39], [40] Haar like features are used for the task of hand detection. Haar like features focus more on the information within a certain area of the image rather than each single pixel. To improve classification accuracy and achieve real time performance, the AdaBoost learning algorithm, that can adaptively select the best features in each step and combine them into a strong classifier, can be used. The training algorithm based on AdaBoost learning algorithm takes a set of “positive” samples, which contain the object of interest and a set of “negative” samples, i.e., images that do not contain objects of interest.

Invariant features allow modeling the hand as collection of characteristic parts. Key points or characteristic regions are extracted. Using such features the hand gesture is decomposed in simpler parts which are easier to recognize. This approach has major advantages: even if some parts are missing a gestures still can be recognized, so there is robustness to partial occlusions, changes in view point and considerable deformations. Bag of Words methods and compositional methods become more and more popular in hand gesture recognition. These techniques have been studied in many diverse fields such as linguistics, logic, and neuroscience, but compositionality is especially evident in the syntax and semantics of language where a limited number of letter scan form a huge variety of words and sentences. In computer vision these techniques are used in the context of a general problem: categorization. Using these techniques the semantic gap that exists between the low level features and high level representations can be addressed in a principled way. The hand posture is no longer modeled as a whole. The characteristic regions of the hand are assembled to form compositions; these compositions, in turn, can be grouped in compositions of compositions and so on. These techniques allow incorporating in the design the Gestalt laws of visual perception. The Gestalt laws are a set of visual rules that guide the construction process of groupings and yield compositions, establishing causal relationships between grouping constituents, and tends to emulate better, the way our brain-view processor works.

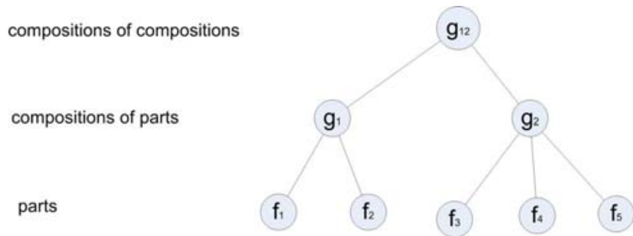

Fig.7 Compositional hierarchy: intermediate compositions of parts

III. Gesture Classification

There are several classifications for hand gestures in the literature.

A list in which their association with speech declines, language properties increase, spontaneity decreases, and social regulation increases can be seen below:

- *Gesticulation*: spontaneous movement of hands and arms, accompanying speech. These spontaneous movements constitute around 90% of human gestures. People gesticulate when they are on telephone, and even blind people regularly gesture when speaking to one another.
- *Languagelike* gestures: gesticulation integrated into a spoken utterance, replacing a particular spoken word or phrase.
- *Pantomimes*: gestures depicting objects or actions, with or without accompanying speech.
- *Emblems:* familiar signs such as “V for victory,” or OK, or Thumbs Up.

A taxonomy which is probably more suitable for HCI applications divides hand gestures into three groups: communicative gestures, manipulative gestures, and controlling gestures.

- Communicative gestures are intended to express an idea or a concept. These gestures are either used together with speeches or are a substitute for verbal communications which on the other hand requires a high structured set of gestures such as those defined in sign languages.
- Manipulative gestures are used for interaction with objects in an environment. These gestures are mostly used for interaction in virtual environments such as tele-operation or virtual assembly systems however; physical objects can be manipulated through gesture controlled robots.
- Controlling gestures are the group of gestures which are used to control a system or point and locate and object. Finger Mouse is a sample application which detects 2D finger movements and controls mouse movements on the computer desktop. Analyzing hand gestures is completely application dependant and involves analyzing the hand motion, modeling hand and arm, mapping the motion features to the model and interpreting the gesture in a time interval.

IV. Application Areas

There is a large variety of applications which involve hand gestures. Hand gestures can be used to achieve natural human computer interaction for virtual environments, or there can be used to communicate with deaf and dumb people. An important application area is that of vehicle interfaces. In this section an overview of few application areas is given.

Virtual Reality: Gestures for virtual and augmented reality applications have experienced one of the greatest levels of uptake interactions [45] or 2D displays that simulate 3D interactions [46].

Robotics and Telepresence: When robots are moved out of factories and introduced into our daily lives, they have to face many challenges such as cooperating with humans in complex and uncertain environments or maintaining long-term human-robot relationships. Telepresence and telerobotic applications are typically situated within the domain of space exploration and military-based research projects.

The gestures used to interact with and control robots are similar to fully-immersed virtual reality interactions, however the worlds are often real, presenting the operator with video feed from cameras located on the robot [47]. Here, gestures can control a robots hand and arm movements to reach for and manipulate actual objects, as well its movement through the world. Hand gesture recognition for robotic control is presented in [33, 48]

Desktop and Tablet PC Applications: In desktop computing applications, gestures can provide an alternative interaction to the mouse and keyboard [49]. Many gestures for desktop computing tasks involve manipulating graphics, or annotating and editing documents using pen-based gestures [50]. Recently eyeSight introduced gesture recognition Technology for Android Tablets and Windows-based Portable Computers [59].

Sign Language: Sign language is an important case of communicative gestures. Since sign languages are highly structural, they are very suitable as test beds for vision algorithms [51]. At the same time, they can also be a good way to help the disabled to interact with computers. Sign language for the deaf (e.g. American Sign Language) is an example that has received significant attention in the gesture literature [52], [53], [54] and [55].

Vehicle interfaces: A number of hand gesture recognition techniques for human vehicle interface have been proposed in [56], [57]. The primary motivation of research into the use of hand gestures for in-vehicle secondary controls is broadly based on the premise that taking the eyes off the road to operate conventional secondary controls can be reduced by using hand gestures.

Healthcare: Wachs et al. [58] developed a hand-gesture recognition system that enables doctors to manipulate digital images during medical procedures using hand gestures instead of touch screens or computer keyboards. A sterile human-machine interface is of supreme importance because it is the means by which the surgeon controls medical information, avoiding patient contamination, the operating room and the other surgeons. The gesture based system could replace touch screens now used in many hospital operating rooms which must be sealed to prevent accumulation or spreading of contaminants and requires smooth surfaces that must be thoroughly cleaned after each procedure – even that sometimes they are not. With infection rates at hospitals now at unacceptably high rates, the hand gesture recognition system offers a possible alternative.

V. Conclusion

In this paper a review of vision based hand gesture recognition methods has been presented. In the last years remarkable progress in the field of vision based hand gesture recognition, from hardware and software point of view, has been done. New type of cameras gives us depth information which is now included in algorithms, local invariant features allow us to use methods which may enclose Gestalt laws of visual perception. The Gestalt laws are a set of visual rules that guide the construction process of groupings, establishing causal relationships between grouping constituents, and tends to emulate better, the way our brain-view processor works. Further research in the areas of feature extraction, classification methods and gesture representation are required to realize the ultimate goal of humans interfacing with machines on their own natural terms.

It is likely that, in the near future hand gesture recognition will play a major role in our lives. Probably sooner that one may think, the surrounding devices will be hand gesture interfaced.

References

[1] H. Zhou, T.S. Huang, "Tracking articulated hand motion with Eigen dynamics analysis," *In Proc. Of International Conference on Computer Vision*, Vol 2, 2003, pp. 1102-1109

[2] J. Lee, T. Kunii, "Constraint-based hand animation, Models and Techniques," *in Computer Animation*, Springer, Tokyo (1993): 110–27

[3] M. Bray, E. Koller-Meier, P. Muller, L.V Gool., N.N. Schraudolph," 3D Hand tracking by rapid stochastic gradient descent using a skinning model," *.First European Conference on Visual Media Production* (2004): 297-302

[4] K. Nirei., H. Saito, M. Mochimaru, S.Ozawa, "Human hand tracking from binocular image sequences," *In 22th International Conference on Industrial Electronics, Control, and Instrumentation*

[5] J.J Kuch, T.S Huang, "Human computer interaction via the human hand: a hand model," *Twenty-Eighty Asilomar Conference on Signal, Systems, and Computers* (1994): 1252– 56

[6] J. Rehg, T. Kanade, "Visual tracking of high DoF articulated structures: An application to human hand tracking," *In European Conference on Computer Vision and Image Understanding* (1994): 35–46

[7] Ali Erol, George Bebis, Mircea Nicolescu, Richard D. Boyle, Xander Twombly., "Vision-based hand pose estimation: A review," *Computer Vision and Image Understanding* 108 (2007), pp 52–73

[8] A. J. Heap, D. C.Hogg, "Towards 3-D hand tracking using a deformable model," *In 2nd International Face and Gesture Recognition Conference* (1996), pp 140–45

[9] B. Stenger, P. R. S. Mendon, R Cipolla, "Model-Based 3D Tracking of an Articulated Hand," Proc. British Mac.hine Vision Conference 1 (2001): 63-72

[10] B. Stenger, A. Thayananthan, P.H.S. Torr, R. Cipolla, " Model-based hand tracking using a hierarchical Bayesian filter," *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2006)

[11] Jinshi Cui, Zengqi Sun, "Model-based visual hand posture tracking for guiding a dexterous robotic hand," *Optics Communications* 235 (2004) 311–318

[12] M Bay, Koller-Meier, L.V. Gool, "Smart particle filtering for 3D hand tracking," in: *Sixth IEEE International Conference on Automatic Face and Gesture Recognition*, Los Alamitos, CA, USA, 2004, pp 675

[13] M. de La Gorce, N. Paragios, D. J. Fleet, "Model-Based Hand Tracking with Texture, Shading and Self-occlusions," *IEEE Conference on Computer Vision and Pattern Recognition,* Alaska, 2008

[14] M. de La Gorce, D.J. Fleet, and N. Paragios, "Model-Based 3D Hand Pose Estimation from Monocular Video," *IEEE Transactions On Pattern Analysis And Machine Intelligence*, 2011

[15] Chutisant Kerdvibulvech, Hideo Saito, "Model-Based Hand Tracking by Chamfer Distance and Adaptive Color Learning Using Particle," Filter *EURASIP Journal on Image and Video Processing* 2009

[16] E. Kollorz and J. Hornegger , "Gesture recognition with a time-of-flight camera," *Dynamic 3D Imaging (Workshop in Conjunction with DAGM'07)*

[17] Iason Oikonomidis, Nikolaos Kyriazis, Antonis A. Argyros, "Efficient model-based 3D tracking of hand articulations using Kinect," in *British Machine Vision Conference (BMVC) 2011*.

[18] S. Malassiotis, F. Tsalakanidou, N. Mavridis, V. Giagourta, N. Grammalidis, M.G Strintzis, "A face and gesture recognition system based on an active stereo sensor," *In: Proceedings 2001 ICIP*, Thessaloniki, Greece, 7-10 Oct. 2001, vol.3. (2001) 955–8

[19] S. Perrin, A. Cassinelli, M. Ishikawa, " Gesture recognition using laser-based tracking system," In: *Proceedings. FGR2004*, Seoul, South Korea, 17-19 May 2004, IEEE Comput. Soc (2004) 541–6

[20] P. Breuer, C. Eckes, S. M¨uller, "Hand Gesture Recognition with a novel IR Time-of-Flight Range Camera - A pilot study," *Proceedings of Mi-rage 2007, Computer Vision / Computer Graphics Collaboration Techniques and Applications*, pp.247–260

[21] Iason Oikonomidis, Nikolaos Kyriazis, Antonis A. Argyros," Full DOF Traking of Human Interacting with an Object by Modeling Occlusions and Physical Constraints," *IEEE International Conference on Computer Vision,* 2011, pp. 2088-2095

[22] J. R. New, E. Hasanbelliu, M. Aguilar, "Facilitating User Interaction with Complex Systems via Hand Gesture Recognition," *In Proc of Southeastern ACM Conf*., Savannah 2003

[23] M. H. Yang, N. Ahuja, M. Tabb, "Extraction of 2-D Motion Trajectories and its Application to Hand Gesture Recognition," *in PAMI.*, 29(8) (2002) 1062–1074

[24] Z. Mo, J.P. Lewis, U. Neumann, "Smartcanvas: a gesture-driven intelligent drawing desk system," *In 10th International Conference on Intelligent User Interfaces*, ACM Press (2005): 239-43

[25] J Martin, V. Devin, J.L. Crowley, "Active hand tracking," *3rd. International Conference on Face & Gesture Recognition*, IEEE Computer Society (1998): 575

[26] R Kjeldsen., J. Kender, "Toward the use of gesture in traditional user interfaces," *International Conference on Automatic Face and Gesture Recognition* (1996): 151–56

[27] R.G. O'Haga., A. Zelinsky, S. Rougeaux, "Visual gesture interfaces for virtual environments," *Interacting with Computers* 14 (2002): 231–50

[28] E. Sánchez-Nielsen, L. Antón-Canalísvisual, "Interaction through Hand Gesture Recognition using Hausdorff Matching," *Wseas Transactions On Computers*, Issue 2, Volume 3, April 2004, ISSN 1109-2750, pp 353-359

[29] Sue Han Lee, Soon Nyean Cheong, Chee Pun Ooi, Wei Heng Siew, "Real Time FPGA Implementation of Hand Gesture Recognizer System," *Recent Researches in Computer Science*, ISBN: 978-1-61804-019-0, pp 217-222

[30] F.M. Mona, M.R. Mursi Ghazy, Assassa Abeer Alhumaimeedy, Khaled Alghathbar, "Automatic Human Face Counting in Digital Color Images," *Proceedings of the 8th Wseas International Conference on Signal Processing, Robotics and Automation*, pp 269-275

[31] L. Bretzner, I. Laptev,T. Lindeberg, "Hand gesture recognition using multiscale color features, hieracrchichal models and particle filtering,"

in Proceedings of Int. Conf. on Automatic face and Gesture recognition, Washington D.C., May 2002

[32] M. Black, D. Jepson, “Eigen tracking: Robust matching and tracking of articulated objects using a view-based representation,” *In European Conference on Computer Vision*, 1996.

[33] C C Wang, K C Wang, “Hand Posture recognition using Adaboost with SIFT for human robot interaction,” Springer Berlin, ISSN 0170-8643, Volume 370, 2008

[34] N. Dardas, Qing Chen, N.D. Georganas, E.M Petriu, “Hand gesture recognition using Bag-of-features and multi-class Support Vector Machine,” *Haptic Audio-Visual Environments and Games (HAVE)*,2010

[35] Yuelong Chuang, Ling Chen, Gangqiang Zhao, Gencai Chen, “Hand Posture Recognition and Tracking Based on Bag-of-Words for Human Robot Interaction,” *IEEE International Conference on Robotics and Automation* Shanghai International Conference Center May 9-13, 2011, Shanghai, China

[36] G. Simion, V. Gui, M. Otesteanu, “A Compositional Tehnique for Hand Posture Recognition: New Results," *Wseas Transactions on Communications* 8 (8) (2009): 805-21

[37] Michael Donoser, Horst Bischof, “Real Time Appearance Based Hand Tracking Pattern,” *ICPR* 2008

[38] R. Lienhart and J. Maydt, “An extended set of Haar-like features for rapid object detection,” *in Proc. IEEE Int. Conf. Image Process*., 2002 vol. 1, pp. 900–903

[39] A.L. C. Barczak, F. Dadgostar, “Real-time hand tracking using a set of co-operative classifiers based on Haar-like features,” *Res. Lett Inf. Math Sci*., 2005, Vol. **7**, pp 29-42

[40] Qing Chen , N.D. Georganas, E.M Petriu, “Real-time Vision based Hand Gesture Recognition Using Haar-like features,” *IEEE Transactions on Instrumentation and Measurement*, 2007

[41] D.G. Lowe, “Object recognition from local scale-invariant features,” *Proceedings of the International Conference on Computer Vision*, 1999,Vol 2, pp. 1150–1157

[42] H. Bay, A. Ess, T. Tuytelaars and L. Van Gool, “SURF: Speeded up robust features,” in *CVIU*, 2008, 110(3), pp. 346-359

[43] V. Bruce, P. R. Green, M.A Georgeson, “*Visual Perception: Physiology, Psychology, and Ecology*,” Psychology Press East Sussex, UK, 3rd edition 1996

[44] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. Van Gool, “A comparison of affine region detectors,” *International Journal of Computer Vision*, 65(1-2):43–72, 2005

[45] Sharma, R., Huang, T. S., Pavovic, V. I., Zhao, Y., Lo, Z.,Chu, S., Schulten, K., Dalke, A., Phillips, J., Zeller, M. & Humphrey, W., Speech/Gesture Interface to a Visual Computing Environment for Molecular Biologists, In: *Proc. of ICPR’96,* Vol 2, pp 964-968

[46] Gandy, M., Starner, T., Auxier, J. & Ashbrook, D. “The Gesture Pendant: A Self Illuminating, Wearable, Infrared Computer Vision System for Home Automation Control and Medical Monitoring”. Proc. of *IEEE Int. Symposium on* Wearable Computers. (2000), 87-94

[47] S. M. Goza, R. O. Ambrose, M. A. Diftler, I. M Spain,” Telepresence Control of the NASA/DARPA Robonaut on a Mobility Platform,” *In: Conference on Human Factors in Computing Systems.* ACM Press, (2004) 623–629

[48] A. Malima, E. Ozgur, M. Cetin “A fast algorithm for Vision based hand gesture recognition for robot control,” *14th IEEE conference on Signal Processing and Communications Applications*, April 2006

[49] D. Stotts, J. Smith, M, K. Gyllstrom Facespace: Endoand Exo-Spatial Hypermedia in the Transparent Video Facetop, *In: Proc. of the Fifteenth ACM Conf. on Hypertext & Hypermedia*. ACM Press, (2004) 48–57

[50] G. M. Smith, M. C. Schraefel, “The Radial Scroll Tool: Scrolling Support for Stylus-or Touch-Based Document Navigation,” *In Proc. 17th ACM Symposium on User Interface Software and Technology*. ACM Press, (2004) 53–56

[51] C. Valli, C. Lucas, “*Linguistics of American Sign Language: An Introduction*," Washington, D. C.: Gallaudet University Press, (2000)

[52] A. Martinez., B. Wilbur, R. Shay, A. Kak, “Purdue RVL-SLLL ASL Database for Automatic Recognition of ASL,” *In IEEE Int. Conf. on Multimodal Interfaces*, (2002) 167–172

[53] T. Starner, J. Weaver, A. Pentland, ”Real-Time American Sign Language Recognition using Desk and Wearable Computer Based Video,” *PAMI*, 20(12) (1998) 1371–1375

[54] C. Vogler, D. Metaxas, “A Framework for Recognizing the Simultaneous Aspects of American Sign Language” *Comp. Vision and Image Understanding*, 81(3) (2001) 358–384

[55] M. Waldron, “Isolated ASL Sign Recognition System for Deaf Persons,” *IEEE Transactions on Rehabilitation Engineering*, 3(3) (1995) 261–271

[56] Dong Guo Yonghua, “Vision-Based Hand Gesture Recognition for Human-Vehicle Interaction,” *International Conference on Control, Automation and Computer Vision*, 1998

[57] C.A. Pickering, A. Burnham, K. J. Richardson, M. J. Jaguar “A research Study of Hand Gesture Recognition Technologies and Applications for Human Vehicle Interaction,” *3rd Conference on Automotive Electronics*, 2007

[58] P. J. Wachs, H. I. Stern, Y. Edan, M. Gillam, J. Handler, C. Feied, M. Smith, ”A Gesture-based Tool for Sterile Browsing of Radiology Images,” *Journal of the American Medical Informatics Association*, 2008

[59] http://www.eyesight-tech.com/news-and events /#news