Pointing Gesture Recognition based on 3D-Tracking of Face, Hands and Head Orientation

Kai Nickel
nickel@ira.uka.de

Rainer Stiefelhagen
stiefel@ira.uka.de

Interactive Systems Laboratories
Universität Karlsruhe (TH)
Germany

ABSTRACT

In this paper, we present a system capable of visually detecting pointing gestures and estimating the 3D pointing direction in real-time. In order to acquire input features for gesture recognition, we track the positions of a person's face and hands on image sequences provided by a stereo-camera. Hidden Markov Models (HMMs), trained on different phases of sample pointing gestures, are used to classify the 3D-trajectories in order to detect the occurrence of a gesture. When analyzing sample pointing gestures, we noticed that humans tend to look at the pointing target while performing the gesture. In order to utilize this behavior, we additionally measured head orientation by means of a magnetic sensor in a similar scenario. By using head orientation as an additional feature, we observed significant gains in both recall and precision of pointing gestures. Moreover, the percentage of correctly identified pointing targets improved significantly from 65% to 83%. For estimating the pointing direction, we comparatively used three approaches: 1) The line of sight between head and hand, 2) the forearm orientation, and 3) the head orientation.

Categories and Subject Descriptors

I.4.8 [**Image Processing and Computer Vision**]: Scene Analysis

General Terms

Algorithms, Experimentation, Human Factors

Keywords

pointing gestures, gesture recognition, person tracking, computer vision

ICMI'03, November 5–7, 2003, Vancouver, British Columbia, Canada.

1. INTRODUCTION

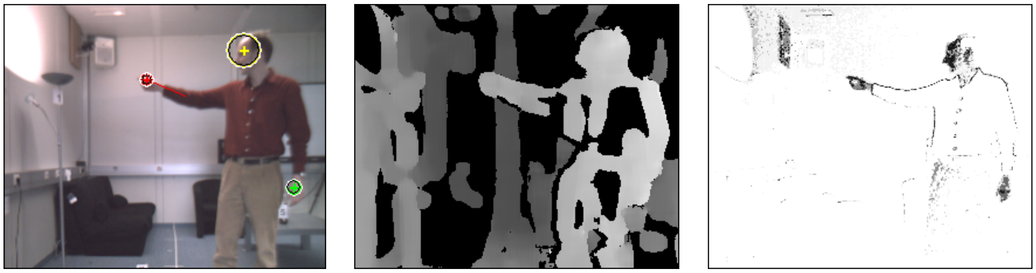
In the concept of multimodal user interfaces, users are able to communicate with computers using the very modality that best suits their current request. Apart from mouse or keyboard input, these modalities include speech, handwriting or gesture. Among the set of gestures intuitively performed by humans when communicating with each other, pointing gestures are especially interesting for applications like smart rooms, virtual reality or household robots.

A pointing gesture in the context of this paper is a movement of the arm towards a pointing target. Humans perform the gesture in the communication with others to mark a specific object, location or direction. Pointing gestures are often used in combination with speech, as they can help to resolve ambiguities and specify parameters of location in verbal statements ("Switch *that* light on!"). In the recognition of pointing gestures, two problems have to be addressed: the detection of the occurrence of the gesture in natural arm movements and the estimation of the pointing direction.

In this paper, we present a system that is able to detect pointing gestures and to determine the 3D pointing direction in real-time. To obtain input features for gesture recognition, we first track a person's head and hands in 3D. To detect the occurrence of pointing gestures in the input feature stream, we use Hidden Markov Models which were trained on different phases of sample pointing gestures (begin, hold, end). We evaluate this approach in a scenario with 10 different persons.

A body of literature suggests that people naturally tend to look at the objects with which they interact. Maglio et al. [1] for instance investigated how people use speech and gaze when interacting with an "office of the future". They report that the subjects nearly always looked at a speech-enabled office device before addressing it. Similar results are reported by Brumitt et al. [2]. They investigated how people use different interfaces to control room lights. They also report that subjects typically looked at the lights they wanted to control.

To analyze whether this behavior could be used to improve our pointing gesture recognition system, we tracked head and hand positions as well as people's head orientations in a second experiment. By using additional features derived from head orientation in our feature vector, we could observe significant gains in both recall and precision of pointing gesture detection. In addition, the percentage of correctly identified pointing targets improved significantly.



a. Left camera image b. Disparity map c. Skin-color map

Figure 1: Tracking of head and hands is based on skin-color classification (dark pixels represent high skin-color probability) and stereoscopic range information.

The main contributions of this paper are the following: 1) We present a robust approach for real-time 3D tracking of head and hands using color and range information. 2) We describe how Hidden Markov Models can be trained to detect the occurrence of a pointing gesture in the 3D input feature stream. 3) We experimentally show the usefulness of head orientation tracking to improve both pointing gesture detection and estimation of pointing direction.

The remainder of this paper is organized as follows: In Section 2 we describe our approach to track a user's head and hands in 3D from stereo images. In Section 3 we describe how we use Hidden Markov Models to detect pointing gesture, describe the used features and present experimental results. In Section 4 we describe experiments with additionally used head orientation features. Section 5 compares three different approaches for the estimation of pointing direction: using a) the line of sight between head and hand, b) the estimated forearm orientation and c) head orientation. Finally, we conclude the paper in Section 6.

1.1 Related Work

There are numerous approaches for the extraction of body features by means of one or more cameras. In [3], Wren et al. demonstrate the system *Pfinder*, that uses a statistical model of color and shape to obtain a 2D representation of head and hands. Azarbayejani and Pentland [4] describe a 3D head and hands tracking system that calibrates automatically from watching a moving person. An integrated head and silhouette tracking approach based on color, dense stereo processing and face pattern detection is proposed by Darrell et al. [5]. Compared to these works, the characteristic of our approach lies in the early assignment of dense stereo to skin-color information, thus allowing for combined clustering.

Hidden Markov Models have been used for years in continuous speech recognition [14], and have also been applied successfully to the field of gesture recognition: In [6], Starner and Pentland were able to recognize hand gestures out of the vocabulary of the *American Sign Language* with high accuracy. Becker [7] presents a system for the recognition of *T'ai Chi* gestures based on head and hand tracking. In [8], Wilson and Bobick propose an extension to the HMM framework, that addresses characteristics of parameterized gestures, such as pointing gestures. Poddar et al. [9] recognize different hand gestures (including pointing gestures) performed by a TV weather person. They combine an HMM-based detection of gestures on head and hand movements with spoken keywords.

Kahn et al. [10] demonstrate the use of pointing gestures to locate objects. Their system operates on various feature maps (intensity, edge, motion, disparity, color). Jojic et al. [11] detect and estimate pointing gestures solely in dense disparity maps. Unlike these approaches, we model the dynamic motion of pointing gestures and not the static posture.

2. TRACKING HEAD AND HANDS

In our approach we combine stereoscopic range information and skin-color classification in order to achieve a robust tracking performance. The setup consists of a fixed-baseline stereo camera connected to a standard PC. We use a commercially available library [12] to calculate a dense disparity map made up of pixel-wise disparity values, and to provides 3D coordinates for each pixel (Fig. 1b). A histogram-based model represents the distribution of human skin-color in the chromatic color space (rg color space). In order to initialize and maintain the model automatically, we search for a person's head in the disparity map of each frame. Similar to an approach proposed in [5], we first look for a human-sized connected region, and then check its topmost part for head-like dimensions. Pixels inside the head region contribute to the skin-color model. The result of the classification by means of the the model is the skin-color map (Fig. 1c), that provides the skin-color probability for each pixel.

In order to find potential candidates for the coordinates of head and hands, we search for connected regions in the morphologically filtered skin-color map. For each region, we calculate the centroid of the associated 3D pixels. If the pixels belonging to one region vary strongly with respect to their distance to the camera, the region is split by applying a k-means clustering method. We thereby separate objects that are situated on different range levels but accidentally merged into one object in the 2D image. In the clustering procedure, all pixels are weighted by their individual skin-color probabilities.

2.1 Search for best Hypothesis

The task of tracking consists in finding a good hypothesis s_t for the positions of head and hands at time t. The decision is based on the current observation O_t (the 3D skin-pixel clusters) and the hypothesis for the preceding frame s_{t-1}. With each new frame, all combinations of the clusters' centroids are evaluated to find the hypothesis s_t that maximizes the product of the following 3 scores:

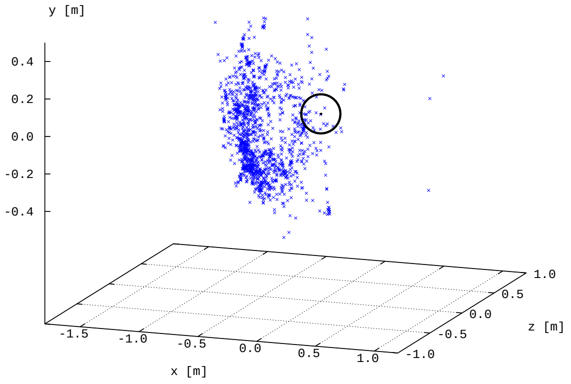


Figure 2: Observed positions of the right hand relative to the head (depicted by a circle) over a time of 2 minutes.

- The *observation score* $P(O_t|s_t)$ is a measure for the extent to which s_t matches the observation O_t. The calculation of this score is based on the weighted sum of the skin-pixels that are inside a certain radius around the predicted head and hand positions. $P(O_t|s_t)$ increases with each skin-pixel that complies with the hypothesis.

- The *posture score* $P(s_t)$ is the prior probability of the posture. It is high if the posture represented by s_t is a frequently occurring posture of a human body. It is equal to zero if s_t represents a posture that breaks anatomical constraints. To be able to calculate $P(s_t)$, a model of the human body was built from training data. The model consists of the average height of the head above the floor, a probability distribution of hand-positions relative to the head (see Fig. 2), as well as a series of constraints like the maximum distance between head and hand.

- The *transition score* $P(s_t|s_{t-1})$ is a measure for the probability of s_t being the successor of s_{t-1}. It is higher, the closer the positions of head and hands in s_t are to their positions in s_{t-1}.

2.2 Results

Our experiments indicate that by using the method described, it is possible to track a person robustly, even when the camera is not fixed and when the background is cluttered. The tracking of the hands is affected by occasional dropouts and misclassification. Reasons for this can be temporary occlusions of a hand, a high variance in the visual appearance of hands and the high speed with which people move their hands. Due to the automatic updates of the skin-color model, the system does not require manual initialization.

3. RECOGNIZING POINTING GESTURES

In this paper, we define a pointing gesture as a movement of one arm towards a pointing target. Hand posture or finger positions are not considered. When looking at a person performing pointing gestures, one can identify three different phases in the movement of the pointing hand:

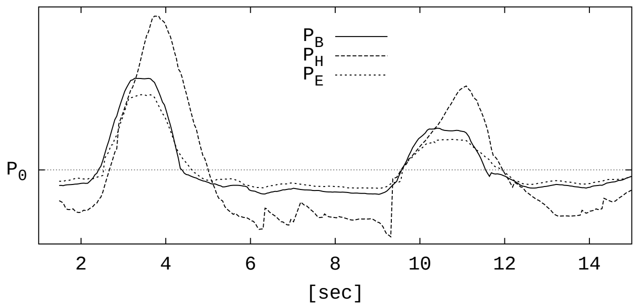


Figure 3: Log-probabilities of the phase-models during a sequence of two pointing gestures

- Begin (B): The hand moves from an arbitrary starting position towards the pointing target.

- Hold (H): The hand remains motionless at the pointing position.

- End (E): The hand moves away from the pointing position.

We examined pointing gestures performed by different persons, and measured the length of the separate phases. The average length of a pointing gesture was 1.8sec. Among the three phases, the hold phase shows the highest duration variance (from 0.1sec up to 2.5sec).

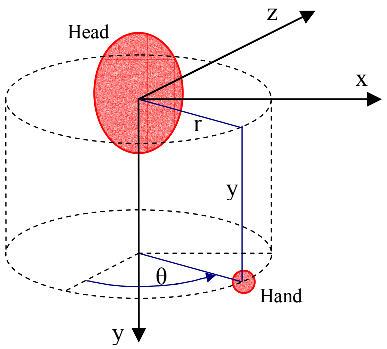
For estimating the pointing direction, it is crucial to detect the hold phase precisely. Therefore, we model the three phases separately: Three dedicated HMMs (M_B, M_H, M_E) were trained exclusively on data belonging to their phase. We choose the same HMM topology (3 states, left-right) for each of the three models. For each state, a mixture of 2 Gaussian densities represents the output probability. To get a reference value for the output of the phase models, we train a *null model* M_0 on short feature sequences ($0.5sec$) which do *not* belong to a pointing gesture. For M_0, we choose an ergodic HMM with 3 states and 2 gaussians per state. The models were trained with hand-labeled BHE-phases using the Baum-Welch reestimation equations (see [14]).

3.1 Classification

As we want to detect pointing gestures on-line, we have to analyze the observation sequence each time a new frame has been processed. The length of the BHE-phases varies strongly from one gesture to another. Therefore, we classify not only one, but a series of subsequences $s_{1..n}$, each one starting at a different frame in the past and ending with the current frame t_0 (see also [7]). The lengths of the sequences are chosen to be within the minimum/maximum length of a pointing gesture. For each of the phase models, we search for the subsequence $\hat{s}_{B,H,E}$ that maximizes the probability of being produced by the respective model. As $P(\hat{s}|M_0)$ represents the probability, that $\hat{s}$ is *not* part of a pointing gesture, we use it to normalize the phase-models output probabilities[1]:

$$\begin{aligned} \hat{s}_{B,H,E} &= argmax\ logP(s_{1..n}|M_{B,H,E}) \qquad (1) \\ P_{B,H,E} &= logP(\hat{s}_{B,H,E}|M_{B,H,E}) - logP(\hat{s}_{B,H,E}|M_0) \end{aligned}$$

[1] Note that in order to avoid numerical underflow, we use log probabilities rather than probabilities.



$$r = \sqrt{x^2 + z^2}$$
$$\theta = \arctan\left(\frac{x}{z}\right)$$
$$y = y$$

Figure 4: The hand position is transformed into a head-centered cylindrical coordinate system.

In order to detect a pointing gesture, we have to search for three subsequent time intervals that have high output probabilities P_B, P_H and P_E. Ideally, the respective model would significantly dominate the other two models in its interval. But as Fig. 3 shows, M_H tends to dominate the other models in the course of a gesture. That is why we detect a pointing gesture whenever we find three points in time, $t_B < t_H < t_E$, so that

$$\begin{aligned} P_B(t_B), P_H(t_H), P_E(t_E) &> 0 \\ P_E(t_E) &> P_B(t_E) \\ P_B(t_B) &> P_E(t_B) \end{aligned} \qquad (2)$$

3.2 Features

The raw input features for the gesture models are the 3D cartesian coordinates of the hands, as provided by the tracking module. We evaluated different coordinate system transformations of the feature vector. In our experiments it turned out that cylindrical coordinates of the hand (see Fig. 4) produce the best results for the pointing task. The origin of the coordinate system is set to the center of the head, in order to achieve invariance to the person's location.

The radius r represents the distance between hand and body, which is an important feature for pointing gesture detection. Unlike its counterpart in spherical coordinates, r is independent of the hand's height y.

Since we want to prevent the model from adapting to absolute hand positions – as these are determined by the specific pointing targets within the training set – we use the *deltas* (velocities) of θ and y instead of their absolute values The final feature vector is $(r, \Delta\theta, \Delta y)$.

Acceleration features have not been evaluated, since our 3D-trajectories turned out to be too noisy to generate meaningful second derivatives. For a comparison of different feature vector transformations for gesture recognition see [15].

3.3 Experiments and Results

In order to evaluate the performance of our system, we prepared an indoor test scenario with 8 different pointing targets. Ten test persons were asked to imagine the camera was a household robot. They were to move around within the camera's field of view, every now and then showing the camera (the "robot") one of the marked objects by pointing on it. In total, we captured 206 pointing gestures within a period of 24 min.

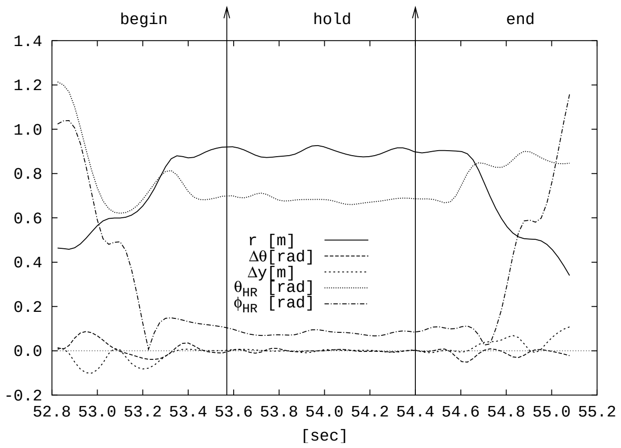


Figure 5: The feature sequence of a typical pointing gesture.

Two measures were used to determine the quality of the gesture detection:

- the detection rate (*recall*) is the percentage of pointing gestures detected correctly,
- the *precision* of the gesture detection is defined as the ratio of the number of correctly detected gestures to the total number of detected gestures (including false positives).

We evaluated by means of the *leave-one-out* method to make sure that the models were evaluated on sequences that were not used for training. In this experiment, the estimation of the pointing direction is based on the line of sight between head and hand (see section 5 for details on pointing direction estimation). Table 1 summarizes the results.

While the detection rate is similar in both cases (88%), the person-dependent test set has a lower number of false positives compared to the person-independent test set, resulting in a higher precision.

In addition, the estimation of the pointing direction is more accurate in the person-dependent case, so that 97% of the targets were identified correctly. This indicates that it is easier to locate the H-phase correctly when the models are trained individually for each subject. However, even in the person-independent case, 90% of the targets were identified correctly.

	person dependent	person independent
Detection rate (recall)	88%	88%
Precision	89%	75%
Avg. error angle	13°	21°
Targets identified	97%	90%

Table 1: Evaluation of the quality of pointing gesture recognition. The person-independent results are the average results on ten subjects. For the person-dependent case, average results on three subjects are given (see text for details).

4. USING HEAD ORIENTATION FOR POINTING GESTURE RECOGNITION

In our recorded data, we noticed that people tend to look at pointing targets in the begin- and in the hold-phase of a gesture. This behavior is likely due to the fact that the subjects needed to (visually) find the objects at which they wanted to point. Also, it has been argued before that people generally tend to look at the objects or devices with which they interact (see for example the recent studies in [1] and [2]).

To analyze whether this behavior could be used to improve the performance of pointing gesture recognition, we measured head rotation (θ_{Head}, ϕ_{Head}) by means of a magnetic sensor[2]. We calculate the the following two features:

$$\begin{aligned} \theta_{HR} &= |\theta_{Head} - \theta_{Hand}| \quad (3) \\ \phi_{HR} &= |\phi_{Head} - \phi_{Hand}| \end{aligned}$$

θ_{HR} and ϕ_{HR} are defined as the absolute difference between the (magnetically measured) head's azimuth/elevation angle and the (visually extracted) hand's azimuth/elevation angle in a spherical head-centered coordinate system. The features were chosen so as to prevent any adaption of the HMMs to the target positions in our test scenario.

Fig. 5 shows a plot of all features values during the course of a typical pointing gesture. As can seen in the plot, the values of the new features θ_{HR} and ϕ_{HR} decrease in the begin-phase and increase in the end-phase. In the hold-phase, both values are low, which indicates that the hand is "in line" with head orientation.

4.1 Results

In a new test scenario (similar to the one in section 3.3), we captured both video and head orientation. Four test persons performed a total number of 118 gestures. While the magnetic sensor was attached to their head, they were still able to walk around in the camera's field of view.

We comparatively trained the models a) using only the hand features ($r, \Delta\theta, \Delta y$), and b) using the head orientation features in addition to the hand features: (r, $\Delta\theta$, Δy, θ_{HR}, ϕ_{HR}). Again, we evaluated the quality of gesture recognition with the leave-one-out method. Table 2 shows the average results on four subjects.

As we can see, the performance improved significantly, when head orientation was included to the feature vector: The recall value increased from 78% to 87% while precision improved from 83% to 86%. Moreover, the error in determining the pointing direction was reduced from 37° to 28°, resulting in a higher percentage of correctly identified targets (1 out of 8).

	without head-orientation	with head-orientation
Detection rate (recall)	78%	87%
Precision	83%	86%
Avg. error angle	37°	28°
Targets identified	65%	83%

Table 2: Evaluation of pointing gesture recognition with head orientation as an additional feature.

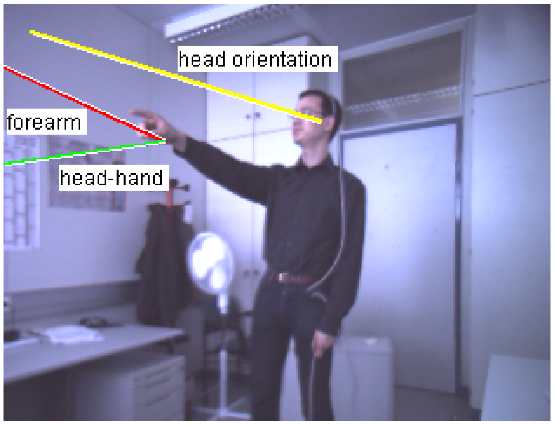


Figure 6: Different approaches for estimating the pointing direction. (The lines were extracted in 3D and projected back to the camera image.)

In both cases the head-hand line during the detected hold-phases was used for pointing direction estimation. We can therefore conclude, that the higher accuracy is the result of the increased ability of the HMMs to locate the gesture's hold-phase precisely.

5. ESTIMATING THE POINTING DIRECTION

We explored three different approaches (see Fig. 6) to estimate the direction of a pointing gesture: 1) the line of sight between head and hand, 2) the orientation of the forearm, and 3) the head orientation. While the head and hand positions as well as the forearm orientation were extracted from stereo-images, the head orientation was measured by means of a magnetic sensor.

5.1 Estimating the Forearm Orientation

In order to identify the orientation of the forearm, we calculate the covariance matrix C of the 3D-pixels that lie within a 20cm radius around the center of the hand. The eigenvector e_1 with the largest eigenvalue (the first principal component) of C denotes the direction of the largest variance of the data set. As the forearm is an elongated object, we expect e_1 to be a measure for the direction of the forearm (see Figure 7).

This approach assumes that no other objects are present within the critical radius around the hand, as those would influence the shape of the point set. We found that in the hold phase, the distance between hand and body and between hand and target object is mostly high enough for not influencing the measurement. Nevertheless, we reject the forearm measurement, when the ratio e_1/e_2 of the first and the second principal component is < 1.5.

5.2 Results

In order to evaluate the accuracy of the pointing direction estimation, we used the test-set described in section 4. The estimate of the pointing direction is based on the mean value of the measurements within the hold-phase of the respective gesture.

Because the gesture phases were manually labeled, this evaluation is not influenced by the gesture detection module,

[2] Flock of Birds Tracker, Ascension Technology Corporation.

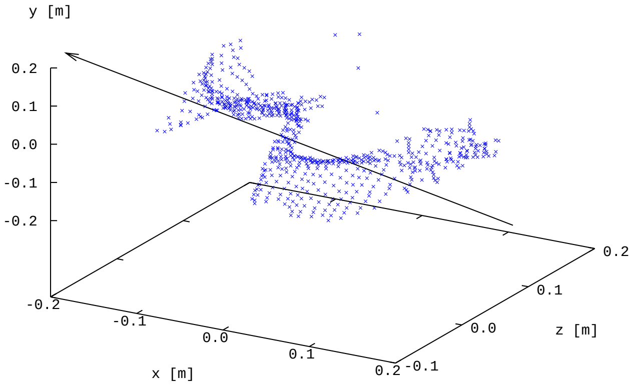


Figure 7: A principal component analysis of the 3D pixels around the center of the hand reveals the orientation of the forearm (arrow).

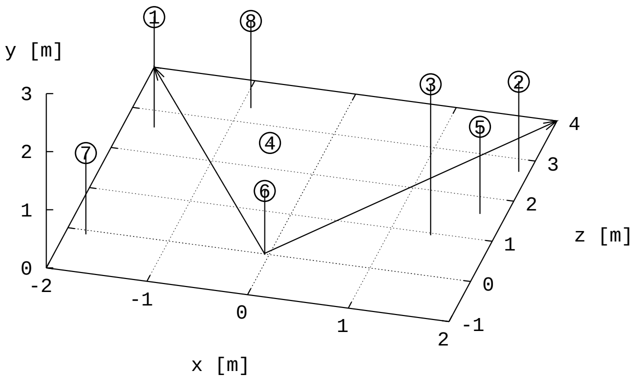


Figure 8: Target positions in the test set. Target #6 is equal to the camera position. The arrows indicate the camera's field of view.

which sometimes may fail to locate the hold-phase precisely. Nevertheless, there is an error induced by the stereo vision system as the camera's coordinates do not comply perfectly with the manual measurements of the target positions (see Figure 8).

Three measures are used to compare the different approaches: a) the average angle between the extracted pointing line and the ideal line to the target, b) the percentage of gestures for which the correct target (1 out of 8) was identified, and c) the availability of measurements during the hold-phase. The results (see table 3) are the average results on four subjects.

The good result of the head-hand line (25° error, 90% target identification) indicates that most people in our test set intuitively relied on the head-hand line when pointing on a target. For 98% of the gestures, head and hand positions were available during the hold-phase, so that the pointing direction could be estimated.

By means of the forearm line, 73% of the targets were identified. The test persons were pointing with an outstretched arm almost every time, thus reducing the potential benefit even of a more accurate forearm measurement[3]. Note that forearm measurements were available only for 78% of the gestures.

As head orientation was measured by means of an attached sensor, the results cannot be compared directly with the head-hand resp. forearm line, which were extracted from video. Nevertheless, the results indicate, that pure head orientation is inherently a good estimate for pointing direction estimation.

	Head-hand line	Forearm line	Head orientation
Avg. error angle	25°	39°	22°
Targets identified	90%	73%	75%
Availability	98%	78%	(100%)

Table 3: Comparison of three different approaches for pointing direction estimation.

6. CONCLUSION

We have presented a 3D vision system which is able to track a person's head and hands robustly, detect pointing gestures, and to estimate the pointing direction. The system was designed to function in natural environments, to operate in real-time[4], and to be person- and target-independent.

By using dedicated Hidden Markov Models for different gesture phases, high detection rates were achieved even on defective trajectories. In a person-independent evaluation with 10 test persons, our system achieved a gesture detection rate (recall) of 88% and a precision of 75%. For 90% of the gestures, the correct pointing target (one out of eight targets) could be identified. When training the HMMs on individual subjects (person-dependent), we noticed a significant performance gain.

In a second experiment, we measured head orientation by means of a magnetic sensor and used the absolute difference between head orientation and hand position as additional features. We found, that by using these additional feature derived from the user's head orientation, the precision and recall values of the gesture recognition as well as the accuracy of the pointing direction estimation increased significantly.

For estimating the pointing direction, we compared three different approaches: the line of sight between head and hands, the forearm orientation and the head orientation. Here, the head-hand-line turned out to deliver the most reliable estimate for pointing direction (90% correctly identified targets).

Since our results strongly indicate the usefulness of head orientation features for pointing gesture and target detection, we are now working on integrating purely vision-based estimation of a user's head orientation in our system.

Acknowledgements

Part of this work was carried out within the European Union funded project FAME (No. IST-2000-28323) and within the Sonderforschungsbereich 588 "Humanoid Robots" funded by the German Research Foundation (DFG).

[3] Unlike the relatively stable head position, the forearm measurements vary strongly during the H-phase.

[4] The system runs at 10 frames per second on a 2.8GHz Pentium 4 PC.

7. REFERENCES

[1] P.P. Maglio, T. Matlock, C.S. Campbel, S. Zhai, and B.A. Smith. Gaze and speech in attentive user interfaces. *Proceedings of the International Conference on Multimodal Interfaces*, 2000.

[2] B. Brumitt, J. Krumm, B. Meyers, and S. Shafer. Let There Be Light: Comparing Interfaces for Homes of the Future. *IEEE Personal Communications*, August 2000.

[3] C. Wren, A. Azarbayejani, T. Darrell, and A. Pentland. Pfinder: Real-Time Tracking of the Human Body. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, Vol. 19, No. 7, 1997.

[4] A. Azarbayejani and A. Pentland. Real-time self-calibrating stereo person tracking using 3-D shape estimation from blob features. *Proceedings of 13th ICPR*, 1996.

[5] T. Darrell, G. Gordon, M. Harville, and J. Woodfill. Integrated person tracking using stereo, color, and pattern detection. *IEEE Conference on Computer Vision and Pattern Recognition*, Santa Barbara, CA, 1998.

[6] T. Starner and A. Pentland. Visual Recognition of American Sign Language Using Hidden Markov Models. M.I.T. Media Laboratory, Perceptual Computing Section, Cambridge MA, USA, 1994.

[7] D.A. Becker. Sensei: A Real-Time Recognition, Feedback and Training System for T'ai Chi Gestures. M.I.T. Media Lab Perceptual Computing Group Technical Report No. 426, 1997.

[8] A.D. Wilson and A.F. Bobick. Recognition and Interpretation of Parametric Gesture. *Intl. Conference on Computer Vision ICCV*, 329-336, 1998.

[9] I. Poddar, Y. Sethi, E. Ozyildiz, and R. Sharma. Toward Natural Gesture/Speech HCI: A Case Study of Weather Narration. *Proc. Workshop on Perceptual User Interfaces (PUI98)*, San Francisco, USA. 1998.

[10] R. Kahn, M. Swain, P. Prokopowicz, and R. Firby. Gesture recognition using the Perseus architecture. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 734–741, 1996.

[11] N. Jojic, B. Brumitt, B. Meyers, S. Harris, and T. Huang. Detection and Estimation of Pointing Gestures in Dense Disparity Maps. *IEEE International Conference on Automatic Face and Gesture Recognition*, Grenoble, France, 2000.

[12] K. Konolige. Small Vision Systems: Hardware and Implementation. *Eighth International Symposium on Robotics Research*, Hayama, Japan, 1997.

[13] J. Yang, W. Lu, and A. Waibel. Skin-color modeling and adaption. Technical Report of School of Computer Science, CMU, CMU-CS-97-146, 1997.

[14] L.R. Rabiner. A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. *Proc. IEEE*, 77 (2), 257–286, 1989.

[15] L. W. Campbell, D.A. Becker, A. Azarbayejani, A.F. Bobick, and A. Pentland. Invariant features for 3-D gesture recognition. *Second International Workshop on Face and Gesture Recognition*, Killington VT, 1996.