A ROBUST HAND GESTURE TRACKING SYSTEM USING A MARKED GLOVE

Alexander Klaeser [1,2], John K. Tsotsos [1]

[1] Center for Vision Research, York University
[2] University of Applied Sciences Bonn-Rhein-Sieg

alexander.klaeser@epost.de, tsotsos@cs.yorku.ca

ABSTRACT

In this paper we present a simple, yet effective and robust way of hand gesture tracking in an environment with cluttered background. The approach used is a tracking of six markers attached to a glove, one marker on each fingertip as well as on the thumb and one on the wrist. After initiation, each marker is tracked independently, what makes this system robust to rotation. To make the tracking more reliable, model knowledge of the hand is used. As a demonstration of the possibilities of this approach, a music application is presented which controls sound and effects, thus turning a PC into a visual-guided instrument.

1. INTRODUCTION

Various interfaces exist for human-computer interaction. Definitely the most intuitive interface for this interaction is still the human hand, since the hand is naturally used for manipulation, controlling, communication and interaction between human beings and machines.

There is a great deal of effort made on capturing hand motion and gestures, either with the help of data gloves or with the help of visual approaches. One big advantage of visual approaches in capturing hand motion is that there is no need for devices directly attached to the hand. Thus interaction is more comfortable and more intuitive for the user. The visual approach also moves the main focus away from developing a functional hardware device towards developing a functional software component, which is less expensive and easier to comprehend.

Of course the visual approach also creates new problems and difficulties. Reliable real-time hand gesture tracking on image sequences of human hands with a cluttered background is still too complex and therefore too demanding with respect to computing power.

The idea of this work is therefore to make a sensible trade-off between the comfort of a deviceless but computer power demafnding and a hardware based hand gesture tracking system. The approach presented in this paper is the tracking of a marked glove, which is still very comfortable to wear for the user, but at the same time it also is much easier to track a marked glove rather than a real human hand in a software application.

Another idea of this work is to be able to capture the natural motion of a human hand and not to recognize a more or less fixed set of predetermined gestures. In that way it is possible to use the obtained information in various ways. One way is to apply a controlling function to each finger individually as it is shown in the sound and effect controlling demo application at the end of this paper. Another way could also be to recognize gestures not only based on the current position but also on the previous motion of the fingers as well.

Fig. 1. *Glove with six markers on the fingers, thumb and the wrist.*

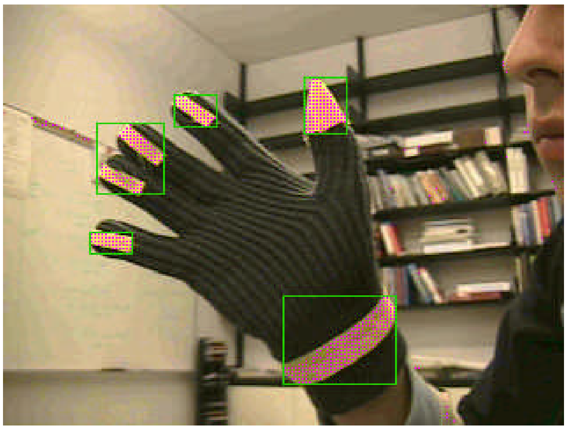

Fig. 2. *Sample result of a rectangular merging algorithm.*

2. RELATED WORK

Probably the most related work to ours is presented in [4] and [5]. [4] uses a similar kind of glove for the hand gesture recognition, but a gesture in their work is defined as a particular movement of the fingers, we focused on a general hand tracking. [5] divides the hand in two dimensional segments which are connected to each other (palm and finger segments) and tries then to match the defined segments to the hand, so this work focuses on a general hand tracking as well, but with a different approach. Our work also aims to realize a hand tracking that works in a realistic, cluttered environment.

Apart from related papers, [1], [2] and [3] give a very good overview, what work has been done in the past.

3. GLOVE AND HAND MODEL

The glove that is used is a simple black cotton glove marked with six yellow stripes of cloth (see also Fig. 1). It is quite similar to the one used in [4] where a glove with white markers on the fingertips and the thumb not on the wrist is used. On each finger (little, ring, middle and index finger) is a stripe attached going from the front over the tip to the back. In that way there is always yellow cloth visible when a finger bends down. In case the fingers are close to each other, there is still enough space so that the markers are visually separated from each other. As for the tip of the thumb the whole tip has been covered with a piece of yellow cloth, since the thumb can be moved very independently. The last marker is attached on the wrist.

The fact that just one color is used for all markers on the hand makes it easier to find the markers in the camera image. If multiple colors would have been used for marking it would be more difficult and more work to calibrate the different interesting color spaces for each color that is used.

The resulting hand model is then a fairly simple model of the hand consisting of all six features marked on the glove (see also Fig. 3). All fingers and the thumb are connected to the wrist.

4. IMAGE SEGMENTATION

In order to achieve a good color segmentation of the image, a HSV color space describing the color of the markers on the glove defined by minimum and maximum H, S and V values is used. All interesting colors that lie in the range of the above specified color space then have to be merged together to regions of interest or segments. This task is done by traversing the picture and using a region-growing algorithm to merge connected pixels together to one segment. After the whole image has been segmented, all segments are ordered by their size from high to low. Segments with a size under a certain threshold are deleted.

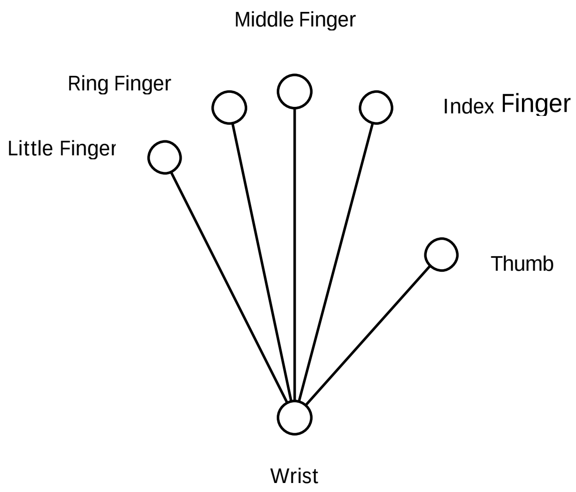

Fig. 3. *Used hand model based on the feature markers.*

One big advantage of this method compared to a bounding box method is that it separates all interesting regions very precisely, with no regard to its shape. So in this way it is no problem to be able to segment markers of a rotated hand, whereas with the bounding box method rotated markers which are close to each other would be merged into one box (see also Fig. 2).

5. HAND INITIALISATION

To be able to track the hand means that there must be knowledge of which marker represents which hand feature (i.e. index, middle, ring and little finger as well as the wrist). To gain this knowledge there is currently a very simple initiation performed.

It is assumed that the wrist feature is the heaviest segment (i.e. the segment with the biggest amount of pixels) of the image. After that the next heaviest five regions are being taken and labeled as little, ring, middle and index finger from left to right (see also Fig. 4, picture in the top left). This method assumes that all finger tips point upward.

A more sophisticated and rotation-independent initiation method has been developed, but not yet implemented. This method again assumes that the heaviest segment in the image is the wrist. Starting from that segment it investigates the next heaviest segments in regard to their angles. If those segments have approximately the same distance to the wrist segment and a valid angle (i.e. not too big) to each other it can be assumed that this is a possible hand. If this combination is not valid other segments can be chosen for the fingers and the thumb and even for the wrist. In that way various possible permutations could be investigated in regard to their validity.

6. HAND TRACKING

The tracking of the hand means to use the known feature positions of the last video frame (either known by initialization or by tracking) and find their corresponding segments in the current frame. The tracking in this work has been realized by a computation of a restricted number of possible permutations and comparing them. Furthermore model knowledge of the hand is added to the process in order to exclude tracking results that are not possible.

6.1. Finding the right permutation

Assuming all possible tracking permutations from the features of the last frame to the segments of the current frame are known. Then within all possible permutations there is just one permutation that describes the right tracking from the last to current frame. If the distance of each tracking permutation to the feature position of the last frame, i.e. the distance of each feature position of the last frame to its corresponding position in the current permutation, is known, then the one permutation with the least overall distance is most likely the right permutation.

It is not necessary to compute all possible permutations, since it can be assumed that a correspondence between a feature of the last frame and a segment of the current frame that is very far away from the feature is not really possible. So the bigger the distance between a feature of the last frame and a segment of the current frame the smaller the possibility of a correspondence. Based on this assumption our results showed that it is sufficient to take the closest three segments to each feature and compute all possible permutations over this subset.

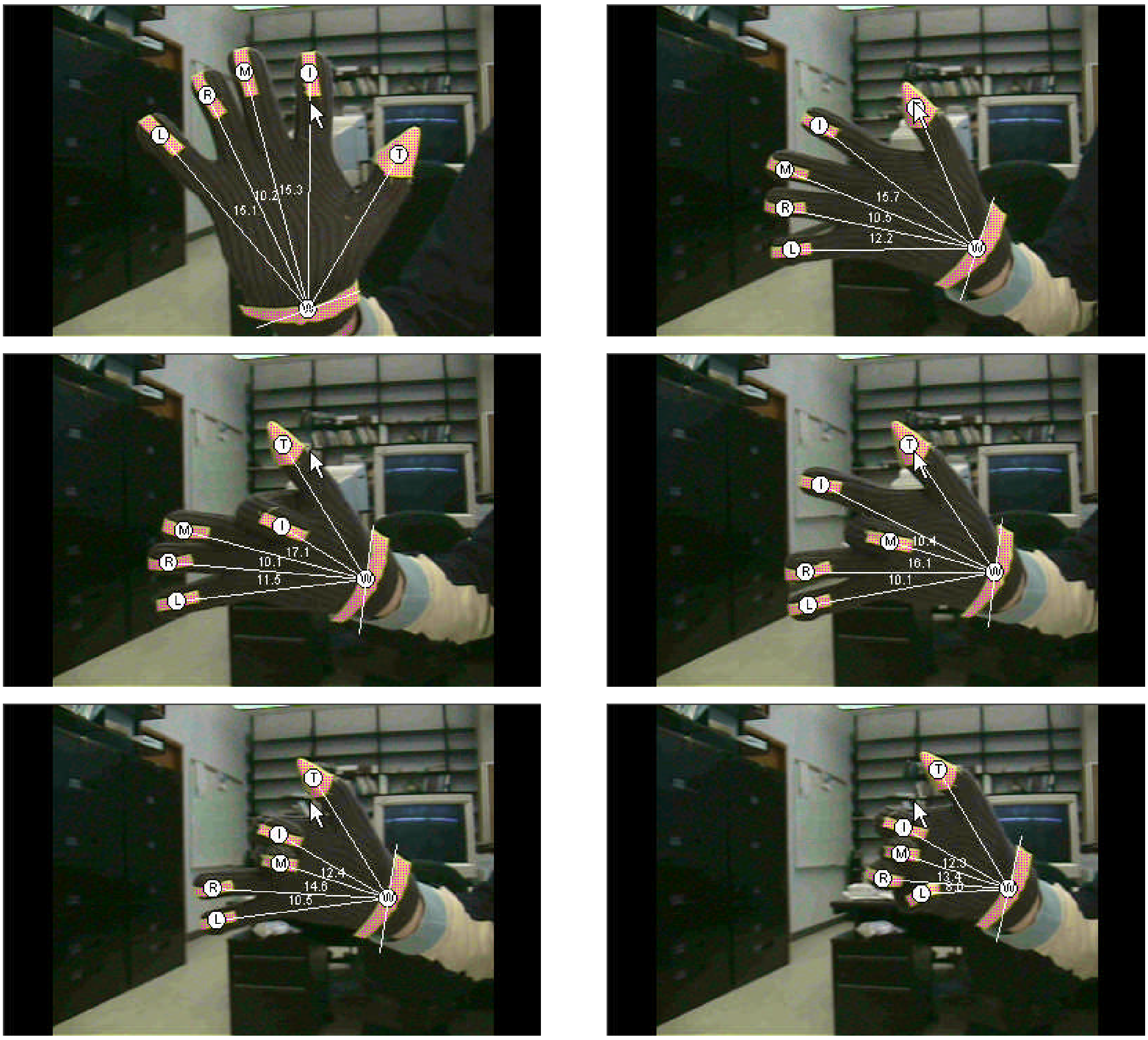

Fig. 4. *Some results of the hand tracking. L (little finger), R (ring finger), M (middle finger), I (index finger), T (thumb) and W (wrist) label the features.*

In difference to the finger and thumb features, the wrist feature has no very close adjacent features. So instead of calculating all possible permutations in regard to the wrist feature it can be assumed that the segment in the current frame that is the closest one to the wrist feature in the last frame is its right corresponding segment as well. This assumption can't be made in regard to the finger and thumb features. For instance a very fast hand rotation is possible so that the segment of an adjacent finger is closer to a finger feature of the last frame then the right corresponding finger segment.

With those two assumptions that limit the number of calculations it is possible to find the right permutation very efficiently.

6.2. Using model knowledge

Once the ideal permutation is found this permutation is investigated for validity. For this evaluation it is assumed that the angle between every finger feature and its adjacent finger feature lies in a certain range. In order to compute the angle, the vector starting at the wrist and ending at the finger feature is used.

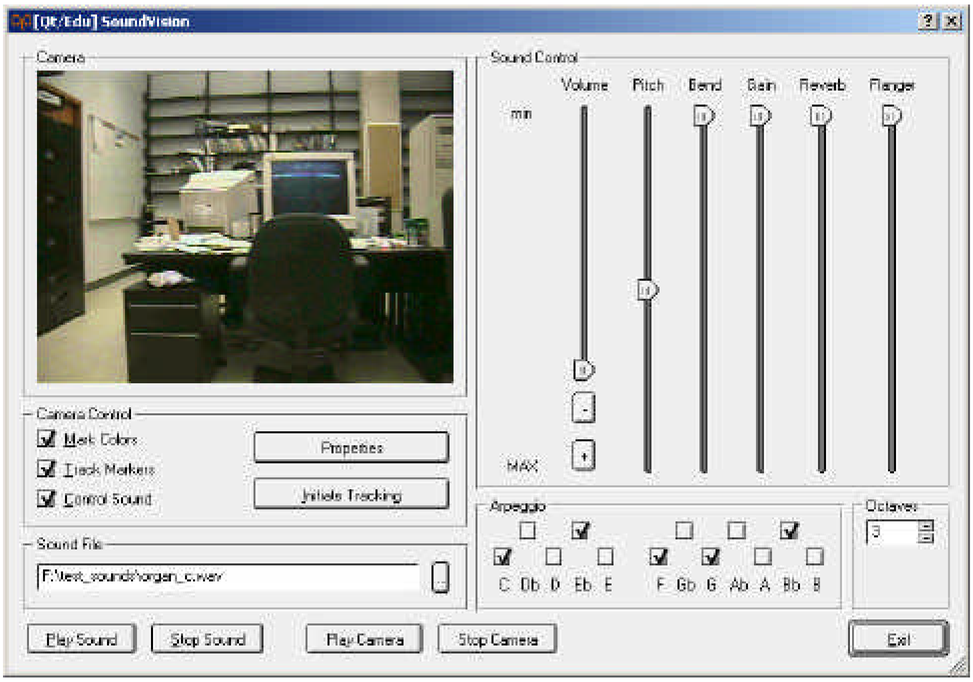

Fig. 5. *Screenshot of the demo sound application. In the left top there is the camera window, the right shows control slides for the effects as well as the piano keyboard on the bottom.*

Results in this work showed that a range from -10 to 30 degrees worked well. This assumption is just applied on the fingers, but not on the thumb, since the thumb is too independent in its motion.

7. RESULTS

7.1. Hand tracking results

Despite its simplicity, the results of this presented hand gesture tracking system are very good (see also Fig. 4). A reliable real-time hand tracking even with fast hand and finger movements is possible. For the experiments we used frame rates around 15 frames/second and the system worked pretty well on a Pentium III with 933 MHz and 256 RAM. For the video camera grabbing we used a Panasonic KX-DP702 video camera, an ATI Radeon graphics card with a video input as capture device and Microsoft DirectShow (DirectX 8) as software platform.

Problems occurred when makers were not visible in the camera image (through occlusion) or when markers overlapped each other so that they were considered to be one region of interest. But as soon as the hand was detectable again, the system was able to continue capturing. Another problem occurred with very fast hand motion and rotation. The system then lost track of the markers and labeled them wrong.

Nevertheless the results showed that this system works very well and reliable. As for the described problems, there are various approaches to solve them.

7.2. A demo application

We developed an application to practically test the hand tracking and see how intuitive and powerful controlling with our hand gesture tracking is (see also Fig. 5). The application we developed is a sound application with which it is possible to control the pitch of a note as well as four different effects at once.

The application is based on Qt developed by Trolltech and it uses Microsoft DirectAudio (DirectX 8) to control the sound as well as to generate and control all effects. This application was tested on a Pentium III with 933 MHz and 256 RAM.

7.2.1. Description

To make the pitch control more interesting (and more musical too) we added a small piano keyboard in the range of one octave on which those notes can be marked which shall be played. After that the range of octaves can be chosen, i.e. over how many octaves the chosen notes shall be played. To control the note pitch, the horizontal wrist position is used.

Four different effects are then applied to the fingers. A distortion effect controlling the amount of distortion is assigned to the index finger, a flanger effect controlling its frequency to the middle finger, a reverb effect controlling the depth of the reverberation to the ring finger and an one full tone downward bending effect to the little finger. A stretched finger corresponds to the minimum and a finger fully bended towards the wrist corresponds to the maximum of each effect. Within its range each effect can be freely adjusted.

In order to capture movements towards and away from the camera as well, it is assumed that the thumb always roughly stays in its position. In this way it is possible to continue controlling the effects while changing the distance to the camera.

7.2.2. Results

Again the results satisfactory, the gesture tracking worked well and the controlling of the tone was intuitive, it took people just a short period of time to get used to the interface. But then it really was a lot of fun to experiment with tone, especially experimenting with the variety of manipulating the tone. It was even possible to improvise over for instance a blues playback.

A real problem with the controlling didn't occur. Just the controlling with the little finger was not as easy as controlling the other effects with other fingers, since the little finger of the human hand is not as independent as the other fingers, it is usually always connected to the ring finger in its movement.

8. CONCLUSION

In this paper we presented a simple but effective approach to capture natural hand motion based on a tracking of a marked glove. We tested the system with a sound application that controls pitch and effects of a tone and proofed its reliability and intuition. We also pointed out the variety of tasks that could be done with this system. Further work will involve especially the problems discussed in 7.1, but also a development of other applications that are actually using the hand gesture tracking system.

REFERENCES

[1] Y. Wu, T. Huang (2001): "Hand modeling, analysis and recognition"

[2] V. Pavlovic, R. Sharma, T.S. Huang (1997): "Visual interpretation of hand gestures for human computer interaction: A review"

[3] T. Huang, V. Pavlovic (1995): "Hand Gesture Modeling, Analysis, and Synthesis"

[4] J. Davis, M. Shah (1994): "Visual gesture recognition"

[5] Ying Wu, John Y. Lin and Thomas S. Huang (2001): "Capturing Natural Hand Articulation"

[6] R. O'Hagan, A. Zelinsky (2000): "Visual Gesture Interfaces for Virtual Environments"

[7] J. Rehg, T. Kanade (1993): "DigitEyes: Vision-Based Human Hand Tracking"

[8] T. Heap and D. Hogg (1996): "Towards 3D hand tracking using a deformable model"

[9] R. Herpers, K. Derpanis, W. MacLean, G. Verghese, M. Jenkin, E. Milios, A. Jepson, J.K. Tsotsos: "SAVI: an actively controlled teleconferencing system"