IDIAP RESEARCH REPORT

Gestures for Multi-Modal Interfaces: A review

Sébastien Marcel

IDIAP–RR 02-34

September 2002

first revision : October 2000
second revision : September 2002

Dalle Molle Institute for Perceptual Artificial Intelligence • P.O.Box 592 • Martigny • Valais • Switzerland

phone +41 − 27 − 721 77 11
fax +41 − 27 − 721 77 12
e-mail secretariat@idiap.ch
internet http://www.idiap.ch

IDIAP Research Report 02-34

Gestures for Multi-Modal Interfaces: A review

Sébastien Marcel

September 2002

first revision : October 2000
second revision : September 2002

Abstract. This document presents a review on gestures for multi-modal interfaces and focus on hand gestures. It first introduces the role that the gesture modality plays in human communication. It then describes different types of gestures. Finally, it gives an overview of many techniques for the recognition of hand gestures.

1 Introduction

Nowadays, computers are more and more easy to use, thanks in particular to ergonomic and intuitive interfaces mainly based on the screen, the keyboard and the mouse. The user is in physical contact with the interactive system. The touch screen is used to select an area on the screen, the mouse to navigate in a menu, or a keyboard short cut to quickly access a functionality. These devices are either limited in speed (one will prefer the keyboard instead of the mouse for a fast action) or in easy of use (the keyboard short cuts are not intuitive and ask for a memory effort). We must thus imagine richer user interfaces which do not impose a physical contact.

A new form of interaction, based on techniques of speech recognition, uses the vocal channel and allows the user to control systems integrating an interpretation function. Machines would be easier to use if we could control them through natural language or even through gestures. Thus, many researchers have begun to study the gesture channel, because it's a very expressive method of non-verbal communication which allows a more natural interaction and which complements the verbal channel. The interfaces of the future will be multi-modal interfaces, which will integrate the traditional techniques of interaction such as those based on the voice and the gesture.

We will study the role that the gesture modality plays in human communication in order to determine the various gesture representation models. The different techniques for gesture modeling, analysis and recognition, will then be introduced and the selected techniques for hand gesture recognition will be presented.

2 Non-verbal Communication

Humans use gestures consciously and sub-consciously for non-verbal, natural and expressive communication. The dynamic aspect of gestures allows the user to express a command in only one action, and variations in the dynamic aspect can specify parameters such as the command range or the objects relating to it.

Figure 1: Size expression

To refresh the mind let us look at a random list of examples of hand movements: praying (two flat hands up together), begging (flat hand), expressing anger (raising a fist), derogation (middle finger up), accusation (index pointing), live or die decisions in the Roman amphitheater (thumb up/down), hitch hiking (thumb up, hand moving sideways), legal and business transactions (handshake, judge hammering), waving and saluting, counting (fingers and/or hand), pointing to real and abstract objects and concepts (index, hand), conducting of an orchestra (variety of both gestures with arms and body), traffic control of cars and airplanes (hands flat pointing or moving), shaping of imagined objects (hands tracing out curves and shapes), martial arts, fighting (variety of movements of arms and body), dance (Balinese dancing), gesturing by singers (hand and body movements), stock exchange operations (various hand shapes), affective gestures (hand touching), rejective (index up moving left & right) / appreciative (hand clapping) gestures, game playing (hand signs to communicate with partner in card games), game scoring (cricket, basketball, soccer, rugby, football), dinner-table actions (commanding waiter to refill wine glass), positioning of real (remote or close) and abstract objects, control panel operations (mousing, steering a vehicle), moving, touching and interacting with objects, silent and non-verbal communication (shrugging, holding one's own earlobe, scratching), "Italianate" gestures (two hands open shaking), mimicry and pantomime (actions and objects are depicted with hand/body movements), sign language (a complete linguistic communication system).

Moreover, gestures are intuitive and often universal (confirmation or disapproval , localization, scale (Figure 1) or distance estimation).

Using gestures removes the cognitive overload relating to the training of a new device. Thus, gestures offer many applications a more natural, concise and efficient interaction than the mouse.

The analysis and recognition of gestures can be very useful in many application fields:

- robotic,
- augmented virtual reality[1] [43],
- multi-modal interfaces controlled by voice and gesture [99] [18],
- sporting images, conductors or choreographic sequences analysis [16],
- automatic translation of signs language [86].

2.1 Taxonomy of Hand Gestures

A conscious gesture is motivated by the intention to communicate or complete a task (to indicate, to reject, to take, to draw). It is the physical expression of a mental concept [90]. Thus, in human society, we communicate using the word, but also using our body, our hands and our eyes. We use some of their movements to express or emphasize an idea, feeling or attitude. For examples, we raise our shoulders when in doubt, or we point to the things that we wish to explore.

Non-verbal communication is done mainly using hand gestures. Thus we will focus on hands. Cadoz [14] defines three hand functions which he considers complementary and interlinked: the epistemic function, the ergotic function and the semiotic function.

- The Epistemic Function
  The hand is an organ of perception. The tactile-kinesthetic perception gives informations about the form, the orientation, the distance and the dimension of objects using touch or exploratory movements. The proprioceptive perception gives informations on the weight, the trajectories and the movements of objects.

- The Ergotic Function
  The hand interacts with the environment to transform it, for instance, to move or to deform an object.

- The Semiotic Function
  The hand is an organ which emits informations to the environment, i.e. to the visual or tactile perception of one or more communication agents. The man-machine communication preferably exploits the semiotic function [10].

  Various authors find deeper distinctions between gestures. McNeill [64] considers gestures as iconic[2], metaphoric[3] or as beat-like [4], whereas Kendon [49] differentiate autonomous gestures (speech independent) and gesticulations (associates to the speech).

[1] In increased virtual reality systems, information coming from the virtual world is superimposed on those of the real world.

[2] Iconic gestures are used to represent an object, an action or an event.

[3] Metaphoric gestures are used to illustrate an abstract concept.

[4] Beat-like gestures are used to mark the rhythm of the speech

Regarding the semiotic function, we can distinguish different categories giving a classification of a human gesture, starting from the less expressive to the more expressive:

1 command gestures,

2 co-verbal gestures.

3 sign language gestures [68],

Command gestures are used for intentional communication acts whose sense is simple and adapted to the context of their use. Thus, we often observe a use of command gestures in activities not adapted to the propagation of the sound waves (skin diving) and where the sound environment is strongly noisy (stock exchange operations or construction sites).

Co-verbal gestures[5] illustrates and complements the verbal channel. Indeed, gesture and speech are combined to transmit several informations simultaneously. For example, if the gesture channel, as we will see, is adapted to the transmission of spatial informations, it's not the same for speech. In the same way, the co-verbal gesture raises ambiguity between some words of natural language, and conversely, the content of the verbal message permits easier interpretation of a gesture.

Sign language forms a real language, with a syntax and even more allowing the dynamic creation of non standard signs. Sign language gestures are much more structured and complex than the natural gestures. In order to simplify our work, we are going to concentrate ourselves on the two others categories.

2.2 Command Gestures

A gestural interface uses command gestures as a conversational interface, but also uses handling gestures to act on real objects (robotic, remote-manipulation) or virtual objects (graphical interfaces, augmented virtual reality). A good example is the "Drag and Drop" which, like the mouse utilization, allows to manipulate graphical objects on the screen. The command gesture is independent from the verbal channel. Nevertheless, it can be useful in a co-verbal context; we talk then about symbolic or emblematic gestures.

2.3 Co-verbal Gestures

Co-verbal gestures can be integrated in multi-modal interfaces. Associated to the verbal channel, co-verbal gestures communicate spatial-temporal informations. They are subdivided in two big categories [64].

- symbolic or emblematic gestures are related to command gestures in co-verbal contexts. Their use are proper to socio-linguistic communities (diver, crane driver, etc.).
- illustrator gestures are associated to verbal channel. We can distinguish different sub-categories.
  - metaphoric gestures illustrate an abstract concept.
  - beat-like gestures mark the rhythm of a speech. Indeed, a gesture is often connected to speech to raise a point or to impose an opinion. But it is also the prerogative of Italian people. "Italianate" gestures are made of two hands open shaking.
  - deictic gestures are movements of pointing using the hand, the face or an artifact. The trace of their movements in space is generally similar to a line segment.
  - iconic gestures, different from the metaphoric gestures, are used to physically dimension an object, an action or an event. For example, the sentence "I fished a large trout like that" is accompanied by a gesture indicating the size of the fish. The iconic gestures can also be subdivided:

[5] We can consider the lip movement as the first co-verbal gesture.

* Spatiographic gestures indicate objects regarding the position of the speaker (in front of, on the right, on the left).
* Pictomimic gestures describe the shape of objects using geometrical primitives such as a line segment, an arc of a circle or a right angle.
* Kinemimic gestures picture an action associated with a lexical unit. For example, a speaker speaks about a road in zigzag, he will complement his speech by a gesture with the hand performing the same zigzag.

Thus, the configuration of the hand gives pictomimic informations, its movement adds kinemimic informations, and the localization of the hand gives spatiographic informations. The deictic and iconic gestures give us spatial informations to perform designation, localization and quantification tasks. Moreover, the dynamic of the movement provides repetition, succession, continuity or stop informations. Therefore, gestures give us temporal information.

In the majority of the cultures, time is located on an axis whose reference is the speaker. Generally, the future is placed in front of you, the past behind you and the present is on your feet. In the same way, on a frontal parallel axis to the body, an event t_1 which is placed on the left of an event t_2 generally indicates that t_1 occurs before t_2.

However, all the cultures do not represent the time in the same way. In the culture of the aborigines from Australia, for instance, the past is in front of you and the future is placed behind you. Logic is as follows: you can see the past because it is known, it is thus in front of you, but the future is not known, it is thus invisible for the eyes and consequently behind you.

Thus, we can choose a deictic gesture to specify the concept of past, present and future performing a gesture of pointing. A spatiographic gesture sets events in time (from the left to the right), a pictomimic gesture indicates the duration of the events, and a kinemimic gesture their unfolding in time.

A gesture is not a simple static position showing the hand under a certain configuration. A gesture is also a dynamic process which can be performed in a particular body space region (sequence of postures) or performed with a large movement in this body space (the hand moving in the body space describes a trajectory). It is thus difficult to reduce the movement to simple geometrical primitives such as lines, arcs or circles, [6], because we then lose all the expressibility of a natural gesture. The dynamic of a gesture is unforeseeable. Thus, problems arise, because the execution of a gesture can vary according to its author (stylistic problem) and according to the context, the mood or the tiredness of the person (situational problem) [69].

3 Gesture Modeling

A hand gesture realization can be seen as a stochastic process in the parameter gesture space during a suitable interval time [43]. Indeed, hand gestures are dynamic processes which follow a characteristic scheme in space and time. We distinguish three phases in the achievement of a simple gesture [49]:

1. the preparation (preparatory movement starting from a home position)
2. the kernel (hand postures, gesture trajectory),
3. the retraction (return movement to a rest position).

The properties of this scheme are universals and can be used to describe the majority of hand gestures (static, dynamic localized and dynamic in movement), except beat-like gestures.

It is necessary to differentiate static gestures, localized dynamic gestures and non-localized dynamic gestures. The static hand gesture is characterized by its posture, i.e. by a particular configuration

[6] Some gestures behave in time like sinusoids [21]. Indeed, many gestures are oscillatory movements that humans do in critical or potential dangerous situations

of the fingers and the palm. Dynamic gestures are characterized by the sequence of postures or the global shape of the movement.

For static gestures, the kernel of the Kendon's scheme is formed by a hand posture. The preparation phase brings the hand in a workspace and the retraction phase ends the gesture. It is the same for located dynamic gestures, but the kernel is formed by a sequence of hand postures. For the non-localized dynamic gestures, the kernel is formed by the trajectory realized during the gesture and possibly by the sequence of postures performed during the movement.

Thus, problems arise such as the temporal gesture segmentation regarding to non-intentional movements or the determination of the execution time interval of a gesture. Indeed, it is necessary to be able to distinguish isolated dynamic gestures, connected dynamic gestures and linked dynamic gestures [10]. Isolated gestures are the simplest to recognize because they are naturally segmented by gestural gaps. Connected gestures are gestures executed the one after the others without covering. For linked gestures, the end of a gesture could influence the beginning of a new gesture and conversely. It is possible to compare this phenomenon with co-articulation in speech. Indeed, when a speaker speaks, the phonemes (sound elements of the language) can be modified by the next phoneme which will be pronounced.

To simplify some of these problems, a gesture scheme including the rules for the segmentation has been drawn [74].

- Gestures are movements in three phases. From a rest position, they start with a slow movement. Then, they continue in a phase increasing in speed. After all, they end in a fast return to a rest position. This is the Kendon scheme.
- Hands have a particular configuration during the execution of the movement.
- The slow movements between rest positions are not gestures.
- Hand gestures are restricted to a specific workspace.
- Static hand gestures need a fixed time period to be recognized. The repetitive movements can be gestures.

The resolution of the segmentation problem for the dynamic gestures can be done by various trajectory-feature extraction techniques like pauses [15] or inflections and changing points [80].

3.1 The Hand

Hand has adaptabilities and a great dexterity [87]. Its natural character makes it advantageous in urgent situations which need an instinctive reaction. Its adaptability refers to its facility to be able to pass slowly and quickly from a function to another. Its dexterity enables it to perform a complex task in an optimal way. The hand is an object highly deformable which owes its qualities to the association of the muscles and the numerous articulations connecting the bones of its skeleton. Hand has, according to authors, between 27 and 29 degrees of freedom [77] [87]. The numerous muscles and tendons interconnections and interactions give a big complexity to the movement. Most of the muscular mass is related to the forearm with long tendons which transmit the force to the fingers. The hand is thus an excellent compromise of lightness, flexibility and force.

We know the length of the segments (Figure 2), as well as articulation angles respecting specific constant and dynamic constraints [77]. The fingers perform movements of inflection, extension, abduction and adduction (Figure 3). The inch is disunited from the palm. Indeed, the three degrees of freedom of the trapezoidometacarpal[7] articulation allows the inch to do a longitudinal rotation in opposition with the other fingers.

Gestures are not limited to hand gestures. Indeed, body gestures have also a significant place in the gesture channel. They complement hand gestures and they often give informations on the emotional

[7]the trapezoidometacarpal articulation is located at the base of the inch.

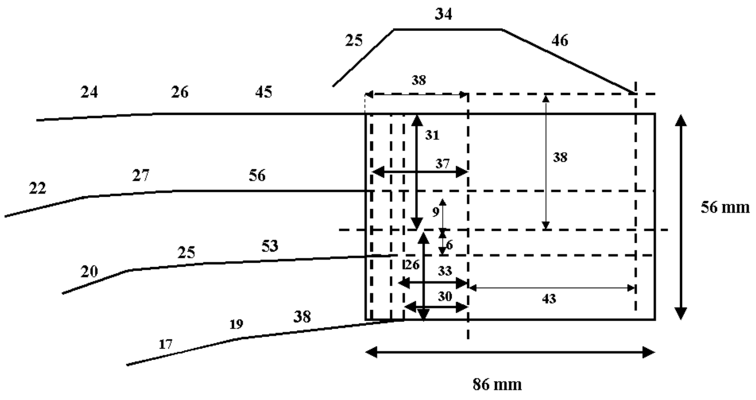

Figure 2: Anthropometric hand model

and cognitive state of the subject, through visible emotions on the face or the activity relating to the change of body posture.

3.2 The Body

Hand gestures tend to be centered on the body [87]. The most used body representations are models in three dimensions (Figure 4), generally for applications of images synthesis and computer animations [59].

However, we can be satisfied with simpler models in two dimensions. We can thus represent the body by the length of the segments of his skeleton (Figure 5) which are expressed according to the total height of the body. The proportions result from anthropometric studies [104].

The use of such a model could permit to analyze some body postures (sited, standed, dropped), but it could also help us to seek the hands and to interpret their gestures.

3.3 Discretization of Body Space

Some authors propose to divide the neighborhood of the body following zones where gestures evolve [64], in order to do a discretization of the representation space of gesture parameters. Indeed, all the gestures are not performed at the same locations and in the same way in the body space. This space can be built by taking the center of the body as a reference, the axis of body symmetry as principal axis and the maximum distances that the hands can reach as limits.

The body space is divided in rectangular zones numbered and concentric around the bust (Figure 6). The **center-center** is positioned on the chest. The **center**, slightly larger than the **center-center**, includes the shoulders. The periphery includes the center to contain the face and the hands at rest. The extreme periphery extends the periphery to the maximum distance that the hands can reach.

This space discretization is a first stage toward static gesture segmentation. Indeed, in order to be able to detect effectively the user intention to address himself to the system, it will be supposed that any recognized commands should result only from an intentional effort from the user [5]. At the time of such an effort, the hands move in a particular zone of the body space, named active window [61]. Moreover, postures recognized in a active window must belong to a specific vocabulary of intuitive commands which have little chance to be realized by inadvertence, and which claim a short and weak effort of the user to not to tire him. Indeed, a hand gesture uses arm muscles, and not always have support points.

Thus, while restricting to an active window and by defining a vocabulary, we limit the immersion syndrome, i.e. the impossibility for the user to interact in the real world when the computer collects

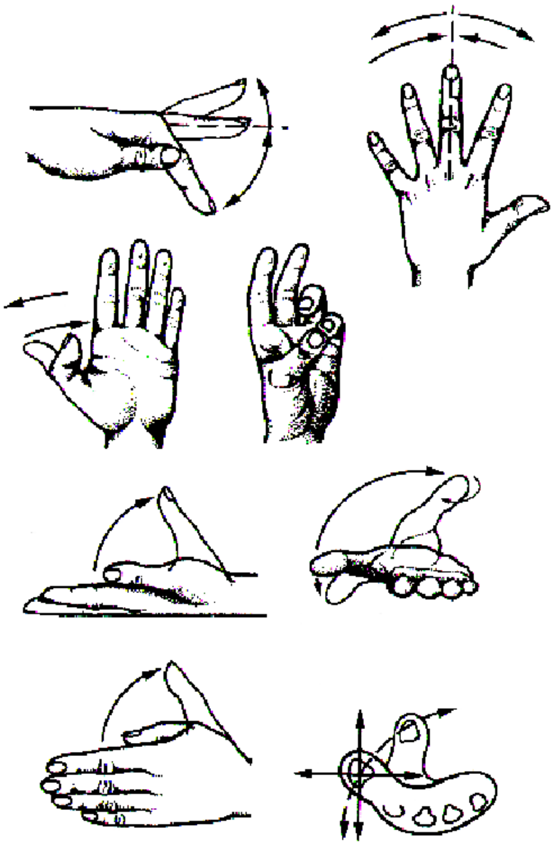

Figure 3: Fingers movements

and interprets all his movements. To illustrate this syndrome, let us imagine a user interacting with a computer, suddenly he stops and performs a non-intentional movement because a person came in the room and spoke to him. Unfortunately, the non-intentional movement is interpreted by the machine and can then cause an incident.

4 Gesture Analysis

Gesture analysis must take into account spatial and temporal dimension. It must implement relevant extraction feature techniques, regarding the source of the data and adapted to the recognition method. We can distinguish two types of gesture interfaces: instrumental interfaces and the pure gesture interfaces [5]. The instrumental interface analyzes the trace left by the gesture using a mouse or a pen and the pure gesture interface interprets the gesture with a numerical glove or of a camera.

We will place ourselves within the framework of a pure gesture interface which constitutes for us a major interest. Indeed, it is considered that only the sensors such as a numerical glove or a camera enable us to exploit the gesture channel to realize advanced man-machine interfaces.

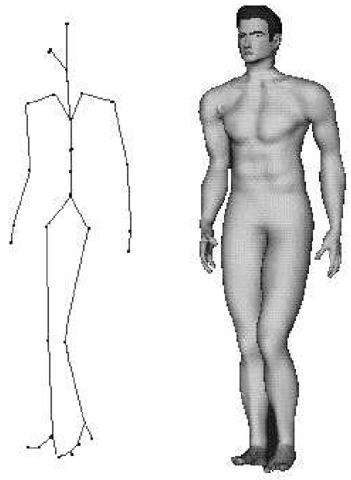

Figure 4: A three dimension body model

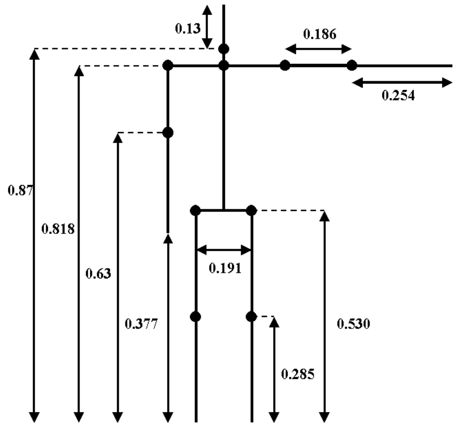

Figure 5: Anthropometric body model: The length of the segment are express according to the total height of the body

4.1 Glove based Analysis

Special devices containing gloves were developed to analyze the hand configuration. One of the application is the navigation into virtual worlds. The user is provided with a helmet, a numerical glove and is immersed in a virtual world. The glove allows the system to locate the hand in space and to reconstruct it in virtual reality. Thus, the subject can interact with the virtual scene, generally by taking objects, while pressing on buttons or by moving sliders. Numerical gloves use mechanical or optic sensors which translate the fingers inflection and abduction movements into electric signals. The hand position and orientation are given by electromagnetic or acoustic sensors. There are different kinds:

- Digital Data Entry Glove [39],
- VPL DataGlove [98],
- Exos Dexterous HandMaster (Little Inc, 1987),

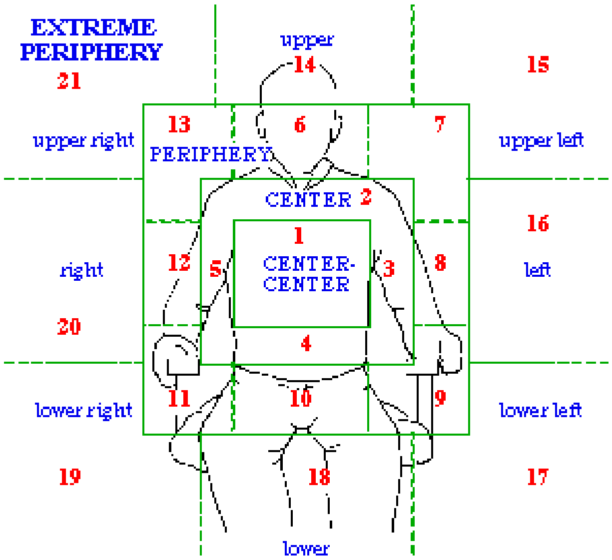

Figure 6: Space discretization for hand location

- Power Glove (Mattel Inc, 1989),
- Virtex CyberGlove [53],
- Space Glove [45].

The more used glove is DataGlove from VPL Research. It is composed by sixteen degrees of freedom (ten inflection informations and six position informations). However, the acquisition frequency of the numerical gloves remains weak and produces relatively noisy data. Moreover, the user is connected physically by cables to the computer, and it denatures the gesture performed. Numerical gloves are cumbersome and complex to connect. That's why, various techniques based on image processing and analysis were introduced to enrich the natural aspect of the interaction.

4.2 Image-based Analysis

The body posture or hand gesture analysis, based on the image is the most natural way for the construction of man-machine gesture interfaces, but it is also the most difficult to do. The image is a non-intrusive modality for gesture recognition [17]. New technologies allow to do image processing in real time. Thus, the interaction based on the image becomes possible. However, this approach raises many technical problems: to segment the body or the hand from the scene, to analyze postures and track the object segmented in a sequence of images.

4.2.1 Image Segmentation

To simplify the segmentation, some systems use passive markers (colored points), active markers (luminous diodes) or skin color. When the subject is motionless, the camera fixed and the background stable and uniform, we can then localize the markers in the image. Moreover, the hand has a particular color which could be discriminant. Thus, we use techniques based on color histograms [81] or on chromatic look-up tables (HSI[8] or RGB[9]) based on pixels of skin color. However, these localization techniques are limited by the variability of the skin color in different illumination conditions.

[8] tint, saturation and intensity.
[9] red, green and blue.

4.2.2 Posture Analysis

An approach to analyze postures and gestures, is to perform a three dimensions reconstruction. A 3D model is matched with one or more images to estimate the parameters (orientations, angles of the articulations). However, the problem of occlusion limits this technique when using only one sensor, which can be partially solved by the use of several cameras. Models used in three dimensions can be volumetric or skeletons.

Volumetric models are used to describe the visual appearance of the objects in computer animation, and thus, to perform an analysis by synthesizing the tracking and the posture [52] [105]. In those models, we use geometrical structures, such as cylinders, spheres or ellipses, to represent the form of certain parts of the body (fingers, arm, forearm, legs) [72] [28] [30] [38] [71]. The parameter estimation of these models is complex. Indeed, it is necessary to be able to consider the initial parameters and to update them during time. This method is very expensive and is difficult to implement in real time. Moreover, the knowledge of the exact posture of the hand, for example, is useless for communication gestures [74], but is well adapted for manipulation gestures.

Skeleton models are based on the morphological and bio-mechanical characteristics [95] [91] to represent the articulatory segments and angles [2] [78] [76] [97].

Another approach of analysis is based on a similarity measure with elements of a set of established postures or gestures. Many methods use deformable 2D hand or body models [56] [50] [22] [37], in which we adjust and associate a pattern with an existing model using an interpolation function. These patterns generally consist of a succession of points coming from the contours, from the finger tips curve extraction [19] [26] or from the calculation of the Gabor's filters in particular points of the image [93] [94]. Deformable models can provide sufficient informations for the handling and communication gestures, but they require a high image resolution and have a high computational cost. Other methods are correlation techniques of a image transformation such as a decomposition into eigen-vectors and the accumulation of spatial-temporal informations. These techniques are much easier to compute, but they give less precise informations, which makes them less adapted to exploit handling gestures.

4.2.3 Tracking

We specified that a gesture is a process as much temporal than spatial. This implies to be able to know at each time, i.e. in each image, the position of the hand previously recognized, in order to establish its trajectory. There are various techniques to predict the next location of the hand or to estimate the parameters of a 2D or 3D model, by using a dynamic model [78] [55] or Kalman's filters [97] [73] [106] [63].

5 Gesture Recognition

Gesture recognition consists in classifying the data resulting from the analysis phase. Gesture recognition (dynamic gestures) requires an invariance in time, whereas on its side, the recognition of postures must be invariant in scale and in orientation. Natural gesture recognition is difficult because the gestures have a very great variability from one person to another. Moreover, the gesture recognition suffers from problems of co-articulation. For example, the movement which the author of the gesture plans to perform after a gesture can influence the end of the current gesture. In the same way, during the execution of a localized dynamic gesture (sequence of postures), an transitory non-intentional posture can be recognized, which could alter the recognition.

5.1 Static Gesture Recognition

Posture recognition can be based on gestural action models (archive of movements in an image [29] [8] formed by the accumulation of the movement), on contours [27] [85] [48], on size functions [96] [36], on mathematical moments [89] (signatures [11], Zernike's moments [44], steerable filters [33]), on eigen-vectors [66] or on Gabor's coefficients [93]. Zernike's moments [83] have invariant magnitudes in rotation and the orientation histograms [35] are invariants in lighting conditions. But, these techniques need an large image resolution to be discriminant.

Posture recognition can be done by a decision tree [15] and a research of the maximum of likelihood [79] where a distance measure is used to evaluate the correlation of a feature vector with posture examples [24]. Posture recognition can also be done by Neural Networks [53].

5.2 Dynamic Gesture Recognition

Features used to recognize gestures are numerous. We generally use the trace of the movement, acceleration, speed, polar speed, angular speed [17], the length of the bounding box, the overall length or angle [79]. These features, more or less invariant, are extracted in a window of fixed or variable size.

5.2.1 Dynamic Time Warping

This method uses techniques of dynamic programming to temporally align two gesture sequences of different duration [17]. The dynamic programming tries to solve a task of correspondence by searching the minimal cost of a path in a graph where the nodes generally represent the states of a gesture [103].

5.2.2 Neural networks

Gesture recognition can be done by artificial Neural Networks. When gestures have a fixed size (postures) or when gesture features are of finite length, it is generally possible, to use a multi-layer perceptron [60] [41]. Other authors will prefer to use Kohonen topological maps [9]. However, the gestures are not always fixed length, their duration can be very variable. Researchers then use recurrent Neural Networks [70] or TDNN (Time Delay Neural Networks) [99].

5.2.3 Hidden Markov Models

Hidden Markov Models (HMM) are intensively used in gesture recognition [86] [82]. They are stochastic processes [75] made up of hidden states, distributions of transitions between the states and distributions of emissions to emit observations. A HMM model is assign to each gesture to be recognized, then the parameters of these models are determine by training. The training of the HMM consists in calculating the probabilities of transitions and emissions by optimization of the maximum of likelihood through a set of training examples. For that, we use procedures of likelihood maximization like the Baum-Welch algorithm [6]. At the time of the recognition, the HMM model which has the greatest probability of having generated the gesture observed is selected and then the recognition is accomplished.

But first of all, it is necessary to be able to model the observations which, in gesture recognition, are the elementary components of a gesture (postures, trajectories) represented by vectors of features. To model the observations, we use Gaussian Mixtures [67], Neural Networks or quantization vector techniques (Geometrical moments, Zernike moments, Eigen-Vectors).

5.3 Existing Systems

5.3.1 Glove based Systems

An implementation of a virtual world combining hand gestures, texts, sounds and stereoscopic images exists [20]. This system simulates a room containing nonrigid objects which can be created or moved in real-time by two users.

Glove-talk [31] is an interface including the hand of a user and a voice synthesizer. It uses five artificial Neural Networks to recognize a vocabulary of two hundred and three hand gestures associated with words.

VirtualPanelArchitecture [88] combines a hardware and software control panel. The users can respectively turn, move and point virtual buttons, sliders and screens.

The pointing system VirtualEnd-Effector [100] is used to train and direct robots using hand gestures. It uses Neural Networks based on a skeleton transformation.

The GIVEN system [32] (Gesture-based Interaction in Virtual Environments) introduced techniques allowing the user to take and surround virtual objects and thus gets a more precise interaction. The use of tactile sensors was proposed to increase the sensitivity in virtual spaces.

One of the most recent concepts is an alternative between multimedia systems and virtual reality systems [54]. Such systems represents an virtual working environment containing objects and virtual control tools. Objects are manipulated on a real desk by various users collaborating on a same task.

In the CHARADE system [4], hand gestures are used to control navigation in a hypertext. It runs in real time and recognizes six command gestures. All these gestures include three phases (starting posture, dynamic phase, final posture). Discrimination between the various commands is done with the first two steps. However, the posture vocabulary is complex to realize, which does not facilitate the use of the system.

5.3.2 Image based Systems

- Systems using 3D hand models:
  An approach used in hand gesture recognition consists in building a three-dimensional model of the human hand. The model is matched with images. Then, the parameters corresponding to the orientation and to the angles of the articulations are considered and classified.

  Such a system exists [28]. It recovers human members from a sequence of images using an articulated cylindrical human model. The images are obtained on uniform black background from a monochrome camera. The matching of the model on the image is based on a perspective projection of the cylindrical model and on the comparison with the contours obtained in the image. The system then provides kinematic informations of the angles of the articulations.

  A cylindrical model was also developed at ATR Research laboratory [30]. The system uses stereoscopic cameras and an algorithm which partition an object in a hierarchical set of cylinders to model the human hand. The contours points extracted from the image are used to find the axis and the size associated with cylinders.

  DigitEyes, a complete hand gesture recognition system, uses a kinematic 3D cylindrical model with 27 degrees of freedom [78]. Finger tips are extracted from stereoscopic images on restricted background and are used as features to find the correspondence with the model. The tracking of the features and the estimation of the parameters of the model are done by minimization of the residual error. The system was designed to run a three-dimensional mouse application using only one camera and to estimate the angles of the articulations of the hand using stereoscopic cameras.

  A hand gesture analysis system based on a skeletal model in three dimensions was developed by Lee and Kunii [57]. This model has 27 degrees of freedom and incorporates constraints resulting from the kinematic of the hand to reduce the search of space of the parameters of the model. The system was used for the analysis of 16 symbols of the ASL[10]. It produces very weak errors, but the computing time is very high.

  In the same way, Kuch formalized a hand model with 26 degrees of freedom including six kinematic constraints of the hand [55]. This model can follow complex hand gestures in a full sequence of images coming from only one camera. The system was used for the tracking of ASL gestures and for applications such as “Virtual Guns”.

- Systems using markers or colored gloves:
  The human hand has a highly not-convex form and detecting its configuration starting from

[10] American Sign Language.

images is very difficult. One of the techniques is thus to place distinct colored markers on the hand and more generally on the finger tips.

Torige and Kono conceived a system which indicates the direction of the moving hand [92]. They use stereoscopic cameras, black gloves and colored markers on the shoulder, the elbow, the wrist and the finger tips. They calculate the position of the fingers and the parameters of movement to control a robot manipulator.

Davis and Shah [25] use also markers on finger tips. The system calculates the trajectories of the markers, then it uses them to determine the beginning and the end of a gesture. The system perform a temporal segmentation of seven hand gestures at four images per second.

Maggioni [58] uses the images from only one camera and a glove whose areas are differently colored. The system calculates several geometrical parameters and uses them to estimate the position and the orientation of the hand.

Cipolla, Okamoto and Kuno [19] use a glove, whose finger tips are marked, to determine the translation and the rotational movement of the hand in order to change the position and the orientation of a virtual object.

- Systems using the properties of the image:
  These systems are based on the extraction of certain features associated with hand posture images. This techniques can simply work on geometrical moments of the image (Hu moments [65], Fourier-Mellin moments [3], Zernike moments [89]) or with artificial Neural Networks. However, the purpose of these methods are not to consider the extracted parameters of the hand like the angles of the articulations, their objective is either to perform a tracking of the hand, or a classification of the posture.

  Darrell and Pentland developed a system working at ten images per second using a set of views [23]. Each hand gesture is represented by its own set of various views. These views are matched with the gestures of an image sequence by using temporal correlation techniques (Dynamic Time Warping).

  Segen [84] uses contours extraction techniques starting from simple silhouettes to distinguish in real-time ten distinct postures.

  Ahmad and Tresp [1] propose a Neural Network to classify a hand posture. This one is described by the polar coordinates of the finger tips and the center of the hand in the image. This Neural Network is conceived to deal with missing input features. If all the characteristics are present, it obtains an error rate of 5%. Ahmad also developed a three-dimensional real-time tracking system of the hand in complex environments [2]. He uses a histogram segmentation and three geometrical moments to extract and localize in the image the finger tips and the angles of the articulations.

  Starner and Pentland [86] use geometrical parameters of a uniform colored hand. Then, a Hidden Markov Model with five states performs the classification of some ASL gestures.

  Schlenzig, Hunter and Jain use Zernike moments as features [83]. These features are calculated from images of hand postures against a uniform background. A HMM model performs the recognition of six hand gestures at the rate of $\frac{1}{2}$Hz.

  Kjeldsen [51] conceived a system to control a window based interface using hand gestures. It uses a histogram segmentation, a Neural Network for the differentiation of postures and a set of rules for gesture recognition. Freeman and Roth [34] proposed a simple and fast system for posture recognition using histograms of local orientations. The system is robust to the local illumination variations, but requires a uniform black background.

  Turk [42] proposed a state based approach to gesture learning and recognition. Using spatial clustering and temporal alignment, each gesture is defined to be an ordered sequence of states in spatial-temporal space. The 2D image positions of the centers of the head and both hands of the user are used as features; these are located by a color based tracking method. From training data of a given gesture, the spatial information is first learned without doing data segmentation and alignment, and then the data is grouped into segments that are automatically associated with information for temporal alignment. The temporal information is further integrated to build a Finite State Machine (FSM) recognizer. Each gesture has a FSM corresponding to it. The computational efficiency of the FSM recognizers achieves real-time online performance. An experimental system was built that plays a game of "Simon Says" with the user.

Marcel [62] uses a hybrid approach of Neural Networks and Hidden Markov Models, called Input-Output Hidden Markov Models, to recognize four dynamic hand gestures. The skin color is filtered and the resulting pixels are processed to segment the hand from the image. Thus, the center of gravity of the hand is provided to the recognizer. This approach gives a good recognition rate and also achieves the rejection of non-trained gestures.

Bretzner [12] presents an algorithm and a prototype system for hand tracking and hand posture recognition. Hand postures are represented in terms of hierarchies of multi-scale color image features at different scales, with qualitative inter-relations in terms of scale, position and orientation. In each image, detection of multi-scale color features is performed. Hand states are then simultaneously detected and tracked using particle filtering, with an extension of layered sampling referred to as hierarchical layered sampling. Experiments have shown that the performance of the system is substantially improved by performing feature detection in color space and including a prior with respect to skin color. These components have been integrated into a real-time prototype system, applied to a test problem of controlling consumer electronics using hand gestures. In a simplified demo scenario, this system has been successfully tested by participants at two fairs during 2001.

- “Intelligent Surfaces” (Magic Boards and Digital Desks):
  The MagicBoard [7] project aims at augmenting a perfectly ordinary whiteboard-like surface with electronic capabilities, via a video projector and a pan-tilt-zoom camera. The user works on the board as in the usual way, drawing or writing with ordinary marker pens. Whenever he chooses, the user can "grab" an electronic copy of the things that have been drawn or written with the marker pen. This copy is projected back onto the board, precisely overlaying the original markings with the appropriate color. The physical ink may then be erased and the electronic version manipulated on the board's surface: it can be duplicated, moved, enlarged or reduced, printed, or hidden for a moment before being recalled. Meanwhile, the user may add to her designs with the marker pen as before. At any time, these new markings can be turned into digital form to merge with the electronic version of her work. An horizontal implementation, called the MagicTable, of the board exists. Red tokens (small disks made of plastic) are used instead of fingers.

  The DigitalDesk [101, 102] is a desk with a computer-controlled camera and projector above it. The camera sees where the user is pointing, and it reads portions of documents that are placed on the desk. The projector displays feedback and electronic objects onto the desk surface. This DigitalDesk adds electronic features to physical paper, and it adds physical features to electronic documents. The system allows the user to interact with paper and electronic objects by touching them with a bare finger (digit). Instead of "direct" manipulation with a mouse, this is tactile manipulation with a finger.

6 Conclusion

This document had as an ambition to introduce gestures and to provide the reader with an non-exhaustive overview of hand gesture recognition systems and applications.

Actual gesture interfaces are limited (small number and low complexity of gestures to recognize) and generally do not deal with bi-manual gestures [40] [13] [47] [46].

Multi-Modal interfaces of the future should achieve real-time hand gesture recognition. These interfaces will be able first to detect and identify persons and then to recognize many different hand gestures (including bi-manual hand gestures).

Acknowledgments The author wishes to thank D. Moore, J. Moore-Schulz and P. Wellner for fruitful discussions and collaboration.

References

[1] S. Ahmad and V. Tresp. Classification with missing and uncertain inputs. In *International Conference on Neural Networks*, volume 3, pages 1949–1954, 1993.

[2] Subutai Ahmad. A usable real-time 3d hand tracker. *Asilomar Conference on Signals, Systems and Computer*, (28), 1994. Interval Research Corporation, www.interval.com.

[3] H. Arsenaud and Y. Sheng. Properties of the circular harmonic expansion for rotation-invariant pattern recognition. In *Applied optics*, volume 25 of *18*, pages 3225–3229, september 1986.

[4] Thomas Baudel and M. Beaudouin-Lafon. Charade: Remote control of objects using free-hand gestures. *Communications of the ACM*, 36(7):28–35, 1993.

[5] Thomas Baudel and Annelies Braffort. Reconnaissance de gestes de la main en environnement reel. *Actes de Informatique'93 - L'interface des mondes reels et virtuel*, pages 207–216, 1993. Montpellier.

[6] L. Baum. An inequality and associated maximization technique in statistical estimation of probabilistic functions of markov processes. *Inequalities*, 3:1–8, 1972.

[7] Magicboard, 1999. CLIPS-IMAG, http://iihm.imag.fr/demos/magicboard/.

[8] A. Bobick and J.W. Davis. Real-time recognition of activities using temporal templates. In *International Conference on Automatic Face and Gesture Recognition*, February 1996.

[9] Klaus Boehm, Wolfgang Broll, and Michael Sokolewicz. Dynamic gesture recognition using neural networks: A fundament for advanced interaction construction. *SPIE, Conference Electronic Imaging Science and Technology*, February 1994. San Jose California, USA.

[10] Annelies Braffort. *Reconnaissance et comprehension de gestes, application a la langue des signes.* PhD thesis, Paris XI, Juin 1996.

[11] Ulrich Brckl-Fox. Real-time 3d interaction with up to 16 degrees of freedom form monocular video image flow. In *International Workshop on Automatic Face and Gesture Recognition*, pages 172–178, June 1995. Zurich, Switzerland.

[12] L. Bretzner, I. Laptev, and T. Lindeberg. Hand gesture recognition using multi-scale colour features, hierarchical models and particle filtering. In *Conference on Automatic Face and Gesture Recognition*, 2002.

[13] W. Buxton and B. Myers. A study in two-handed input. In *Proceedings of the Conference on Human Factors in Computing Systems*, pages 321–326, 1986.

[14] C. Cadoz. Le geste canal de communication Homme/Machine - la communication instrumentale. *Techniques et Sciences Informatiques*, 13:31–61, 1994.

[15] J.M. Cagin. *Une tude sur la reconnaissance de formes dynamiques - Application la langue des signes.* ENSTA Paris, 1993. Projet de fin d'tude.

[16] Lee Campbell and Aaron F. Bobick. Recognition of human body motion using phase constraints. In *International Conference on Computer Vision*, pages 624–630, 1995. Cambridge MA, MIT TR number 309.

[17] Lee W. Campbell, David A. Becker, Ali Azarbayejani, Aaron F. Bobick, and Alex Pentland. Invariant features for 3d gesture recognition. In *International Workshop on Automatic Face and Gesture Recognition*, pages 157–162, October 1996. Killington, Vermont.

[18] J. Cassell, T. Bickmore, M. Billinghurst, L. Campbell, K. Chang, H. Vilhjlmsson, and H. Yan. Embodiment in conversational interfaces: Rea. In *ACM Conference on Human Factors in Computing Systems CHI'99 Conference Proceedings, Pittsburgh, PA*, 1999.

[19] R. Cipolla, Y. Okamoto, and Y. Kuno. Robust structure form motion using motion parallax. In *International Conference on Computer Vision*, pages 374–382, 1993.

[20] C. Codella, R. Jalili, L. Koved, and al. Interactive simulation in a multi-person virtual world. In *ACM Conference on Human Factors in Computing Systems*, volume 35, pages 329–334, 1992. CHI'92.

[21] Charles J. Cohen, Lynn Conway, and Koditschek. Dynamical system representation, generation, and recognition of basic oscillatory motion gestures. In *International Workshop on Automatic Face and Gesture Recognition*, pages 60–65, October 1996. Killington, Vermont.

[22] T.F. Cootes, C.J. Taylor, D.H. Cooper, and J. Graham. Active shape models - their training and applications. *Computer Vision and Image Understanding*, 61:38–59, January 1995.

[23] T. Darrell and A. Pentland. Space-time gestures. In *Computer Vision and Pattern Recognition CVPR'93*, pages 335–340, 1993.

[24] Trevor J. Darrell and Alex P. Pentland. Recognition of space-time gestures using a distributed representation. Technical Report 197, M.I.T. Media Lab Perceptual Computing, 1992. Annee a verifier.

[25] J. Davis and M. Shah. Gesture recognition. Technical report, Department of Computer Science, University of Central Florida, 1993. CS-TR-93-11.

[26] J. Davis and M. Shah. Recognizing hand gestures. In *ECCV94*, pages A:331–340, 1994.

[27] R Deriche. Using canny's criteria to derive a recursively implemented optimal edge detector. *International Conference on Computer Vision*, 1(2):167–187, May 1987.

[28] A.C. Downton and H. Drouet. Image analysis for model-based sign language coding. *International Conference on Image Analysis and Processing*, (6):637–644, 1991. Progress in image analysis and processing II.

[29] I. Essa and S. Pentland. Facial expression recognition using a dynamic model and motion energy. In *IEEE International Conference on Computer Vision*, 1995.

[30] M. Etoh, A. Tomono, and F. Kishino. Stro-based description by generalized cylinder complexes from occluding contours. *Systems and Computers in Japan*, 22(12):79–89, 1991.

[31] S.S. Fels and G.E. Hinton. Glove-talk: A neural networks interface between a data-glove and a speech synthesizer. *IEEE Transactions on Neural Networks*, 4:2–8, January 1993.

[32] M. Figueiredo, K. Bhm, and J. Teixera. Advanced interaction techniques in virtual environments. *Computers and Graphics*, 17(6):655–661, 1993.

[33] William T. Freeman and Edward H. Adelson. The design and use of steerable filters. In *IEEE Transaction Pattern Analysis and Machine Intelligence*, volume 13 of *9*, pages 891–906, september 1991. MIT - Media Lab and Dept of Brain and Cognitive Sciences.

[34] William T. Freeman and Craig D. Weissman. Television control by hand gestures. In *International Workshop on Automatic Face and Gesture Recognition*, pages 179–183, June 1995. Zurich, Switzerland.

[35] W.T. Freeman and M Roth. Orientation histograms for hand gesture recognition. In *International Workshop on Automatic Face and Gesture Recognition*, pages 296–301, june 1995. Zurich, Switzerland.

[36] P. Frosini. Measuring shapes by size functions. In *Proceedings SPIE on Intelligent Robotic Systems*, 1991. Boston.

[37] D.M. Gravila. Hermite deformable contours. In *International Conference on Pattern Recognition*, 1996. ICPR'96, Vienna.

[38] D.M. Gravila and L.S. Davis. Towards 3-d model-based tracking and recognition of human movement: a multi-view approach. In *International Conference on Automatic Face and Gesture Recognition*, number 1, pages 272–277, June 1995.

[39] Gary J. Grimes. Digital data entry glove interface device. Technical report, Bell Telephone Laboratories, November 1987. United States Patent 4, 414, 537, Murray Hill, NJ, November 8, 1983.

[40] Y. Guiard. Asymetric division of labor in human skilled bimanual action : The kinematic chain as a model. *Journal of Motor Behavior*, 19(4):486–481, 1987.

[41] P.A. Harling. Gesture input using neural networks. Technical report, Dept of Computer Science University of York, 1993. BSc degree in Computer Science.

[42] P. Hong, Turk M., and Huang T. Gesture modeling and recognition using finite state machines. In *Conference on Automatic Face and Gesture Recognition*, 2000.

[43] Thomas S. Huang, Vladimir I. Pavlovic, and Rajeev Sharma. Gestural interface to a visual computing environment for molecular biologists. In *International Workshop on Automatic Face and Gesture Recognition*, pages 30–35, October 1996. Killington, Vermont.

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

[45] W Industries. *Press release*. W Industries, 1991. Leicester, UK.

[46] P. Kabbash, W. Buxton, and A. Sellen. Two-handed input in a compond task. In *Proceedings of the Conference on Human Factors in Computing Systems*, pages 417–423, 1994.

[47] P. Kabbash, I.S. MacKenzie, and W. Buxton. Human performance using computer input devices in the preferred and non-preferred hands. In *Proceedings of the Conference on Human Factors in Computing Systems*, pages 474–481, 1993.

[48] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. *International Conference on Computer Vision*, pages 259–268, 1997.

[49] A. Kendon, J.-L. Nespoulous, P. Peron, and A.R. Lecours. Current issues in the study of gesture. *The Biological Foundations of Gestures: Motor and Semiotic Aspects*, pages 23–47, 1986. Lawrence Erlbaum Assoc.

[50] C. Kervrann and F. Heitz. Learning structure and deformation modes of nonrigid objects in long image sequences. In *International Workshop on Automatic Face and Gesture Recognition*, June 1995.

[51] R. Kjeldsen. Visual hand gesture interpretation. *IEEE Computer Society Workshop on Non-Rigid and Articulate Motion*, 1994. Austin, TX.

[52] R. Koch. Dynamic 3d scene analysis through synthetic feedback control. In *IEEE Pattern Analysis and Machine Intelligence*, volume 15, pages 556–569, 1993.

[53] James Kramer and Larry Leifer. The talking glove: An expressive and receptive "verbal" communication aid for deaf, deaf-blind and nonvocal. Technical report, Department of Electrical Engineering, 1989. Stanford University.

[54] M.W. Krueger. Environmental technology: Making the real world virtual. *Communications of the ACM*, 36:36–37, July 1993.

[55] J.J. Kuch and T.S. Huang. Vision based hand modeling and tracking. In *International Conference on Computer Vision*, June 1995. Cambridge, MA.

[56] A. Lanitis, C.J. Taylor, T.F. Cootes, and T. Ahmed. Automatic interpretation of human faces and hand gestures using flexible models. In *International Workshop on Automatic Face and Gesture Recognition*, pages 98–103, June 1995. Zurich, Switzerland.

[57] J. Lee and T.L. Kunii. *Constraint-based hand animation*, pages 110–127. Models and techniques in computer animation, 1993. Tokyo: Springer-Verlag.

[58] C. Maggioni. A novel gestural input device for virtual reality. In *IEEE Annual Virtual Reality International Symposium*, pages 118–124, 1993.

[59] N. Magnenat-Thalmann and D. Thalman. *Computer Animation: Theory and Practice*. New York: Springer Verlag, 2nd rev edition, 1990.

[60] S. Marcel. Hand posture recognition in a body-face centered space. In *Conference on Human Factors in Computing Systems CHI'99*, 1999.

[61] S. Marcel, O. Bernier, and D. Collobert. Reconnaissance de la main pour les interfaces gestuelles. In *CORESA'99*, 1999.

[62] S. Marcel, O. Bernier, J.E. Viallet, and D. Collobert. Hand gesture recognition using input/output hidden markov models. In *Conference on Automatic Face and Gesture Recognition*, 2000.

[63] Jerome Martin, Vincent Devin, and James L. Crowley. Active hand tracking. *3rd IEEE Conference on Automatic Face and Gesture Recognition*, 1(3):573–578, April 1998. FG'98, Nara, Japan, 14-16 April 1998.

[64] D. McNeill. *Hand and Mind: What gestures reveal about thought*. Chicago Press, 1992.

[65] H. Ming-Kuel. Visual pattern recognition by moment invariants. *IRE Transaction on Information Theory*, 2:179–187, 1962.

[66] Baback Moghaddam and Alex Pentland. Maximum likelihood detection of faces and hands. In *International Workshop on Automatic Face and Gesture Recognition*, pages 122–128, June 1995. Zurich, Switzerland.

[67] Baback Moghaddam and Alex Pentland. Probalistic visual learning for object representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):696–710, July 1997.

[68] B. Moody. *La langue des signes. Histoire et grammaire*, volume 1. Paris, 1983.

[69] P. Morrel-Samuels. Clarifying the distinction between lexical and gestural commands. *International Journal of Man-Machine Studies*, 32:581–590, 1990.

[70] K. Murakami and H. Taguchi. Gesture recognition using recurrent neural networks. In *Conference on Human Interaction*, pages 237–242, 1991. Fujitsu Lab, Conference on Human Factors in Computing Systems CHI'91, ACM, New Orlean, Louisiana.

[71] O'Rourke and N.L. Badler. Model-based image analysis of human motion using constraint propagation. In *IEEE Transaction Pattern Analysis and Machine Intelligence*, volume 2, pages 522–536, 1980.

[72] Alex Pentland, Ali Azarbayejani, and Wren Christopher. Real-time 3d tracking of human body. Technical Report 374, M.I.T. Media Lab Perceptual Computing, May 1996. Appears in Proceedings of IMAGE'COM 96, Bordeaux, France.

[73] Alex Pentland and Andrew Liu. Modeling and prediction of human behavior. Technical Report 433, M.I.T. Media Lab Perceptual Computing, September 1995.

[74] Francis K.H. Quek. Toward a vision-based hand gesture interface. Technical report, University of Illinois at Chicago, August 1994. Virtual Reality Software and Technology Conference.

[75] Lawrence R. Rabiner. A tutorial on Hidden Markov Models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–285, February 1989.

[76] J. Rehg and T. Kanade. Model-based tracking of self-occluding articulated objects. In *ICCV95*, pages 612–617, 1995.

[77] James M. Rehg. *Visual Analysis of High DOF Articulated Objects Application to Hand Tracking*. PhD thesis, Carnegie Mellon University, April 1995.

[78] James M. Rehg and Takeo Kanade. Digiteyes: Vision-based human hand tracking. Technical report, Carnegie Mellon University, december 1993. Proceedings of European Conference on Computer Vision, May 1994, Stockholm, Sweden.

[79] D. Rubine. *The automatic recognition of gestures*. PhD thesis, Carnegie Mellon University, 1991.

[80] H. Sagawa, H. Sakou, and M. Abe. Sign language translation system using continuous dp matching. In *MAV'92 - IAPR Workshop on Machine Vision Applications*, pages 339–342, 1992. Tokyo.

[81] David Saxe and Richard Foulds. Toward robust skin identification in video images. In *International Workshop on Automatic Face and Gesture Recognition*, pages 379–384, October 1996. Killington, Vermont.

[82] J. Schlenzig, E. Hunter, and R. Jain. Recursive identification of gesture inputs using hidden markov models. In *WACV94*, pages 187–194, 1994.

[83] J. Schlenzig, E. Hunter, and R. Jain. Vision based hand gesture interpretation using recursive estimation. In *Proceedings of the 28thAsilomar Conference on Signals, Systems and Computer*, 1994.

[84] J. Segen. Controlling computers with gloveless gestures. In *Virtual Reality Systems*, April 1993.

[85] Jun Shen and Serge Castan. An optimal linear operator for edge detection. *Image Vision and Computing*, 1986.

[86] Thad Starner and Alex Pentland. Real-time american sign language recognition form video using hidden markov models. Technical Report 375, M.I.T. Media Lab Perceptual Computing, 1995. ISCV'95.

[87] David Joel Sturman. *Whole-hand Input*. PhD thesis, Massachusetts Institute of Technology, February 1992.

[88] S.A. Su and R. Furuta. Virtual panel architecture: a 3d gesture framework. *IEEE Annual Virtual Reality International Symposium*, pages 387–393, 1993.

[89] M. Teague. Image analysis via the general theory of moments. *Journal of the Optical Society of America*, 70:920–930, 1980.

[90] S. Thieffry. *"Hand gestures"in The Hand*, pages 488–492. Tubiana, R., 1981. Philadelphia, PA: Sanders, 1981.

[91] D. Thompson. Biomechanics of the hand. *Perspectives in Computing*, 1:12–19, October 1981.

[92] A. Torige and T. Kono. Human-interface by recognition of human gestures with image processing recognition of gesture to specify moving directions. *IEEE International Workshop on Robot and Human Communication*, pages 105–110, 1992.

[93] Jochen Triesch and Christoph Malsburg. Robust classification of hand posture against complex backgrounds. In *Proceedings of the second International Conference on Automatic Face and Gesture Recognition*, pages 170–175, october 1996. Killington, Vermont, USA.

[94] Jochen Triesch and Christoph Malsburg. A gesture interface for human-robot-interaction. In *Proceedings of the third International Conference on Automatic Face and Gesture Recognition*, pages 546–551, April 14-16 1998. Nara, Japan.

[95] R. Tubiana. *The Hand*, volume 1. Philadelphia, PA: Sanders, 1981, 1981.

[96] Claudio Uras and Alessandro Verri. Hand gesture recognition form edge maps. In *International Workshop on Automatic Face and Gesture Recognition*, pages 116–121, June 1995. Zurich, Switzerland.

[97] Rgis Vaillant and David Darmon. Vision based hand pose estimation. In *International Workshop on Automatic Face and Gesture Recognition*, pages 356–361, june 1995. Zurich, Switzerland.

[98] Inc VPL Research. *DataGlove model2 users manual*. VPL Research, Inc, 1987. Redwood City, CA, 1987.

[99] Alex Waibel and Minh Tue Vo. A multimodal human-computer interface: Combination of gesture and speech recognition. In *Proceedings of Inter Conference on Human Factors in Computing Systems CHI'93*, april 1993. Amsterdam.

[100] C. Wang and D.J. Cannon. A virtual end-effector pointing system in point-and-direct robotics for inspection of surface flaws using a neural network based skeleton transform. In *IEEE International Conference on Robotics and Automation*, volume 3, pages 784–789, May 1993.

[101] P. Wellner. The digitaldesk calculator: Tangible manipulation on a desk top display. In *Proc. ACM SIGGRAPH Symposium on User Interface Software and Technology*, pages 107–115., 1991. http://citeseer.nj.nec.com/wellner93interacting.html.

[102] Pierre Wellner. Interacting with paper on the DigitalDesk. *Communications of the ACM*, 36(7):86–97, 1993.

[103] Andrew D. Wilson and Aaron F. Bobick. Configuration states for the representation and recognition of gesture. In *International Workshop on Automatic Face and Gesture Recognition*, pages 129–134, June 1995. Zurich, Switzerland.

[104] D.A. Winter. *Biomechanics and Motor*. John Wiley and Sons, 1979. Control of Human Mouvement.

[105] Christopher Wren, Ali Azarbayejani, Trevor Darrell, and Alex Pentland. Pfinder:real-time tracking of the human body. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, volume 19 of *7*, pages 780–785, July 1997.

[106] Christopher Wren and Alex Pentland. Dynamic models of human motion. *3rd IEEE International Conference on Automatic Face and Gesture Recognition*, 1(3):22–27, October 1998. April 14-16, 1998, Nara, Japan.