Survey and Investigation of Hand Motion Processing Technologies for Compliance With Shape Conceptualization

Article · January 2004

DOI: 10.1115/DETC2004-57710

Edit Varga
Delft University of Technology

Imre Horvath
Delft University of Technology

Zoltán Rusák
Delft University of Technology

Bram de Smit
Delft University of Technology

Proceedings of DETC'04 ASME 2004
Design Engineering Technical Conferences
and Computers and Information in Engineering Conference
September 28 – October 2, 2004, Salt Lake City, Utah

DETC2004/CIE-57710

SURVEY AND INVESTIGATION OF HAND MOTION PROCESSING TECHNOLOGIES FOR COMPLIANCE WITH SHAPE CONCEPTUALIZATION

Edit Varga
Imre Horváth
Zoltán Rusák
Bram de Smit
Han Broek
Faculty of Industrial Design Engineering
Delft University of Technology
Landbergstraat 15, 2628 CE Delft, The Netherlands
{e.varga, i.horvath, z.rusak, a.desmit, j.j.broek }@io.tudelft.nl

ABSTRACT

An evergreen topic of human-computer interaction research is multi-modality. It has been considered important for the user interface of future computer aided conceptual design systems and what is hoped is that integration of, for instance, voice control, hand gesture/motion processing, and physical object scanning can increase both the semantic level and the efficiency of the interaction. In the area of computer mediated shape conceptualization, especially human hand motion detection and processing can play an important role. The authors' research focuses on the study of the opportunities offered by hand motion processing in shape conceptualization. As a first step they have studied the state of the art and analyzed the technologies applicable to hand motion processing. This paper reports on the findings. The various technologies have been sorted in four categories: direct incomplete, direct complete, indirect incomplete and indirect complete detection. First, the principles supporting this categorization are explained in Section 2. The next four sections of the paper investigate the hand motion detection and processing technologies. Section 7 discusses the characteristics and operational parameters from the aspect of using hand motion for shape input in conceptual design. Our conclusion is that the currently known technologies do not absolutely support processing of a language of hand motions that is under development for creative shape conceptualization. Therefore, the hand motion language needs to be redesigned and adapted to the best technology.

1. INTRODUCTION

Human-computer interaction (HCI) is an evergreen, but still somewhat neuralgic issue of computer-aided design (Dix, A. et al., 2001). For a rather long time, one-dimensional input devices (e.g., alphanumeric keyboards) and two-dimensional positioning devices (e.g., mouse) determined the way of entering shape information to design support systems. It has been recognized that these conventional input devices pose many, actually too many, constraints when applied in computer aided conceptual design systems, in particular, in creative shape conceptualization systems. Through the past two decades, HCI has pursued a broad and ambitious scientific agenda, progressively integrating its research concerns with the contexts of system development and use (Carroll, J. M., 1993). This has created an unprecedented opportunity to manage the emergence of new technologies so as to support socially responsive objectives. Multi-modality has been considered important for the user interface of future computer aided conceptual design systems (Oviatt, S. and Cohen, P., 2000), (Flippo, F. et al., 2003). A class of collaborative design systems seeks to provide multi-modal input possibilities for the collaborating design participants to enable them to hold virtual shape conceptualization sessions. What is hoped is that integration of, for instance, voice control (Dorozhkin, D. V. and Vance, J. M., 2002), hand gesture and motion processing (Rieger, T., 2003) (Aggarwal, J. K. and Cai, Q., 1999), and physical object scanning and reverse engineering (Várady, T. et al., 1997) can increase both the semantic level and the efficiency of the interaction and shape modeling.

In the area of computer mediated shape conceptualization, which features vagueness and incompleteness, human hand motion based shape externalization and modeling can play an important role. Hand motion is one of the most natural ways for designers to express their shape concept for computers and to communicate with each other during shape conceptualization. Therefore, it seems to be obvious to apply it as a new descriptive input device. Other constituents of highly interactive multi-modal user interfaces have also been developed and studied (Kuczogi, Gy. et al., 2001) (Grasso, M. A. et al., 1998) (Chu, C. P. et al., 1997). In addition to their natural character, the advantage of these means is that they do not require the user to decompose 3D modeling problems to 2D ones.

The ultimate goal of our research is to make it possible to use hand motions in three-dimensional conceptualization of shapes and to express shape related information in collaborative virtual design environments so intuitively and effectively as it is with two-dimensional sketching (Horváth, I. et al., 2003). A Hand Motion Language (HML) has been developed, which can be applied in shape conceptualization. Contrary to the efforts, we are still far from an ultimate solution. The main hurdles are the complexity of the problem and the lack of dedicated hand motion processing technologies. It has been recognized that a hand motion language cannot be optimized without taking into account a given detection and processing technology. Therefore the goal of our research has been twofold. Concurrently, we want (i) to find the best fitting technology and adapt it to our particular application, and (ii) to redesign and optimize the first implementation of our hand motion language with a view to the preferred technology. As a first step, we have studied the state of the art and analyzed the technologies applicable to hand motion processing. This paper reports on the latest findings.

In order to clarify the concept of the HML, we briefly describe its elements and structure. As a formal modeling language, HML is based on the symbolization of hand postures and movements. It comprises three types of constituents: (a) the set of correct letters, (b) the rules of concatenating the letters into words, and (c) the grammar for composing sentences from the words. The letters of the HML are signs produced as hand postures. A purposely sequence of changing postures of the hand(s) is a word of the HML. A sentence is composed of words that are needed to generate the components of a shape, or to manipulate them. According to its semantics, a word can be either procedural, or constructive. The procedural words such as 'stop', 'resume', and 'share' provide information for the process of shape conceptualization in the form of well recognizable starting and end postures that enclose the constructive hand motions. The constructive words provide information for the shape model. They can be grouped as geometric, identification, connectivity, positioning, scaling, and assembling words. Both two-handed and dual-handed words have been specified in order to increase the comfort of the designer and the efficiency of use.

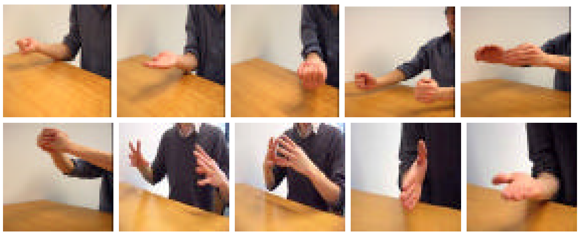

Figure 1. Hand motion language in use

As an example, Figure 1 shows a sentence, which generates a straight cone standing upright and tilts it over. This sentence consists of four words: the first one describes the base surface of the cone, the second one generates the conical surface, the third one merges these surfaces to form the volume of the cone, and the fourth one tilts the cone to rest on the conic surface.

Most survey papers published earlier addressed the problem of hand motion processing (HMP) from a pure technological point of view (Watson, R., 1993) (LaViola, J. J., 1999). Others concentrated only on camera-based recognition (Pavlovic, V. et al., 1997) or on contact methods (Hand, C., 1997). In this paper we investigate HMP from the aspect of real-time information extraction and conversion, which is required by creative shape conceptualization. Our aim was to analyze the available technologies and to compare them based on the operational parameters and the features of information processing. Section 2 introduces a simple reasoning model based on which the technologies have been sorted in four categories and the investigation of the conformance has been systematized. The following Sections identify and analyze the most important technologies and processing approaches. Section 7 compares the various HMP techniques taking into consideration the requirements originating in shape conceptualization.

2. INTRODUCING THE REASONING MODEL

In order to be able to reconstruct surfaces swept by hand motion in the 3D space we have to extract sufficient amount of information from the posture and motion of the human hand. Consequently, extraction of information has been chosen as the primary aspect of HMP. It can be carried out in several ways. Human hands can be completely scanned, or some characteristic points (such as landmark points or silhouette points) can be detected. These two ways of obtaining shape information from the moving hands can be identified as *complete* and *incomplete* information extraction. The way of information extraction represents the first dimension in our reasoning about technologies.

A second aspect of HMP is the way of transferring information from the physical space, in which the hands are moving, to the virtual space, where the shape is modeled. It is also a dimension in our reasoning. The information transfer can be *direct* or *indirect*. Direct transfer means that the positional and motional information obtained from the hands is directly sent to a geometric modeling system, which generates the visual representation of the swept surfaces, for instance, in the form of point clouds. We talk about indirect transfer when the obtained information is first fed into an intermediate hand model with the aim to generate extra information, and an extended set of information is transferred to the geometric modeler and processed as before.

According to our reasoning, the third aspect of categorization is the relationship of the hands and the information extracting devices. Certain devices are mounted on or touch the hand, while other devices can extract information at a distance. These relationships have been described as *contact* or *non-contact*. This aspect has special importance in our research since our intention is to avoid the negative influences of the contact technologies on the comfort and creativity of the designers who interact with the hand motion based shape conceptualization system. Considering the hand-device relationship as a third dimension and combining it with the other two dimensions, we created a frame of reference, which served as a reasoning model for the literature study (Figure 2). This reasoning model combines the aspects of extracting hand motion information and of transferring information to shape representation with the aspect of placing the sensors relative to the hand.

Based on the above introduced framework of reasoning, we can sort the various HPM technologies into four processing categories, which have been called (i) direct incomplete, (ii)

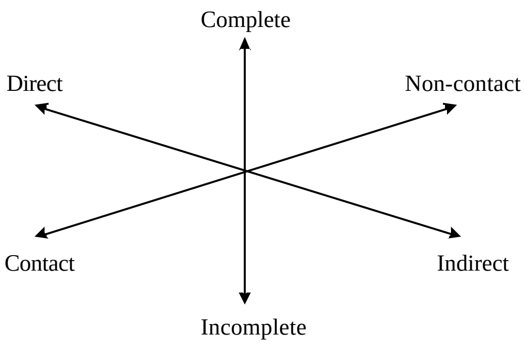

Figure 2. Categorization of HMP technologies

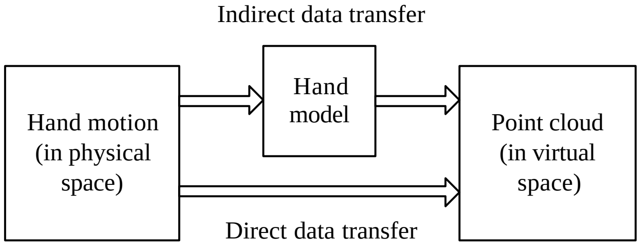

Figure 3. Information transfer during hand motion processing

direct complete, (iii) indirect incomplete and (iv) indirect complete. These categories of processing are graphically illustrated in Figure 3. Again we emphasize that the major difference between the categories is whether an intermediate hand model is generated or not. We are talking about indirect data transfer, when an active hand model is used to extend the detected data for a better representation of the swept surfaces or for a better mapping of a manipulative action. A large quantity of contact and non-contact technologies belong to each of these categories. In the next Sections we focus only on those technologies, which can in principle be applied in a hand motion based shape conceptualization system. In addition to the functionality and the operational parameters, we are also interested in the experiences related to the reported applications. More specifically, we are going to investigate (i) which categories give the best opportunities for HMP, and (ii) which technologies are the most advantageous for HMP within each category. In Section 7 we will investigate the compliance of the various technologies with hand motion based shape conceptualization.

3. DIRECT INCOMPLETE HAND MOTION PROCESSING

This form of HMP technologies can extract or are used to extract only a limited amount of information from the moving hand, which is not sufficient to reconstruct a swept surface. The information from the moving hand is directly transferred to the geometric modeling system where the intended shape in virtually reconstructed. It is usually a curve. It is supported by both the contact technologies and non-contact technologies. The contact technologies can be divided in two subcategories on the basis of they are held in the hand, or they are mounted on the hand.

3.1. Contact technologies

A typical representative of the first subcategory is a mouse, which supports 2D positioning tasks. When included in a graphical user interface it facilitates interactive drawing in 2D. For this functionality other 2D input devices have also been developed, for instance, tablets. However, humans design 3D objects rather than 2D representations; using 2D input devices only projections of 3D objects can be designed. The interpretation of 2D sketches and drawing needs expertise, and reconstruction of these incomplete representations in the form of 3D computer models is only partially solved. This is the reason why a lot of efforts have been invested in technologies that are capable for 3D drawing. They try to develop a digital 3D equivalent of drawing on a piece of paper. First we analyze some of those proposal and implementations, which converts the 2D drawing functionality of a mouse to a 3D drawing functionally. With these technologies the user can navigate in a 3D virtual or physical modeling space and can generate or select spatial points.

An early effort was made by (Clark, J. H., 1976) to design free-form surfaces in 3D. He intended to position the control points of a B-spline surface in the space using a 3D wand, and to generate the surface accordingly. With the evolution of this type of technologies, the focus of research shifted from the functional aspects towards other aspects such as cognizance, ergonomic and efficacy. The ergonomics considerations related to 3D input devices put the technology development in a different context. The consideration of intuitiveness and creativity brought the issue of application of natural two-handed gestures (Leganchuk et al., 1998) in the focus of research. The experimental JDCAD (Liang, J. and Green, M., 1993) and THRED (Shaw, C. and Green, M., 1994) systems represent the first practical technologies. They apply two 3D position and orientation trackers with three buttons, one for each hand. These devices, called 'Bat', supports spatial geometric modeling tasks, such as modeling mechanical components or designing surfaces.

Based on anthropometrics and ergonomics considerations, left and right hand operations are defined and executed according to the importance of the tasks. That is, the most frequently used operations can be done by the dominant hand. The 'Frog', proposed by (Gribnau, M. W. and Henessey, J. M., 1998), is a mouse-like input device with two buttons. A six DOF magnetic tracker was built inside to measure the Frog's location and orientation. The 'Bug' and the 'Dragonfly' are similar input devices using optical tracking (Stefani, O. and Rauschenbach, J., 2003). The Bug has two retro-reflective spheres for the identification and the position detection of the device. The Dragonfly has six retro-reflective spheres to facilitate the detection of its position and orientation. The user can navigate in the virtual environment with two hands, one holds the Bug, and the other holds the Dragonfly. The latter has a virtual ray coming out of it, and the former has buttons. With this combination the problem of menu selection in 3D is solved. The user can point with the ray to a menu item and then press the button to select it. A virtual object can be connected to the Dragonfly to follow its motion while the button on the Bug is pressed.

With the help of these devices users can directly input 3D data into computer systems in real-time, and the modeling time can also be reduced. However, these devices were developed following the navigation principles of traditional computer aided design systems. Conventionally, the user controls the application with menus, and change the shape of objects by clicking on menu items or icons, and the result of the operation is usually displayed on a monitor. Free-form objects require the designers to define control points and then the system interpolates or approximates them to form a curve or a surface.

This approach however involves difficulties in the case of a 3D design tasks, requiring split attention. The control points should be selected in 3D and the menu items and icons in 2D. The former is difficult since the users do not have appropriate spatial visual feedback on the 2D monitor. The latter is also difficult because of the inherent 3D characteristic of the input device. Therefore, researchers had to come to the conclusion that these approaches are not the best in terms of supporting intuitive shape design. To solve the visual feedback problem, they started to apply head-mounted displays (HMDs), which place the user into a virtual environment (Clark, J. H., 1976). It has been observed however that psychologically and ergonomically this is not the best solution since the designers have no contact with the real world while wearing the HMDs (Wesche, G. and Seidel, H. P., 2001) (Keefe, D. F. et al., 2001). Research has also been done towards better solutions for control, and 3D navigation elements such as the 3D cursor and menu (Liang, J. and Green, M., 1993) and the 3D color picker (Keefe, D. F. et al., 2001) have been proposed.

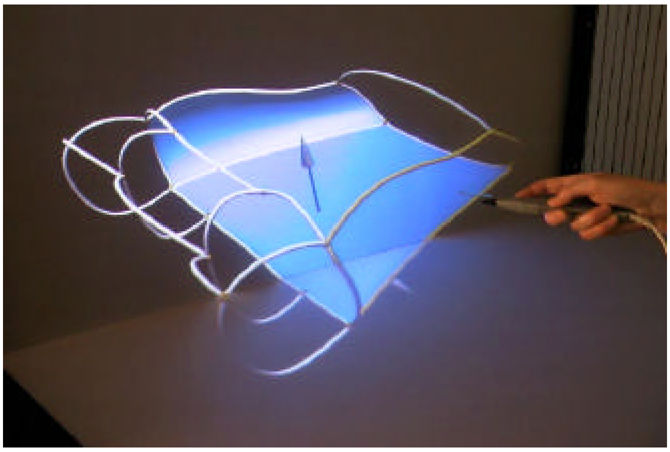

Figure 4. The FreeDrawer system in use (courtesy of Wesche, G. and Seidel, H. P.)

To come through the difficulty of continuous clicking when the user wants to draw a free-form curve, a new device was developed by (Wesche, G. and Seidel, H. P., 2001) for their FreeDrawer system (Figure 4). It comprises a stylus by which the user can input 3D points to the computer system by simply moving it in the 3D space. The orientation of the stylus can be also measured. This method is similar to drawing with pencil on a paper. Developers of the FreeDrawer system also reported that the usage of the system caused some troubles to users without artistic skills. It is a sketching system with which the user draws space curves in a virtual environment with a tracked stylus. In the FreeDrawer system a curve network is created in order to generate surfaces by filling in the connected curves. The developers also observed difficulties with changing among several tasks, and proposed new interaction styles. To solve this problem they proposed a hand held 3D widget with a set of virtual pointers, originating at the stylus. Each pointer corresponds to an editing tool and the user can directly touch the object with the tip of the pointer. By pressing the stylus button, the function is performed on the selected object.

This solution is advantageous in artistic applications, likewise the CavePainting system, in which the user moves a tracked brush in the 3D space and strokes are created according to its motion (Keefe, D. F. et al., 2001). In the CavePainting system the users can define so called strokes by moving the brush by the hand, and it is similar to painting on a huge canvas. It is possible to change the type of strokes by dipping the physical brush into a cup. Conductive cloth was placed on the tip of the brush and along the inside of the cup in order to sense the contact. Users have the possibility to interact with the virtual environment by both hands. While the dominant hand holds the brush, the user can wear a pinch glove on the non-dominant hand. By pinching the thumb and the index finger, some features such as changing the color or the brush size, can be accessed quickly, even during drawing. These features make the CavePainting very adequate for artistic tasks, but it has nothing to do with designing complex mechanical shapes.

The two technologies investigated above belong to the subcategory of technologies based on hand held devices. With this type of interaction many drawing tasks could be successfully realized in the 3D space. The mouse-like devices and the stylus measure either the position, or the position and orientation of the global hand together, but do not care about the local finger motion. There must be a permanent grip and support provided, and, in addition, the shape of the hand can be hardly changed when holding these 3D drawing devices. By tracking one point in the 3D space a free-form 3D curve can be created, but generation of a free form surface is challenging. This is true for those approaches where drawing in 3D is realized based on detecting the motion of a fingertip. These technologies can support only simple navigation and manipulation tasks. Consequently, they cannot be considered in processing of hand motions in shape conceptualization.

Aligned with many more researchers we claim that a better solution can be achieved with non-contact technologies. Typical implementations of these technologies are based on computer image processing or computer vision. In these implementations image sequences are taken from camera(s) and analyzed by image processing algorithms. We investigate these technologies in the next Subchapter.

3.2. Non-contact technologies

Several competitive approaches have been proposed, but many of them suffer from the algorithmic complexity of image processing and the high computational time (Abe et al., 2000). Certain systems restrict the motion of the hand to avoid difficulties and to reduce the problems of occlusion. Many systems are sensitive to variable lighting conditions and backgrounds, and moving in front of changing backgrounds. The pre-calibration of the camera(s) also takes time and presume some knowledge about the position of the hand. Since with one camera the calculation of 3D positions is not possible, researchers apply multiple cameras, typically two, (Segen and Kumar, 1998), or use other additional information, such as shadows (Segen and Kumar, 1999).

We start our investigation with a system, in which only 2D data are considered and a single camera is used. This system, developed by Iannizzotto, G., et al. (2001), substituted the mouse with a vision-based input device. As the user communicate with the computer through the monitor, 2D information is sufficient. The thumb and the index finger are used to emulate the functions of an ordinary mouse. The necessary information is obtained from the motions and contacts of the fingertips. After filtering the image, the portion where the fingers are detected becomes small (app. 60 x 100 pixel). Consequently, the pattern-matching algorithm can work in real-time. They applied a 6 frame per second (fps) tracking speed, considering that their camera was able to capture 25 fps and they approximated the speed of dynamic gestures as 8 Hz. User tests showed that the use of this system could be learnt easily, although the accurate selection of a menu item proved to be difficult. In a similar approach, in the Finger Mouse System, which was developed by Ko, B. K. and Yang, H. S., (1997), the position of a fingertip was tracked and the different functions were detected by three pre-defined gestures. Their Finger Draw application is driven by gestures, but these gestures are detected as sequences of points described by the moving fingertip. Based of these point sequences, primitive objects can be drawn and simple operational task can be realized.

In order to make 3D free-form curve design without hand-attached devices possible, Abe, K. et al. (2000) developed a 3D drawing system. Actually what this system does is using hand motions to complete simple drawing tasks. They used two cameras to detect the moving hand, from which one took top-view and the other took side-view. The restrictions of this system are that the designer must use his right hand, and the back of the designer's hand must always face up. The designer indicates 3D points by positioning the tip of its finger and gives a drawing command by hand pose. The system extracts a pair of fingertips on both images and estimate the 3D position of points by comparing the coordinates and a pair of camera parameters. The average time of hand pose recognition is 10 fps. They reported that they could draw images without feeling delay because of the short processing time.

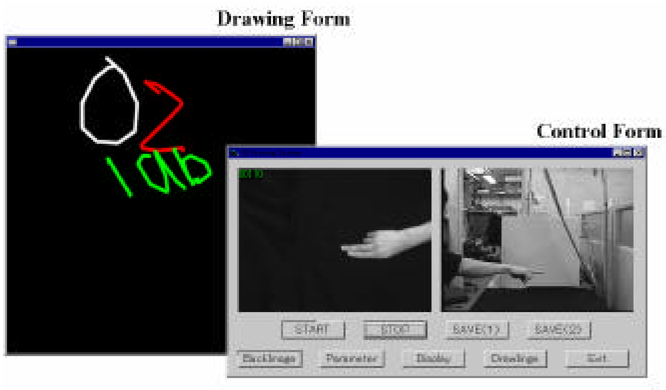

Figure 5 A 3-D drawing system using two cameras for hand motion recognition (courtesy of Abe, K., Saito, H. and Ozawa, S.)

Segen and Kumar (1998) first applied two cameras to estimate the 3D position of the fingertip of the thumb and index finger. Then, they implemented a pilot system based on one camera and a light source, most probably to reduce the processing time (Segen and Kumar, 1999). In the latter case, features derived from the projections of the hand and its shadow was used to compute 3D position and orientation. Both systems operate at the rate of 60 Hz. They tested it with a Fly-through application, where the position of the pointing finger along an axis was used to control the velocity. With the information of two fingertips, picking and moving tasks can be realized. In their Scene Composer application, objects can be picked and moved in a 3D scene.

As it has been shown in this Section, many efforts have been made to use hand motion as an input means. Striving after natural ways of describing shapes, researchers studied the opportunities of obtaining sufficient amount of information directly from the hand, without any hand-held device. Our observation is that in the investigated cases only some dedicated points were detected. That is, these technologies extract less information than that is needed to generate surfaces or complete boundary of objects. They transfer the obtained data either directly to the modeling subsystem, or to the image processing subsystem and then to the modeling subsystem. They do not apply active hand model to generate extra information. In the next Section we investigate those approaches, which try to collect as much information from the hands as possible.

4. DIRECT COMPLETE HAND MOTION PROCESSING

If sufficient amount of information is collected, the swept surfaces and curves can be regenerated with high fidelity. If this extensive set of information is directly transferred to the modeling system, fast response times and real time interaction can be achieved. This is the philosophy behind the direct complete hand motion processing technologies. It cannot be implemented by image processing based technologies, unless additional information provided. At the same time, the technologies for scanning fast moving objects are still in their infancy. This is why we could not find relevant non-contact technologies in this category. Nevertheless there are contact technologies, which can detect large amount of information on the moving hand.

Murakami and Nakajima (2000) developed a hand-held input device for intuitive 3D shape deformation. It has a cubic shape and pieces of conductive foam are embedded on the surface and the inside of it to form a sensor network to cover the cube. It contains 24 sensors on the cube edges, 48 on the faces and 18 inside the cube to measure deformation. A 3D virtual control volume is defined on the screen for deforming the objective shape. By measuring the deformation of the tool electrically in real-time, the system computes the corresponding deformation of the control volume, and then that deformation is mapped to the objective shape.

The conclusion we can derive based on the investigation in this Section is that the currently available contact technologies can be used to obtain large amount of data that is needed to generate complex hand shapes and/or surfaces. However, if an incomplete detection is complemented with the use of active hand models, they can also be considered. Below we investigate the technologies, which have been sorted into this category.

5. INDIRECT INCOMPLETE HAND MOTION PROCESSING

These HMP processing technologies can extract only a limited amount of information from the moving hand due to the inherent technological limitations or to striving for short cycle time. A hand model is used to generate the missing information that is not extracted from the moving hand and to support the construction of the intended shape by the geometric modeling system. We consider these hand models as active models, since they are activated and refreshed whenever new information from the moving hand is obtained. We consider a hand model to be good, if it can be parameterized for each user, because our system requires relatively accurate positional information, not only rough estimation.

5.1. Contact technologies

Data gloves offer the most possibility to capture information about human hand motions, as they were designed especially for this task. The most equipped model contains 22 sensors. It has three flexion sensors per finger, four abduction sensors, a palm-arch sensor, and sensors to measure flexion and abduction. The sensor resolution is 0.5 degrees and the raw sensor data rate is 150 records/s and 112 records/s when it is filtered. With this information hand motion can be captured and a hand model can be driven in real-time. However, position and orientation data can be obtained only with additional tracking devices.

The amount of degrees of freedom tracked can vary. Pure point tracking results in positional (X, Y and Z) information. A number of systems can track position and orientation. (X, Y, Z, A, B and C). In order for the system to be able to follow a point it must be recognizable. To achieve it there are three methods. First, it can be recognized from a geometrical feature. In this case the surroundings of the point must be scanned and interpreted. Second, a passive marker can be attached. This is a piece of material that can be recognized by the system because, for example, it is magnetic. The third option is to attach an active marker. This is a small device that sends out some signal that can be received by the tracking system, for example light pulses. The number of tracked points varies from 1 to 256 in the case of commercially available systems. Different types of media, like light or magnetic fields, can be used for measuring.

The complexity of the tracking task increases if more points have to be followed simultaneously. There must be for example a way to distinguish the different points so that they cannot be mixed up. Depending on the speed of the motions to be tracked one will need a faster or slower system. The speed of the tracking system is determined by the amount of measurements that can be made per second, which is about 100-240 Hz. The accuracy of detection varies from 0.1 mm to 2 mm. Most measuring systems can be influenced by external factors, like light sources or magnetic fields. It depends on the application whether this is a problem or not. A tracking system can influence its surrounding. This can cause unsafe situations. For example high-energy lasers can damage the eye; magnetic field may disturb for example pacemakers. By building up this setup, positional, orientation and also hand shape data can be obtained. Now we analyze systems that use data gloves to obtain hand information for communicating shapes.

A typical technology was developed by Xu, S. et al. (2000), which is able to support the creation and manipulation of free form curves by hand motion. They detect the tip of the user's index finger with the data glove and reconstruct its motion curve from these points. They developed algorithms for filtering the points, for instance, to delete unintentional point and to compress redundant points, in order to create a high fidelity curve from the hand motion.

As an early effort, Weimer and Ganapathy (1989) developed a 3D modeling system, which uses hand gestures to interact with a virtual environment. A data glove detects hand motion and it is used to control a computer model of the human hand. The hand model can be parameterized as either a left or a right hand, and it provides continuous visual feedback, showing the hand's relationship to the virtual world. The selection, viewpoint modification and surface modeling functions are implemented by combining hand motion with voice input. Although the usage of the hand model offers the possibility to define swept surfaces by hand motion data, in this system surfaces are built from swept curves. To take more advantage from the hand model, Matsumiya et al. (2000) presented an interactive technique for modeling 3D freeform surfaces in real-time. In this system users can design 3D objects by using their fingers, which are detected by a data glove. A primitive object is given, which can be manipulated by two operations, denting and pinching. Predefined finger angles switch the types of manipulations. The forefinger, the shape of which is given by a mathematical formula, produces dent deformation. Actually, its shape is subtracted from the primitive object. The pinch deformation adds a cylindrical shape to the primitive object, which is generated among the forefinger, middle finger and thumb.

In the above technology, the shape of only one or two fingers was used. However, it is natural to use the whole hand to deform shapes. The next effort provides a solution for this problem. As a development of the surface deformation technique (Ma et al., 1997), the authors proposed a new method (Wong, J. P. Y. et al., 1998). They addressed problems with the planar projection method. Before manipulation, users were required to set their hand in a coplanar state to create a base plane of the hand. Deviations of finger joints from the base plane lead to errors in the initial mapping. On the other hand, the amount of hand movement could be small, because the mapping was done via the base plane. Because of these reasons, they changed the mapping method from planar projection to ray projection. To achieve this, they constructed a 3D triangulated hand model; therefore the users do not have to set their hand in a coplanar state. The hand model was built upon the points measured by the glove and some additional points, which were defined by simple distance assumptions. Then each point was connected to the neighboring points and a surface was created from triangles and the projection was realized through them. The whole deformation procedure took less than 0.5s. The system was further developed to a distributed virtual sculpting environment (Li, F. W. B., 2003) to support shape communication among geographically separated designers. The main disadvantage of this method is that users got tired during using the glove. They also commented that as they are editing a model with both hands, they can not look around in the virtual environment.

Nishino et al. (1998) proposed a method for modeling 3D objects based on hand motion detection by a data glove. Spatial and pictographic bimanual gestures were used to create and modify 3D objects. An object was defined by a process, in which the user created and combined primitives, and deformed this rough shape to achieve the wanted shape. These steps were iterated until the final shape is obtained. In order to get an adaptable gesture interface, they implemented a gesture learning and recognition algorithm, which allows the users to register their preferred gestures before using the system. The main findings of their experiments were that dynamic adjustment of the quality of visualization and speed of drawing are critical issues for an efficient modeling. The users of this system spent about 20 minutes to produce complex objects, e.g. a teapot.

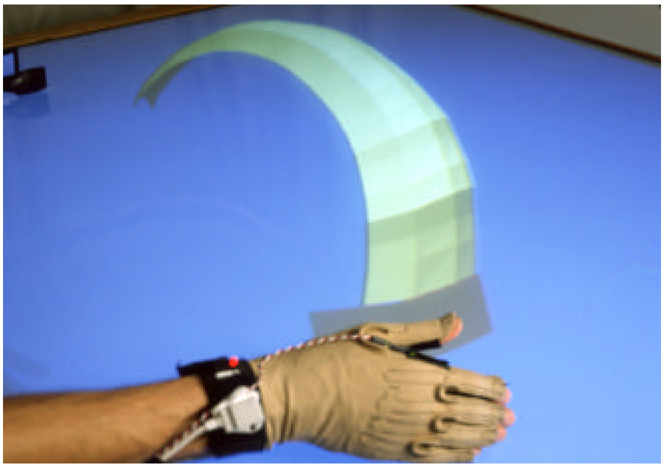

Figure 6 Surface drawing (courtesy of Schkolne, S., Pruett, M. and Schröder, P.)

While in the previous approach users created and deformed primitive objects, Surface Drawing was developed for creating free-form 3D shapes by hand motion (Schkolne, S., 2001) (Figure 6). As the hand is moving in the 3D space, the path of the hand is directly realized as geometry. Their goal was not to create numerically precise models, but to support shape expression and communication. Although surfaces are defined by hand motions detected by a data glove, modifications of the shape should be solved using tangible tools. To give a solution for this problem, Ma et al. (1997) developed a technique, which allows the users to deform surfaces by hand motion. The angles of finger joints and the local coordinate of the palm were measured by a data glove. To calculate the joint positions, they assumed that the distances between the finger joints are given. These points were further used to interpolate a surface, called hand surface. In their method surface deformation was realized as a planar mapping of the hand surface to the surface of the object.

The usage of the hand model also proved to be efficient in the next effort, which is a different application. However, in this case the glove is not equipped with electric wires and sensors, it is in contact with the hand. In the Hand Motion Understanding system, developed by Holden and Owens (2001), a 21 DOF kinematic 3D hand model is used to detect hand motion. The model is a combination of 5-finger mechanism (total 15 DOF) each attached to the wrist base of 6 DOF. They used a color-coded glove in order to locate the joint positions on the images captured by a single camera. The joints are separated by different colors; thus, a color segmentation algorithm can be used to determine markers. Then a correspondence is established between the marker locations and the joints in the model. Computing the size of the markers on the image and predicting its location can solve the occlusion problems. The five previous frames are used to predict the 3D model state and the possible states are projected onto the image. The model is gradually updated until its projection fits the image feature. In order to classify the hand signs, a vector is built from the angles of finger joints, that is, according to the 15 DOFs. A sign is represented by its starting/ending posture and motion information. The starting and ending postures are defined by using the Auslan basic hand postures. The number of directional changes in the movement of fingers is used to represents the motions.

From the point of view of our intended application, the technologies that extracts information from the bare hand, and use neither active nor passive markers, are even more interesting.

5.2. Non-contact technologies

Basically, recognition of hand postures and gestures can be categorized into two approaches considering camera-based systems: model-based approaches and appearance-based approaches. The latter takes advantage of some features extracted from the image of the hand and classifies it into gestures (Utsumi and Ohya, 1999), (Wu and Huang, 2000), (Lamar et al., 2000). These methods are usually based on the assumptions that the appearances are much different among different gestures, but small among different people, so the estimation of the hand configuration does not need to be accurate. It can be true in some applications, such as recognition of a set of predefined gestures, but it is not sufficient in applications, which require multi-DOF input, like in our case. Because of the previous reason, we do not give detailed description of the appearance-based techniques and we focus on the model-based approaches. As our aim is to obtain 3D positional data from the hand, we mainly consider approaches with 3D hand models. In this paper we do not analyze approaches with 2D hand models (Heap and Samaria, 1995) (Rehg and Kanade, 1994).

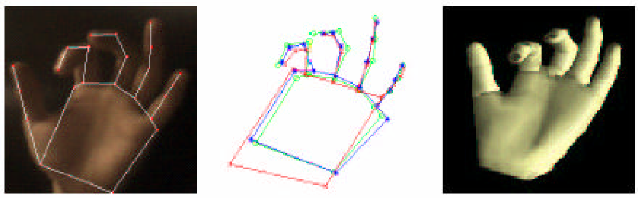

Figure 7 Modeling articulation of the human hand (courtesy of Wu, Y. and Huang, T. S.)

First we demonstrate some examples which use a single video camera to capture the hand. Although these methods cannot provide us with 3D positional information, we analyze them because of the hand models. Wu and Huang (1999) proposed a camera-based non-contact method for capturing articulated human hand motion, which was decoupled to global hand motion and local finger motion (Figure 7). They employed a kinematic hand model, where each finger was modeled as a kinematical chain with the palm as its reference frame. They treated the hand model as a set of 16 rigid objects, in this case sticks. The configuration of the hand was defined by the length of the sticks and the kinematical relation among them. A generic 3D hand model was automatically calibrated to each person in order to derive user-specific models. Feature points were observed from the images taken by the camera. Based on these points and the estimated motion parameters the 3D hand model could be regenerated. The algorithm worked accurately even if the local finger motion between two consecutive frames is large. However, the algorithm failed, when one of the fingertips was occluded.

Heap, T. and Hogg, D., (1996) applied a 3D deformable Point Distribution model of the human hand in order to track the hand moving with 6 DOF. The hand model was constructed from real life examples of hands and it is modeled as a surface mesh from which large amount of position information can be derived. The model is used to track a hand in real-time, 10 fps on a standard 134 MHz Silicon Graphics Indy workstation. The model is projected onto the input images and 3D edge detection is used to move and deform the model. Their system assumes a homogeneous dark background and the user must initialize the hand model. They addressed problems like scale and rotation confusion, planar rotation ambiguities, occlusions and implausible model shapes.

Although positional information is not available in these cases, the hand shape is successfully reconstructed. In the following part of this section some examples are shown using multiple cameras.

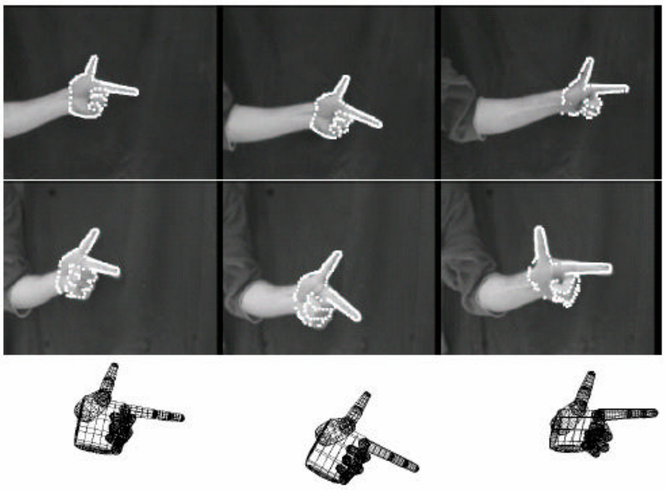

Figure 8 Model-based 3D hand tracking (courtesy of Stenger, B., Mendonca, P. R. S. and Cipolla, R.)

Stenger, B. et al., (2001) used an anatomically accurate 3D hand model for tracking the hand on 2D images (Figure 8). The hand model was created from truncated quadrics, which provides a practical method for generating contours of the model. The hand model was projected and compared with the edges of the hand detected on the images. Their system is scalable from single to multiple cameras and from rigid to articulated model. In their experiments they used two cameras and a 7 DOF hand model. This method can handle self-occlusion. They captured images with 360x288 pixel in front of a dark background. The parameters of the hand model were manually set to match the pose of the hand in the first frame. The global hand motion was tracked with 6 DOF and 1 DOF was given to the configuration of the thumb. They intended to use this model in simulating a point and click interface, using positional, velocity and acceleration data. They addressed that the computational complexity grows linearly with the number of cameras. They reported that the system can not operate in real-time, and they proposed some methods to improve the speed, such as optimizing the code, using a faster machine (in this experiment a Celeron 433MHz machine was used) or a distributed system.

Ogarawa et al. (2003) captured the hand by three infrared cameras. With this technology, silhouette images of the hand can be extracted easily. Then a volumetric representation of the hand is generated from the three silhouette images. In this approach, a kinematic bone hand model is used with 20 DOF of 15 joints, and extra 6 DOF for the entire hand, 3 for translation and 3 for rotation. A surface mesh model of the hand is constructed by measuring a real hand and fitted to the bone model. The position and shape of the reconstructed hand volume is estimated by fitting the articulated hand model in the 3D space. They reported that the accuracy of thumb pose recognition is less than other fingers due to the complex kinematic structure.

However, in these cases they used multiple cameras because of the accurate reconstruction of the hand configuration, with this technology positional data also can be obtained as we could see in Section 3.2.

6. INDIRECT COMPLETE HAND MOTION PROCESSING

This approach is based on technologies that are capable to extract sufficient amount of information from the moving hands; nevertheless, it also uses a virtual hand model in the process of regenerating the swept surfaces. Obviously, the parallel use of these two information sources results in a redundancy in terms of data. On the other hand, some technologies are capable to reconstruct the whole hand surface. We consider these hand surfaces as passive hand models, since they are generated during the detection process. As we do not know anything about the configuration of the hand, other process is needed to extract the required information. For example, if we want to extract the points from the palm side of the hand, it requires further analysis.

6.1. Contact technologies

CT, MRI and ultrasound technologies can capture internal images of 3D volumes and scanners can detect points on the surface of objects. With this information the surface of the 3D

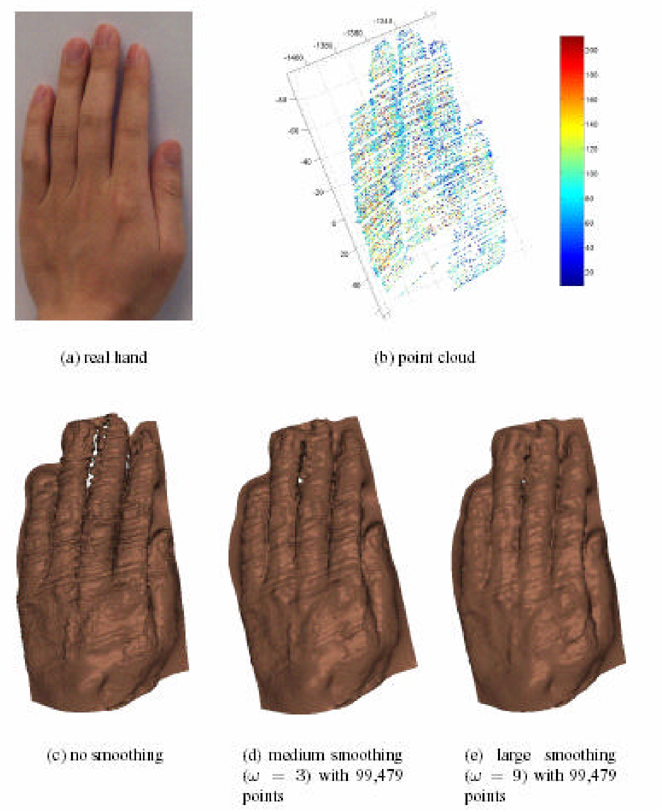

Figure 9 Direct surface extraction from 3D ultrasound images (courtesy of Zhang, Y., Rohling, R. and Pai, D. K.)

object can be reconstructed. However, if we want to measure the 3D position and orientation of the object, in our case the hand, additional equipment is needed. Typically, 3D sensors are attached to the object to be measured, that is the reason we categorized them into the contact technologies.

In medical examinations 3D ultrasound, CT and MRI are a widespread technique. Recently researchers deal with surface reconstruction based on images provided by these technologies (Estépar, R. S. J. et al., 2003), (Cheng, S. W. and Dey, T. K., 1999). Zhang et al. (2002) proposed a technique for extracting surfaces from 3D ultrasound data (Figure 9). They used an open architecture ultrasound machine for measurement and markers were attached to the probe to be able to track its position and motion optically. The accuracy of tracking is high enough (~0.1 mm), but the overall system accuracy (~1mm) depends on the calibration of the probe. During the calibration process the transformation matrix from 2D ultrasound images to the 3D probe is determined. The measured surface is implicitly described as a set of points with three coordinates. Then the surface is defined by a single function, whose parameters are these points. As the number of the obtained points was more than 100000, and it just highlighted the noise, point deletion algorithm was needed. Depending on the application, the amount of remained point can be chosen. With almost 10000 points they achieved 50s fitting time and 73s surfacing time, by using a 1GHz Intel P3 computer with 1GB RAM.

Scanners provide us with large amount of point data. As in case of the previous techniques, we get the whole shape of the hand, but we do not know any specific information about it, that is, we do not know which points belong to different hand parts. Most of the research being done on surface reconstruction upon scanned data (Várady, T. et al., 1997) (Vergeest, J. S. M. et al., 2001). However, in our case this large amount of point information is sufficient, but the problem is making correspondence between points and hand parts.

6.2. Non-contact technologies

Deawele, G et al. (2004) proposed a method for tracking hand motion from 3D point trajectories and a smooth surface model. They developed an articulated hand model, where the skin surface is defined as an implicit surface, and the skin motion is described by skinning techniques used in computer animation. The basic configuration of the hand is a 27 DOF kinematic chain. A stereoscopic sequence of images was taken with the restriction, that the palm should face the cameras. About 500 points of interest were extracted and matched between the left and right view, and the resulting 3D point set served as input for hand tracking. Then the hand structure was modeled interactively, and the third view was used for the validation of tracking results. In our case, this large amount of point information is sufficient, and we do not need an additional hand model.

7. DISCUSSION

In this section we further investigate the various technologies from the aspect of using in a hand motion based shape conceptualization system. To support the discussion, first we characterize the technologies for the information content that they provide for virtual reconstruction of swept surfaces according to the given requirements. Let IH be the amount of information obtained from the moving hand; IM the information that is derived from a virtual hand model, and IS the information needed to reconstruct the swept surface. Figure 10 shows the characterization of the discussed HMP approaches for the kind and amount information they process. All direct incomplete approaches provide less information than needed due to the partial scanning of the hand. Conversely, all indirect complete approaches result in more, actually redundant, information since they completely scan the hands and also manipulate hand models. These features do not promote the application of these two approaches in hand motion based shape conceptualization. Therefore, we excluded them in our research from the further investigations.

The indirect incomplete HMP approach seems to make sense in our specific application since the information obtained by partial scanning of the hand can be extended by the information obtained from a hand model, and thus the swept surfaces can be reconstructed with high fidelity. The direct complete approach is also appropriate since it is supposed to obtain and transfer sufficient amount of information to the virtual modeling space at once. However, the number of actions

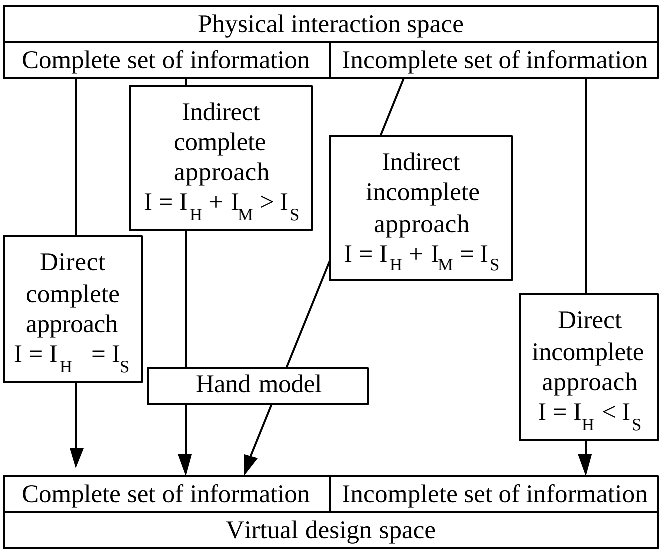

Figure 10 Comparison of HMP approaches

involved in the two processes is different. In the first case, (i) the hands should be detected and scanned, (ii) the information obtained by partial scanning should be transferred to the hand model, (iii) the hand model should be actualized, (iv) the required additional information should be derived, (v) the extended set of information should be sent to the shape modeling system, and (vi) the surface should be displayed. In the second approach, (i) the hands should be detected and fully scanned, (ii) the information should be transferred to the shape modeling system, (iii) the information should be preprocessed, and (iv) the surface should be displayed. A process involving less number of processing can be in principle better, but in our case we have to take into consideration the capabilities of the current technologies. For this reason it is not obvious which approach is finally better. In any case, we give preference to non-contact technologies for the comfort of the designers.

The possibility of real-time processing strongly depends on the time elapsed by the execution of the actions of processing. Therefore the speed of detection, scanning, and computation is also considered as a technology selection criterion. Real-time processing is crucial in a conceptual design system, where ideas may come rapidly after each other and designers need fast visual feedback. Based on the cognitive model of shape conceptualization, the typical cycle time is between 1 and 10 s. It means that the hardware and software platform of the system should be able to provide us with visual feedback in at least ten seconds. The speed of the hand motion (5-8 m/s) is also a challenge for the HMP technology. Our analysis showed that while contact technologies like data gloves could work in almost real-time, camera-based detection systems need more time due to image processing. On the other hand, direct complete HMP approaches elapse more time at scanning the hand than the indirect incomplete approaches, but the latter require additional time to process the virtual hand model.

When designers use their hands to conceptualize shapes in the 3D space, the free movement of the hands is a basic requirement. The designers should not be limited by the applied detection and scanning technologies. Intuitiveness of motion suffers a lot under restrictions such as 'user's hands must always face up'. Furthermore, if the hand movement is constrained by heavy and uncomfortable equipments, or by cables that connect the user to the computer, the space of motion is limited and the comfort is demolished. It requires significant adaptability from the HMP technology that can currently be achieved only with limitations. Specific technologies such as color-based gloves only slightly restrict the hand motion, though they are in a direct contact with the hand. Obviously, the non-contact technologies meet the comfort and adaptability requirements much more. However, data gloves can work properly in different hand positions and orientations, camera-based systems usually restrict the position and orientation of the hand due to the difficulty of handling occlusion problems. In the case of gloves, for instance, also the different hand sizes can cause problems. It is requested that the quality of HMP should not be influenced by the trajectory and speed of hand motion. Some of the low-scale image-based systems have intrinsic limitations to fulfill this requirement.

The constructed surface should properly reflect the details and characteristics of the intended surface - a fact that introduces the requirement of fidelity. The typical magnitude of the macro-geometry of the human hand is 10 – 100 mm, and of the micro-geometry is 0.1 – 1 mm. However, the hand moving in the space sweeps a vague domain rather than a crisp surface. The reasons of this are (i) the multiplicity of points on the hand that generate the surface, (ii) the uncertainty of the best fitting motion trajectories, (iii) shaking and imperfect forming of the hands while moving, and (iv) the interaction of macro- and micro-geometry information at scanning the hand. In general, the typical magnitude of the characteristic uncertain movements is 1 – 10 mm. The HMP technology and the geometric modeler should jointly take care of these. Some progress has been achieved with fuzzy sensing technologies, but much more can be expected from non-nominal shape modeling, such as vague discrete interval modeling (VDIM) (Rusák 2003) and alpha-shape modeling (ASM) (Gerritsen 2001). Unfortunately, in many of the current research projects, hand motion based shape input and shape concept modeling have not yet been addressed concurrently.

8. CONCLUSION

Hand motion is regarded as a prospective input mechanism for computer aided conceptual design systems for initial shape design of consumer durables. Its success or failure in this application largely depends on the enabling detection and processing technologies. A lot of research and development effort has already been invested in this area, therefore a survey and investigation of the current state of the art is not only timely, but also necessary. Our research has a dual focus: we simultaneously investigate the compliance of HMP technologies and redesign and optimize the first implementation of our HML with a view to the preferred technology. This paper reported on the findings about the state of the art and analyzed the technologies applicable to hand motion processing. A novel classification scheme has been introduced to systematize the survey and investigations.

Direct complete and indirect incomplete HMP technologies have the potential to support hand motion based shape conceptualization. In this paper we compared these two categories of technologies from four aspects: contact, speed, adaptability, and fidelity. The advantage of the former technologies is that they do not need extra time to process a virtual hand model, but they do need it to scan the moving hands. If sufficient amount of information is obtained from the hand, fidelity of the generated surface can be high. Indirect incomplete HMP technologies target landmarks or other characteristic points only. They require less scanning time as well as less sensitive and powerful technology, which is an important issue from a technological point of view. Our conclusion is that based on the literature study the HMP technologies can only qualitatively be assessed for applicability in hand motion based shape conceptualization. Consequently, more experimental research is needed in order to decide on a concrete technology, which meets the requirements best.

REFERENCES

Abe, K., Saito, H., Ozawa, S., (2000), "3-D Drawing System via Hand Motion Recognition from Two Cameras", in *Proceedings of the 6th Korea-Japan Joint Workshop on Computer Vision*, 21-22 January, Nagoya, Japan, pp. 138-143.

Aggarwal, J. K., Cai, Q., (1999), "Human Motion Analysis: A Review", *Computer Vision and Image Understanding*, Vol. 73, No. 3, pp. 428 - 440.

Carroll, J. M., 1993, "Creating a design science of human-computer interaction", *Interacting with Computers*, Vol. 5, No. 1, pp. 3-12.

Cheng, S. W. and Dey, T. K., (1999), "Improved Construction of Delaunay Based Contour Surfaces", in *Proceedings of ACM Symposium on Solid Modeling and Applications*, pp. 322-323.

Chu, C. P., Dani, T. H., Gadh, R., (1997), "Multi-sensory User Interface for a Virtual-reality-based Computer-aided Design System", *Computer-Aided Design*, Vol. 29, Nr. 10, pp. 709-725.

Clark, J. H., (1976), "Designing Surfaces in 3-D", *Communications of the ACM*, Vol. 19, No. 8, pp. 454-460.

Deawele, G., Devernay, F., Horaud, R., (2004), "Hand Motion from 3D Point Trajectories and a Smooth Surface Model", *European Conference on Computer Vision*, Prague, 2004

Dix, A., Finlay, J., Abowd, G., Beale, R., (2001), "Human Computer Interaction", 2nd Edition, Pearson Education Limited, Harlon.

Dorozhkin, D. V., Vance, J. M., (2002), "Implementing Speech Recognition in Virtual Reality", in *Proceedings of ASME Design Engineering Technical Conferences and Computer and Information in Engineering Conference – DETC 2002*, Montreal, Sept. 29 – Oct. 2, 2002, ASME, CD-ROM, DETC2002/CIE-34390, pp. 1 - 5.

Estépar, R. S. J., Martín-Fernández, M., Alberola-López, C., Ellsemere, J., Westin, C. F., (2003), „Freehand Ultrasound Reconstruction based on ROI Prior Modeling and Normalized Convolution", in *Proceedings of the 6th International Conference on Medical Image Computing and Computer-Assisted Intervention*, LNCS 2879, pp. 382-390. November 15-18, 2003. Montreal, Canada

Flippo, F., Krebs, A., Marsic, I., (2003), "A Framework for Rapid Development of Multimodal Interfaces", *ICMI'03*, November 5-7, 2003, Vancouver, British Columbia, Canada

Gerritsen, B. H. M., "Using Weighted Alpha Complexes in Subsurface Modelling", Ph.D. Thesis, Delft University Press, Delft, 2001, pp. 1-224.

Grasso, M. A., Ebert, D. S., Finin, T. W., (1998), "The Integrality of Speech in Multimodal Interfaces", in *ACM Transactions on Computer-Human Interaction*, Vol. 5, No. 4, December 1998, pp. 303-325.

Gribnau, M. W., Hennessey, J. M., (1998), "Comparing Single- and Two-Handed 3D Input for a 3D Object Assembly Task", in *Proceedings of CHI'98 Conference on Human Factors in Computing Systems*, 18-23 April, Los Angeles, USA, pp. 233-234.

Hand, C., (1997), "A Survey of 3D Interaction Techniques", *Computer Graphics Forum*, Vol. 16, No. 5, pp. 269-281.

Heap, T.., Hogg, D., (1996), "Towards 3D Hand Tracking using a Deformable Model", in *Proceedings of Second International Conference on Automatic Face and Gesture Recognition*, pp. 140-145.

Heap,J. and Samaria, F., (1995), "Real-Time Hand Tracking and Gesture Recognition using Smart Snakes", in the *Proceedings of Interface to Real and Virtual Worlds*, Montpellier, June

Holden, E. J., Owens, R., (2001), "Visual Sign Language Recognition", R. Klette et. al.: Multi-Image Analysis, LNCS 2032, pp. 270-287

Horváth, I., Rusák, Z., (2001), „Collaborative Shape Conceptualization in Virtual Design Environments", *Communications of the ACM*, Vol. 44, No. 12, pp. 59-63.

Horváth, I., Tromp, N., Daalhuizen, J., (2003) "Comprehending a Hand Motion Language in Shape Conceptualization", in *Proceedings of ASME DETC*, 2-6 Septemb er, Chicago, IL. CD-ROM CIE-48286, pp. 1-15.

Iannizzotto, G., Villari, M., Vita, L., (2001), "Hand Tracking for Human-Computer Interaction with Graylevel VisualGlove: Turning back to the simple way", in *Proceedings of the Workshop on Perceptual User Interfaces*, November 15-16, Orlando, Florida, USA

Keefe, D. F., Feliz, D. A., Moscovich, T., Laidlaw, D. H., LaViola, J. J., (2001), “CavePainting: A Fully Immersive 3D Artistic Medium and Interactive Experience”, in *Proceedings of the 2001 Symposium on Interactive 3D Graphics*, March 19-21, Research Triangle Park, USA, pp. 85-93.

Ko, B. K., Yang, H. S., (1997), “Finger Mouse and Gesture Recognition System as a new Human Computer Interface”, *Computers&Graphics*, Vol. 21, No. 5, pp. 555-561

Kuczogi, Gy., Horváth, I. Vergeest, J. S. M., Rusák, Z., (2001), “Communication Model for the User Interface of a Shape Conceptualization System”, in *Proceedings of ASME Design Engineering Technical Conferences, 9-12 September*, 2001, ASME, New York, DETC2001/CIE-21300, pp. 1 - 14.

Kurata, T., Kato, T., Kourogi, M., Keechul, J., Endo, K., (2002), “A Functionally-Distributed Hand Tracking Method for Wearable Visual Interfaces and Its Applications”, In *Proceedings of IAPR Workshop on Machine Vision Applications*, Nara, Japan, pp.84-89

Lamar, M. V., Bhuyian, Md. S., Iwata, A., (2000), “T-CombNET – A Neural Network Dedicated to Hand Gesture Recognition”, *BMCV 2000*, LNCS 1811, pp. 613-622.

LaViola, J. J., (1999), “A Survey of Hand Posture and Gesture Recognition and Technology”, Technical Report CS-99-11, Brown University, Department of Computer Science, USA

Leganchuk, A., Zhai, S., Buxton, W., (1998), “Manual and Cognitive Benefits of Two-Handed Input: An Experimental Study”, *ACM Tansactions on Computer-Human Interaction*, Vol. 5, No. 4, Dec. 1998, pp. 326-359.

Li, F. W. B., Lau, R. W. H., Ng, F. F. C., (2003), “VSculpt: A Distributed Virtual Sculpting Environment for Collaborative Design”, in *IEEE Transactions on Multimedia*, Vol. 5, No. 4, pp. 570-580.

Liang, J.; Green, M. “Geometric Modeling Using Six Degrees of Freedom Input Devices”, (1993), in *Proceedings of 3rd international Conference on CAD and Computer Graphics*, August 23-26, Beijing, China, pp. 217-222.

Ma, L., Lau, R. W. H., Feng, J., Peng, Q., Wong, J., (1997), “Surface Deformation Using the Sensor Glove”, *ACM VRST ’97*, Lausanne, Switzerland

Matsumiya, M., Takemura, H., Yokoya, N., (2000), “An Immersive Modeling System for 3D Free-form Design Using Implicit Surfaces”, in *Proceedings of the ACM Symposium on Virtual Reality Software and Technology*, October 22-25, Seoul, Korea, pp. 67-74.

Murakami, T., Nakajima, N., (2000), “DO-IT: Deformable Object as Input Tool for 3-D Geometric Operation”, *Computer-Aided Design*, Vol. 32, No.1, pp. 5-16

Nishino, H., Utsumiya, K., Korida, K., (1998), “3D Object Modeling Using Spatial and Pictographic Gestures”, *VRST’98*, Taipei, Taiwan, Nov. 2-5

Ogawara, K., Hashimoto, K., Takamatsu, J., Ikeuchi, K., (2003), “Grasp Recognition using a 3D Articulated Model and Infrared Images”, in *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems*, October 27-31, Las Vegas, USA

Oviatt, S. and Cohen, P., (2000), “Multimodal Interfaces that Process What Comes Naturally”, in *Communications of the ACM*, March 2000, Vol. 43, No. 3, pp. 45-53.

Pavlovic, V., Sharma, R., Huang, T., (1997), “Visual Interpretation of Hand Gestures for Human-Computer Interaction: A Review, *IEEE Transactions on PAMI*, Vol.19, No.7, July, pp. 677 - 695.

Rehg, J. M., Kanade, T., (1994), “DigitEyes: Vision-Based Hand Tracking for Human-Computer Interaction”, in Proceedings of the *IEEE Workshop on Motion of Non-Rigid and Articulated Objects*, Austin, Texas, November, pp. 16-22.

Rieger, T., (2003), “Avatar Gestures”, *Journal of WSCG 2003*, Vol. 11, No. 2, pp. 379-386.

Rusák, Z., (2003), “Vague Discrete Interval Modeling for Product Conceptualization in Collaborative Virtual Design Environments”, Ph.D. Thesis, Millpress, Rotterdam, pp. 1-159.

Schkolne, S., Pruett, M., Schröder, P., (2001), “Surface Drawing: Creating Organic 3D Shapes with the Hand and Tangible Tools”, *CHI 2001*, 31 March – 5 April

Segen, J., Kumar, S., (1998), “Video-Based Gesture Interface to Interactive Movies”, *ACM and ATR Workshop on Technologies for Interactive Movies*, Sep 13, 1998

Segen, J., Kumar, S., (1999), “Shadow Gestures: 3D Hand Pose Estimation using a Single Camera”, *IEEE International Conference on Computer Vision and Pattern Recognition*, Fort Collins, June

Shaw, C., Green, M., (1994), “Two-handed Polygonal Surface Design”, in *Proceedings of the 7th ACM Symposium on User Interface Software and Technology*, November 2-4, Marina del Rey, USA, pp. 205-212.

Stefani, O., Rauschenbach, J., (2003), “3D Input Devices and Interaction Concepts for Optical Tracking in Immersive Environments”, in *Proceedings of the 7th International Workshop on Immersive Projection Technology, 9th Eurographics Workshop on Virtual Environments*, 22-23 May, Zurich, Switzerland, pp. 317-318.

Stenger, B., Mendonca, P. R. S., Cipolla, R., (2001), “Model-Based 3D Hand Tracking of an Articulated Hand”, in *Proceedings of Conference on Computer Vision and Pattern Recognition*, Vol. II, pages 310-315, Kauai, USA, December

Utsumi, A., Ohya, J., (1999), “Multiple-Hand-Gesture Tracking using Multiple Cameras” in Proceedings of CVPR’99

Várady, T., Martin, R. R., Cox, J., (1997), „Reverse engineering of geometric models – an Introduction”, in *Computer-Aided Design*, Vol. 29, No. 4, pp. 255-268.

Vergeest, J. S. M., Spanjaard, S., Horváth, I., Jelier, J. J. O., (2001), “Fitting Freeform Shape Patterns to Scanned 3D Objects”, in *Transactions of the ASME*, Vol. 1, September 2001, pp. 218-224

Watson, R., (1993), “A Survey of Gesture Recognition Techniques”, Technical Report TCD-CS-93-11

Weimer, D., Ganapathy, S. K., (1989), "A Synthetic Visual Environment with Hand Gesturing and Voice Input", in *Proceedings of CHI '89*, May 1989, pp. 235-240.

Wesche, G., Seidel, H. P., (2001), "FreeDrawer - A Free-Form Sketching System on the Responsive Workbench", in *Proceedings of the ACM symposium on Virtual reality software and technology,* November 15-17, Baniff, Alberta, Canada, pp. 167-174.

Wong, J. P. Y., Lau, R. W. H., Ma, L., (1998), "Virtual 3D Sculpturing with a Parametric Hand Surface", in *Computer Graphics International '98*, Germany, pp. 178-186.

Wu, Y. and Huang, T. S., (1999), "Capturing Articulated Human Hand Motion: A Divide and Conquer Approach", in *Proceedings of IEEE International Conference on Computer Vision*, Greece, September 1999, pp. 606-611.

Wu, Y., Huang, T. S., (2000), "View-independent Recognition of Hand Postures", in *Proceedings of IEEE Conference on CVPR'2000*, Hilton Head Island, SC, Vol. II, pp. 88-94.

Wu, Y., Huang, T. S., (1999), "Vision-Based Gesture Recognition: A Review", *International Gesture Workshop*, Gif-sur-Yvette, France, March

Xu, S., Fok, S. C., Tor, S. B., (2000), "Rapid Creation and Direct Manipulation of Free-form Curves Using Data Glove", in *Proceedings of EDA 2000 Conference*, Orlando, Florida, USA, pp. 954-959.

Zhang, Y., Rohling, R., Pai, D. K., (2002), "Direct Surface Extraction from 3D Freehand Ultrasound Images", in *Proceedings of IEEE Visualization 2002*, Oct. 27-Nov. 1, Boston, MA, USA.