

HAND MOTION PROCESSING IN APPLICATIONS: A CONCISE SURVEY AND ANALYSIS OF TECHNOLOGIES

E. Varga, I. Horváth, Z. Rusák and J. J. Broek

Keywords: hand motion processing, hand motion language, detection technologies, information extraction, surface regeneration

1. Introduction

Human-computer interaction is an evergreen issue of computer-aided design. 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 so far. 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. For this reason, new input means such as verbal control, hand gestures and motions, and reverse engineering of objects have been considered as constituents of highly interactive multi-modal user interfaces [Chu 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. 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.

Efforts have been made in research related to hand motion detection, and a Hand Motion Language (HML) has been proposed, which can be applied in shape conceptualization [Horváth 2003]. Contrary to the efforts, we are still far from the ultimate solution. The main reasons are the complexity of the problem and the lack of dedicated hand motion processing technologies.

Most technical and survey papers published earlier addressed the hand motion processing (HMP) problem from a technological point. In this paper we investigate the HMP from the aspect of real-time information extraction and conversion, which is required by creative shape conceptualization. Our aim was not only to analyze the available technologies, but also to compare them based on the operational parameters and the features of information processing. Section 2 introduces a simple reasoning model to categorize the technologies and systematize the investigation of their conformance. The following sections identify and analyze typical technologies and processing approaches in four categories. Section 7 compares the various HMP techniques taking into consideration the requirements originating in shape conceptualization.

2. Introducing the reasoning model

The ultimate goal of our research is to generate swept surfaces by detecting and processing hand motions in 3D space. In order to achieve this goal we have to extract sufficient 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 a dimension in our reasoning about technologies.

A second important 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. 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 (Fig. 1).

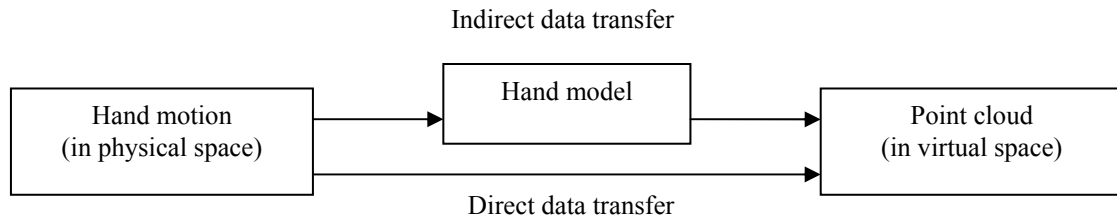


Figure 1. Information processing during hand motion processing

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 the 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. 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.

On the basis of this, the contact and non-contact technologies can be sorted into processing categories, which have been called (i) *direct incomplete*, (ii) *direct complete*, (iii) *indirect incomplete* and (iv) *indirect complete* HMP. These are discussed in the next Sections, in which, besides sorting the technologies into the above categories, 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 the last Section of the paper we will investigate the compliance of the various technologies with hand motion based shape conceptualization. Due to the space limitation in this paper, we could involve only a very limited number of, but characteristic, technologies in the analysis.

3. Direct incomplete hand motion processing

This form of HMP extracts 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 in order to construct the intended shape. Here, we analyze only two typical approaches of direct incomplete HMP, which are representative of the contact and non-contact technologies.

Xu, S. et al. [2000] proposed a method for creating and manipulating free form curves by hand motion. They detect the tip of the user's index finger during its motion with the help of a data glove and reconstruct a curve from these points. They developed algorithms for filtering the points in order to preserve the fidelity of the curve according to the hand motion. This technology requests the designers to wear a data glove, and does not support obtaining surface information.

Abe, K. et al. [2000] developed a 3D drawing system to realize easy drawing tasks with hand motion. 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 fingertip position 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 parameter. The average time of hand pose recognition is 10 fps (0.10 second per one frame). They reported that they could draw images without feeling delay because of the short processing time.

4. Indirect incomplete hand motion processing

As the previous one, this HMP approach extracts only a limited amount of information from the moving hand due to technological limitations or to striving for short cycle time. Usually 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. 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.

Another representative of this category, Wu et al. [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. They employed a kinematical hand model, where each finger was modeled as a kinematical chain with the palm as its reference frame. 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.

5. Direct complete hand motion processing

This form of HMP endeavors extracting sufficient amount of information from the moving hand, and transfers this information directly to the geometric modeling system. Both contact and non-contact technologies have been developed. Nishino et al. [1998] proposed a method for modeling of 3D objects based on hand motion, detected 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. Another proposal, T-CombNET is a non-contact, neural network based system for hand gesture recognition [Lamar 2000]. A single camera was used to capture hand motions, which are elements of a sign language. The system understands static hand postures and also dynamic hand gestures. In order to preserve adaptability, five non-native sign language speakers generated the database of gestures.

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. Heap and Hogg [1996] applied a deformable 3D hand model for tracking hand motion with six degrees of freedom, using a single video camera. The hand model was constructed in order to preserve fidelity, by using information from real hands in various positions. The model is projected onto input images and an algorithm was developed to move and deform the model to fit the image. Based on the hand model, hand motion can be tracked in real-time. Although, a large amount of position information can be obtained from this model, the problem of self-occlusion is not solved. Developed ten years ago, DigitEyes is a system for vision-based hand tracking [Rehg 1994]. It employs a kinematic hand model, which is used to solve a 3D mouse user interface problem. The 3D configuration of the hand is estimated from a sequence of images and with the hand model. The tracking speed is up to 10 fps. In the corresponding 3D graphical mouse application a simplified eight state model was used. The palm was constrained to lie in the plane of a table, and a camera was placed at approximately 45 degrees to the tabletop.

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 I_H be the amount of information obtained from the moving hand; I_M the information that is derived from a virtual hand model, and I_S the information needed to reconstruct the swept surface. Figure 2 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.

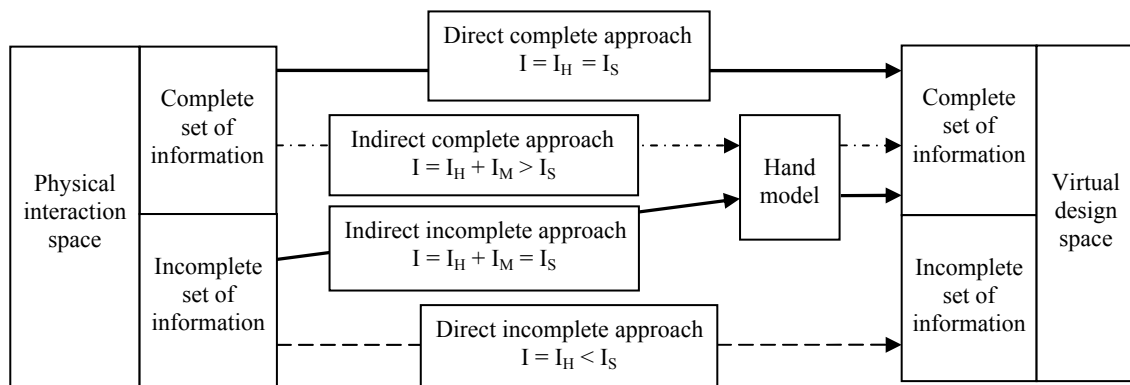


Figure 2. Comparison of HMP approaches

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 involved in the two processes are 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. In principle a process involving less number of actions can be 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.

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 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. 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 movement envelope 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 sensed 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 [Wagner 1988]. 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. Our research has got a dual focus: we are simultaneously investigating the compliance of HMP technologies and redesigning and optimizing 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.

Based on the investigations it can be stated that direct complete (DC) and indirect incomplete (II) HMP technologies offer the largest potential to support hand motion based shape conceptualization. With respect to our goal we give privilege to non-contact technologies. We found that the studied technologies show significant differences in terms of speed, adaptability, and fidelity. The advantage of the DC 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. II HMP technologies target landmarks or other characteristic points only. They require less scanning time and less sensitive and powerful technology, which is an important issue from an operational point of view. Based on the literature study the HMP technologies could only qualitatively be assessed for applicability in hand motion based shape conceptualization. Consequently, more experimental research is needed to decide on which concrete technology meets the operational requirements best.

References

- Abe, K., Saito, H., Ozawa, S., "3-D Drawing System via Hand Motion Recognition from Two Cameras", in *Proceeding of the 6th Korea-Japan Joint Workshop on Computer Vision, 21-22 January 2000*, pp. 138-143.
- Chu, C. P., Dani, T. H., Gadh, R., "Multi-sensory User Interface for a Virtual-reality-based Computer-aided Design System", *Computer-Aided Design, Vol. 29, Nr. 10, 1997*, pp. 709-725.
- Gerritsen, B. H. M., "Using Weighted Alpha Complexes in Subsurface Modelling", *Ph.D. Thesis, Delft University Press, Delft, 2001*, pp. 1-224.
- Heap, T., Hogg, D., "Towards 3D Hand Tracking using a Deformable Model", in *Proceedings of Second International Conference on Automatic Face and Gesture Recognition, 1996*, pp. 140-145.
- Horváth, I., Tromp, N., Daalhuizen, J., "Comprehending a Hand Motion Language in Shape Conceptualization", in *Proceedings of ASME DETC 2003, 2-6 September, 2003, Chicago, IL. CD-ROM CIE-48286*, pp. 1-15.
- Lamar, M. V., Bhuiyan, Md. S., Iwata, A., "T-CombNET – A Neural Network Dedicated to Hand Gesture Recognition", in *Proceedings of BMCV 2000, LNCS 1811, 2000*, pp. 613-622.
- Nishino, H., Utsumiya, K., Korida, K., "3D Object Modeling Using Spatial and Pictographic Gestures", in *Proceedings of VRST'98, Taipei, Taiwan, 2-5 November, 1998*, pp. 51-58.
- Rehg, J. M., Kanade, T., "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 1994*, pp. 16-22.
- Rusák, Z., "Vague Discrete Interval Modeling for Product Conceptualization in Collaborative Virtual Design Environments", *Ph.D. Thesis, Millpress, Rotterdam, 2003*, pp. 1-159.
- Wagner, Ch., "The pianist's hand: anthropometry and biomechanics", in *Ergonomics, Vol. 31, No. 1, 1988*, pp. 97-131.
- Weimer, D., Ganapathy, S. K., "A Synthetic Visual Environment with Hand Gesturing and Voice Input", in *Proceedings of CHI '89, May 1989*, pp. 235-240.
- Wu, Y., Huang, T. S., "Capturing Articulated Human Hand Motion: A Divide-and-Conquer Approach", in *Proceedings of the IEEE International Conference on Computer Vision, Greece, 1999*, pp. 606-611.
- Xu, S., Fok, S. C., Tor, S. B., "Rapid Creation and Direct Manipulation of Free-form Curves Using Data Glove", in *Proceedings of EDA 2000, Orlando, Florida, USA, 2000*, pp. 954-959.

Ms. Edit Varga, PhD student
Faculty of Industrial Design Engineering, Delft University of Technology
Landbergstraat 15, 2628 CE Delft, The Netherlands
Telephone: +31 015 278 9321, Telefax: +31 015 278 1839
E-mail: e.varga@io.tudelft.nl