A Two-Stage Bayesian Network Approach for Robust Hand Trajectory Classification

Frank Dittrich, Stephan Puls and Heinz Woern

Institute for Process Control and Robotics (IPR)
Karlsruher Institute of Technology (KIT)
76133 Karlsruhe, Germany
{frank.dittrich, stephan.puls, woern}@kit.edu

***Abstract*—In this paper, a hand gesture based approach is presented for the unidirectional information transfer from the human worker to the machine worker, designed for industrial applications. In order to allow high variance in respect of the location and orientation of the user relative to the RGB-D sensor, no detailed information about the finger setup is processed for the classification. For the recognition of various gestures, 3D hand trajectories are extracted and injected into a two-stage probabilistic inference mechanism. In our approach, we use Hidden Markov Models to optimize the classification rate for intended gestures, and a Bayesian Network for the robust elimination of unintended gestures. We show how empirical and heuristic information about the stochastic process are combined and used for the probabilistic modeling, in order to improve the overall robustness of the classification. The evaluation of our approached showed that the two-stage scheme allows for a directed rejection strategy on the basis of heuristic driven restrictions, without the loss of classification performance.**

I. Introduction

The here described approach for unidirectional human robot communication is applied in research scenarios related to safe human robot cooperation in the industrial domain. Here, a modular cognitive system provides information about activities and work processes in the workspace, based on the analysis of RGB-D sequences from a multi-sensor-setup. For the cooperation process, certain requirements are imposed, in order to enable the design of processes where the human worker can interact with the robot in a natural and intuitive way. Mainly, these requirements are the shared workspace of human and robot, with no spatial or temporal separation, and the use of the human body for interaction and communication.

For the first part, certain measures must be taken in order to ensure the safety of the human worker during the cooperation process. Details about these measures and the overall Safe Human Robot Collaboration (SHRC) system can be found in Section IV.

The here presented approach deals with the unidirectional communication as part of the natural and intuitive human robot interaction, using the human body as the input device.

A. Basic Concept

For the before mentioned cooperation scenarios, two major interaction classes were determined, namely the command gestures for the input of commands and states, and the movement gestures for the positioning of the robot. In order to communicate commands and states to the robot, various, sometimes numerous gestures have to be learned not only by the SHRC system, but also by the user. Therefore it is important, that they are easy to describe and easy to remember. Also for the robustness of the gesture classification, these body gestures have to be discriminative amongst each other. Considering that this approach is applied in an industrial context with rather large workspaces, it is also required that the gestures can be executed with one hand or arm, and with high variability in matters of position and orientation.

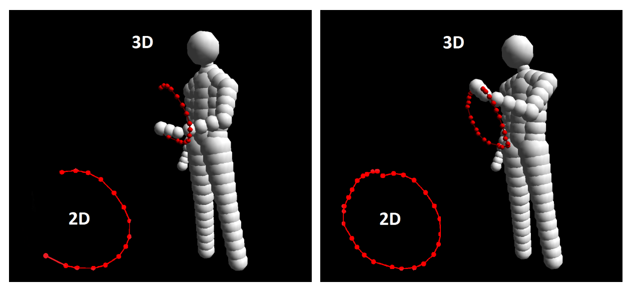


Fig. 1: A circular trajectory of the left hand in the 3D space, describes a circle on the 2D plane of the upper body.

To meet the demands, we use the concept of a virtual drawing board. Here, the user has to imagine a plane right in front of him and parallel to the upper body, where he can draw various signs with one hand. The hand is hereby represented by one point in the 3D space, allowing for a high variability in respect to the position and orientation of the user relative to the RGB-D sensor, since no additional information about the finger setup is needed. For the classification of various symbols, the 2D hand trajectory is examined (Fig. 1). Therefore, the gestures can be described to the user via simple 2D drawings (Fig. 2), which are easy to memorize because of the inherent ability and training of humans to use 2D symbols for handwriting.

For the robust gesture recognition, two types of probabilistic models are used to infer from data in the 2D plane and 3D space, in a two-stage decision process. In the first step the probabilistic evaluation of the 2D trajectories is conducted by Hidden Markov Models (HMMs), which showed good

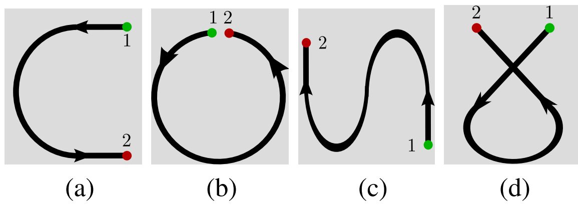


Fig. 2: Visualization of distinct hand and respectively command gestures.

results here and in earlier applications of on-line handwriting recognition. In the second step, a Bayesian network is deployed for the robust classification of intentionally executed gestures by the user, and the elimination of false detections of unintended gestures. For modeling and training, both empirical and heuristic information are used.

B. Related Work

Gestures are part of the nonverbal communication between humans. Figurative motions of certain body parts are used to clarify assertions or can be used as main form of communication [7]. Consequently, gestures are actions which have assigned denotation.

The recognition of spontaneously occuring gestures was already investigated by Eickerler and Rigoll in 1998 [3]. Differences of images of a grey-scale camera were analyzed and changing regions were identified. Thus, the gesturing person had to stand still while only moving gesture related body parts.

In [2], skin color detection is used to extract hand motions. Based on HMMs, these motions are classified and gestures identified. The gestures are used to command a mobile robot. A similar approach is used in [4], which detects one hand only for recognition of *air-written* alpha-numerical signs. The recognition of hand signs is investigated in [1]. In that context, the relationship between face and hands is important. Due to the fact that sign language is similarly expressive as spoken languages resultung ambiguities and complexity hinder the simple and intuitive interaction between humans and machines. Also, these approaches have in common, that the interaction space in which human movement is allowed is very restrictive, usually attributed to the used camera setups.

Progress in the entertainment electronics has recently yielded in widely available 3D sensing camera systems, e.g. Microsoft Kinect or ASUS Xtion. These sensors allow for direct interaction with computer systems without the need for additional auxiliary means. In [8], a method is proposed which maps human kinematics onto a humanoid robot system using a Kinect sensor. Consequently, the human motions can be captured and reproduced directly through the robot. This method is mainly applicable due to strong similarities of human and robot kinematics. For general industrial robots this method cannot be applied. Furthermore, such direct mapping does not contain information for guiding (partly) automated processes.

II. Robust Recognition of Arm Gestures

As shown in the schematic layout of the recognition system (Fig. 3), the gesture classification is based on the estimate sequence $\mathbf{S}_{1:t}$ of a full body tracking approach. In our SHRC system we use the Random Decision Forest based approach from the Microsoft KINECT SDK. In order to reduce noise in the estimation results, an optical flow driven inference scheme is deployed similar to [6].

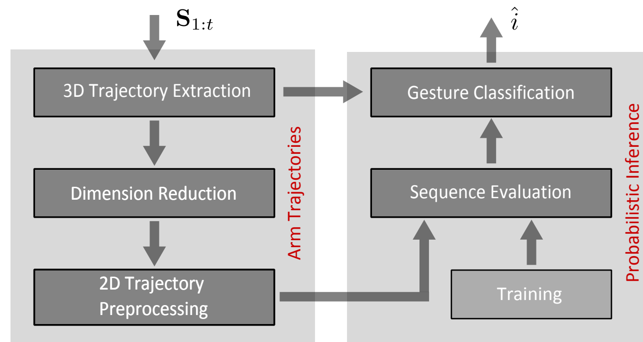


Fig. 3: Schematic layout of the gesture recognition system.

A. Arm Trajectories

1) 3D Trajectories Extraction: In order to extract trajectories from the hand movement, the user has to stop the motion of the hand at the beginning and the end of the gesture execution. In this way, the gesture trajectory can be extracted with an activity recognition approach. Even though measures are taken in order to remove noise from the tracking results, the estimates of the hand are still unstable and distributed around the true hand position, when the whole arm is not moving. Therefore we assume activity when a certain distance d_{min} is covered in the last Δ sampling steps, not by the sum over all h trajectory segments, but by the position difference of the hand positions at time steps t and $t - \Delta$:

$$\mathrm{d}(\mathbf{h}_t, \mathbf{h}_{t-\Delta}) \geq d_{min} \ . \tag{1}$$

Here, the vector $\mathbf{h}_t$ depicts the hand position at time step t, and the function $\mathrm{d}(\mathbf{h}_t, \mathbf{h}_{t-\Delta})$ is the distance metric induced by the l^2-norm.

The modular SHRC system allows for different configurations, where real-time Optical Flow (OF) estimates can be available. With OF estimates present, the activity recognition is augmented by the following activity measure:

$$\begin{aligned} l &= K_\omega * ||\mathbf{u}||_2 \ , \\ l(\mathbf{x}_{h_t}) &\geq f_{min} \ . \end{aligned} \tag{2}$$

The function l results from the convolution of the optical flow magnitude field $||\mathbf{u}||_2 = ||(u_x, u_y)||_2$ with an isotropic Gaussian with standard deviation ω. The 2D point $\mathbf{x}_{h_t}$ depicts the back-projection of the 3D hand position into the OF estimate at time step t. The scalar f_{min} describes the lower boundary of the local OF, for the activity assumption.

The before defined activity measures are now used for the extraction of 3D trajectories by recognizing and extracting activity sequences.

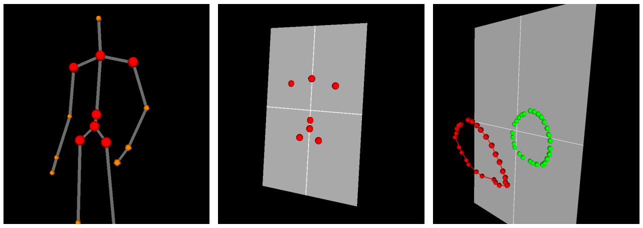


Fig. 4: *Left:* Resulting skeleton information from the full body tracking. The red joints are assigned to the upper body region. *Center:* Estimation result of the upper body plane and the body axes. *Right:* 2D Projection result (green) of a 3D circle trajectory (red).

2) Dimension Reduction: As described in the introduction (Section I-A), the probabilistic evaluation of the trajectories is done partially in 2D. Following the concept of the virtual drawing board, the extracted 3D trajectories have to be projected onto the upper body plane. One appropriate method for this task is the Principal Component Analysis (PCA) [9]. In the first step, the upper body plane is estimated:

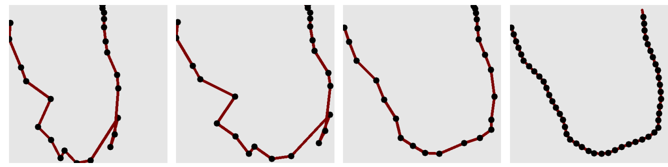
$$\begin{aligned}\tilde{\boldsymbol{\mu}} &= E\{\mathbf{P}|S_u\} \ , \\ \tilde{\boldsymbol{\Sigma}} &= \mathrm{Cov}\{\mathbf{P}|S_u\} = \mathbf{E}\boldsymbol{\Lambda}\mathbf{E}^T \ , \\ &= (\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3) \begin{pmatrix} \kappa_1 & & 0 \\ & \kappa_2 & \\ 0 & & \kappa_3 \end{pmatrix} \begin{pmatrix} \mathbf{e}_1^T \\ \mathbf{e}_2^T \\ \mathbf{e}_3^T \end{pmatrix}, \kappa_1 \geq \kappa_2 \geq \kappa_3 \ .\end{aligned} \tag{3}$$

Here the random variable $\mathbf{P}$ describes the upper body point positions conditioned on $S_u = \{p_{u1_{t-h:t}}, p_{u2_{t-h:t}}, \ldots\}$ which is a set of upper body point samples (Fig. 4) over a certain time frame h. The plane is defined by the expected value $\tilde{\boldsymbol{\mu}}$, describing the translation, and the two eigenvectors $(\mathbf{e}_1, \mathbf{e}_2)$ of the covariance matrix $\tilde{\boldsymbol{\Sigma}}$, describing the rotation. The eigenvectors and the corresponding eigenvalues are contained in the matrices $\mathbf{E}$ and $\boldsymbol{\Lambda}$, which are the result of the eigen-decomposition of $\tilde{\boldsymbol{\Sigma}}$. The procedure described in Eq. 3 is the PCA of the upper body points in a certain time frame. Because of the nature of the PCA and the choice of the body samples, the eigenvector $\mathbf{e}_1$ is aligned with the vertical body axis, and the eigenvector $\mathbf{e}_2$ is aligned with the horizontal body axis (Fig. 4). The estimation of the plane is conducted in every time step and during the gesture execution, therefore smoothness of the result is important because of the temporal context of the trajectory points. Here, the examination of the sample point history promotes the smoothness of the estimates.

Based on the estimation result, the points of the 3D trajectory $\{\mathbf{t}_{3D_i}\}$ are projected on the body plane using $\tilde{\boldsymbol{\mu}}$ and $(\mathbf{e}_1, \mathbf{e}_2)$:

$$\mathbf{t}_{2D_i} = \begin{pmatrix} \mathbf{e}_1^T \\ \mathbf{e}_2^T \end{pmatrix} (\mathbf{t}_{3D_i} - \tilde{\boldsymbol{\mu}}) \ . \tag{4}$$

The resulting 2D trajectory $\{\mathbf{t}_{2D_i}\}$ (Fig. 4) is a 2D point series, where each point is depicted by the coordinates relating to the body axes. It should be mentioned, that because the features description of the 2D trajectory sample points contains absolute angle information, it is important that in addition to the body plane approximation, the body axis are estimated too. This ensures that trajectory coordinates are comparable.



(a) Original (b) Resized (c) Smoothed (d) Resampled

Fig. 5: *From left to right:* Successive processing steps of the 2D trajectory.

3) 2D Trajectory Preprocessing: Because of noise in the hand tracking and the movement of the user while generating the 3D trajectory, the resulting 2D trajectory is flawed (Fig. 5(a)) and has to be preprocessed. The single processing steps are partially similar to the applied methods in [5].

The first step is the size normalization of the trajectory, which is necessary because the features for the HMM observation sequence contain information about the position increment $(\Delta x, \Delta y)$. The ratio of maximum expansion in respect to both axes is hereby not preserved:

$$\bar{p}_x = \frac{p_x - p_{x_{min}}}{p_{x_{max}} - p_{x_{min}}} S_x, \quad \bar{p}_y = \frac{p_y - p_{y_{min}}}{p_{y_{max}} - p_{y_{min}}} S_y \ . \tag{5}$$

Here, the constants S_x and S_y describe the target size, and p_{min} and p_{max} are the position extrema.

The smoothing step removes discontinuous and rough parts with a moving average filter (Fig. 5(c)).

$$\hat{\mathbf{p}}_t = \frac{\bar{\mathbf{p}}_{t-1} + \bar{\mathbf{p}}_t + \bar{\mathbf{p}}_{t+1}}{3} \ . \tag{6}$$

In the resampling step, the sample point count per trajectory is set to a constant value N_S, where all points are equidistant (Fig. 5(d)). This step is very important for the probabilistic evaluation of the trajectory by the HMMs, because the more sample points an observation sequence has, the less likely it becomes. Therefore, comparing sequences with different sizes is not practical.

B. Probabilistic Inference

The results from the previous section are used for the probabilistic inference. This comprises two types of graphical models, separate HMMs λ_i for every gesture $i \in \{1, \ldots, N_g\}$ and one star-shaped Bayesian network for the concluding robust gesture classification.

1) Sequence Evaluation: In our approach, HMMs are used for the stochastic evaluation of 2D trajectories. Here the assumption is, that the 2D trajectory is an observation sequence from a hidden stochastic process with markov properties. The

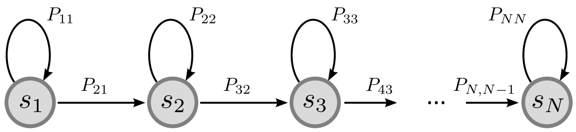


Fig. 6: Restrictive *Left-To-Right* model type for all HMMs, with $P_{ji} = P(s_j|s_i)$

observation itself is a stochastic process conditioned on the hidden states. The HMM can be used to model the stochastic behavior of gesture trajectories: the state transition probability distribution $P_X(s_i|s_j)$ can be used to model the temporal behavior of the sequence, and the emission probability distribution $P_Z(o_i|s_j)$ assigns an uncertain observable behavior to the individual states. In order to reason about an unknown trajectories T, the probability $P(O(T)|\lambda_i)$ is calculated, which describes numerically how good this trajectory can be explained with the HMM λ_i [12].

We chose the same design parameters for all deployed HMMs, each one representing one gesture class. We use a maximum restrictive *left-to-right* model type as depicted in Figure 6. The initial state distribution is set to $P_\pi(s_1) = 1$.

For the description of the trajectory T we use the following features for every sample point $\mathbf{p}_t = (p_x, p_y)$:

$$\begin{aligned} \Delta x_t &= p_{x_{t-2}} - p_{x_{t+2}} \ , \\ \Delta y_t &= p_{y_{t-2}} - p_{y_{t+2}} \ , \\ \gamma &= \tan\left(\frac{\Delta y_t}{\Delta x_t}\right) \ . \end{aligned} \tag{7}$$

The absolute angle information requires, that symbols on the virtual drawing board are always entered in the same orientation, which is comparable to writing on a lined sheet of paper. The advantage of this approach is that symbols like the semi circle (Fig. 2 (a)) can be distinguished, when entered in different orientations. This allows for the description of different commands and states with the same basic symbol type. For instance the semi circle in 90 degree rotation steps can be used to describe commands like *left*, *up*, *right* and *down*.

After the extraction of the feature vectors, a quantization strategy is deployed using k-means clustering. The result is a scalar and discrete observation sequence $O(T)$, which can now be processed with the HMMs.

Prior to the application of the HMMs, all models have to be trained with sample trajectories, in order to approximate the stochastic behavior of the individual gesture classes. Recording sufficient training samples by executing the gesture numerous times is not practical. This is mainly due to the fact, that trajectories are extracted all the time, if they are gestures or not. So it is not easy to control exactly which trajectory should be saved when executing a gesture. Therefore we use 2D mouse trajectories which are entered by a simple software tool. With this method, producing sufficient training data is convenient and takes only a few minutes or even seconds. For the training of the HMM parameters, we use the Baum-Welch method, described in [12].

In order to reason about an unknown trajectory T, we determine the most likely state sequence S_{max}

$$S_{max} = \arg\max_S \underbrace{P(S|O(T), \lambda)}_{P^*} \tag{8}$$

using the Viterbi algorithm [12]. Both S_{max} and P^*_{max} will be processed in the following classification step.

2) Gesture Classification: Most applications use solely the measure P^*_{max} for the gesture classification. In such a classification scheme, if a gesture is executed, the chances are very high that the correct gesture is detected. However, if no further information is injected into the decision process, the chances are also high, that gestures will be recognized when no gesture execution was intended. Therefore, we use the information from the sequence analysis in a subsequent processing step, where it is combined with further information and restrictions in a Bayesian manner, thereby optimizing the classification and increasing the robustness of the decision process against unintended gestures. Here the results of the empirical data driven inference is combined with heuristics which describe constraints based on expert knowledge.

a) 3D Trajectory Input Space: A highly discriminative constraint is the restriction that the input of gestures is only allowed in a certain region relative to the body. The deviation from that constraint is measured by the sum over the distance of all 3D trajectory points from the region, and will be described by the random variable H_1.

b) 3D Trajectory Length: Also the length of the 3D trajectory has to be in a certain range. The trajectory length will be described by the random variable H_2.

c) 3D Trajectory Geometry: In order to make the approach more robust against false classification, discriminative geometric features of the 3D trajectory are extracted:

$$\begin{aligned} \Delta_{SE} &= \mathrm{d}(\mathbf{p}_S, \mathbf{p}_E) \ , \\ \Delta_{CS} &= \mathrm{d}(\mathbf{p}_C, \mathbf{p}_S) \ , \\ \Delta_{CE} &= \mathrm{d}(\mathbf{p}_C, \mathbf{p}_E) \ . \end{aligned} \tag{9}$$

The distinct points $\mathbf{p}_S, \mathbf{p}_E$ and $\mathbf{p}_C$ are the first, last and center 3D trajectory points respectively. The function d is the distance metric induced by the l^2-norm. The geometric features are described by the random variables H_3, H_4 and H_5 respectively.

d) 3D Trajectory Distance Smoothness: In addition, a measure for the smoothness of the trajectory T in respect of the distance to the body plane is determined:

$$\sigma_D = \mathrm{Var}\{D|T\} \ . \tag{10}$$

The random variable D describes the distance of 3D trajectory points to the body plane. The variance σ_D conditioned on T is therefore a measure for how parallel the trajectory runs to the body plane. This measure will be described by the random variable H_6.

e) State Sequence Analysis: In order to make the stochastic modeling of the 2D trajectories by the HMMs less restrictive, transition and observation probabilities which were set to zero in the course of the model training, were assigned very small values afterwards. This measure highly increases the HMM's classification quality, but it also creates another problem. Trajectories which describe almost linear movements are presented to the HMM as a sequence of similar observation symbols. The measure P^*_{max} for those trajectory types can be very high. The examination of such state sequences S_{max} shows that the highest likely explanation for such observations are fast and costly transitions from state s_1 to a state where the observation symbols are highly likely, where it resides until the end of the sequence. Considering the factorization of the joint probability, this explanation is plausible. Because such state sequences are very homogenous, the entropy of S_{max} is an adequate measure:

$$H(X_S) = -\sum_{i=1}^{N} P(s_i)\log_2\left(P(s_i)\right) \ . \tag{11}$$

Here N is the number of states, X_S is the random variable which describes the state sequence and $P(s_i) = P(X_S = s_i)$ is the distribution of the states given the state sequence. Demanding a high negative value for H, combined with the model type (Fig. 6), enforces heterogenous state sequences and thereby observation sequences which follow the temporal modeling of the HMM. The entropy, dependent of the HMM λ_i, is described by the random variable E_i. The random variable S_{P_i} describes the probability measure $P^*_{max_i}$.

f) Bayesian Network for Classification: As depicted in Figure 7 the joint probability factorizes in the following way:

$$P(G, \mathbf{E}, \mathbf{S}_P, \mathbf{H}) = \prod_{k=1}^{M} P(E_k|G)P(S_{P_k}|G)\prod_{l=1}^{6} P(H_l|G)P(G) \tag{12}$$

Here, G is the variable which describes the gestures g_i, $\mathbf{E}$ and $\mathbf{S}_P$ are the vectors of the HMM dependent E_i, and S_{P_i}, $\mathbf{H}$ is the vector of the random variables H_i, and the value M is the number of HMMs or learned gestures.

The restrictive independence assumptions help to simplify the modeling step of the single distributions, using training or expert knowledge. For the classification we first search for the gesture g_i, which maximizes:

$$P(G = g_i|\mathbf{E}, \mathbf{S}_P, \mathbf{H}) = \frac{P(G, \mathbf{E}, \mathbf{S}_P, \mathbf{H})}{P(\mathbf{E}, \mathbf{S}_P, \mathbf{H})} \ . \tag{13}$$

Because $(\mathbf{E}, \mathbf{S}_P, \mathbf{H})$ is given as evidence, and the prior distribution $P(G)$ is assumed as uniformly distributed, the classification can be formulated as:

$$\hat{i} = \arg\max_i \prod_{k=1}^{M} P(E_k|g_i)P(S_{P_k}|g_i)\prod_{l=1}^{6} P(H_l|g_i) \ . \tag{14}$$

In the last step, if $P_{\hat{i}} = P(g_{\hat{i}}|\mathbf{E}, \mathbf{S}_P, \mathbf{H})$ exceeds a certain threshold, we infer that an intentionally executed gesture was detected, and $\hat{i}$ is the gesture index.

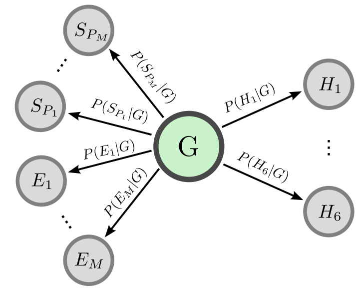


Fig. 7: Bayesian Network for the classification task.

III. Experiments

For the evaluation of our approach all HMMs were trained with 30 hidden state symbols and 40 observation symbols, using the method described in Section II-B1.
In the first experiment, the classification performance of the HMMs was analyzed.

	SC_L	SC_U	SC_R	SC_D	C_L	C_R	S_V
SC_L	1						
SC_U		0.98	0.02				
SC_R			1				
SC_D				1			
C_L					1		
C_R						1	
S_V							1

TABLE I: Confusion matrix of the HMM based classification, using 50 test samples per gesture class. SC_L, SC_U, SC_R, SC_D depict the semi-circle gestures oriented left, up, down and right, C_L and C_R depict the full circle with left and right direction of orientation, S_H depicts the snake shaped gesture in vertical orientation (cf. Fig. 2).

For training and evaluation of the HMMs we created 150 2D trajectory samples, using the mouse as input device (Section II-B1). The set was then divided into a training set with 100 samples, and a test set with 50 samples per class. Table I shows the confusion matrix of the HMM based classification task. The entries show clearly that the HMM based 2D classification, combined with the discriminative symbols and trajectory features, delivers highly reliable information about intended gestures.

For the experimental evaluation of the second step of the decision process, we recorded a full body tracking sequence of random executed trajectories by the user. Here it was the objective not to enter any gestures, but try to imitate conventional movements appropriate to the application in a work environment. The first line of Table II shows that the information of the HMM alone leads to many misclassifica-

	%	SC_L	SC_U	SC_R	SC_D	C_L	C_R	S_V
HMM	36.0	3	4	4	6	9	0	8
2-Stage	0.0	0	0	0	0	0	0	0

TABLE II: Detection of unintended gestures based on 100 extracted arm trajectories.

tion. Here a maximality examination combined with a global threshold was used for the inference. The second line shows no misclassification for the two-stage approach.

In order to conclude the overall robustness of our approach, we evaluated the classification quality of the two-stage approach, based on a sequence of 20 intended gesture per gesture class. The gestures were executed by the user, using his right hand. Table III shows that despite the restrictions modeled in

	SC_L	SC_U	SC_R	SC_D	C_L	C_R	S_V
2-Stage	0.95	1	1	1	1	1	0.90

TABLE III: Classification rate of our two-stage approach in percent, based on 20 intended gesture executions per class.

the second stage, the classification rate is comparable to the results presented in Table I. The performance of the two-stage approach in detection and classification of intended gestures, and the elimination of false gesture detections, shows an improved overall robustness.

IV. Overall System

The here proposed approach to human robot communication is embedded in a research system to enable safe, workspace sharing human robot cooperation. Thus, different modules are incorporated for sensing the robot's environment, analyzing and reasoning about sensor data, and planning the robot's reactions. Consequently, this resembles a cognitive cycle: *Perception*, *Cognition*, and *Action*.

The perception includes, besides full human body tracking, position sensing of the robot and state information of the gripper or handled workpiece.

The cognition consists of the presented approach to gesture recognition and spatio-temporal reasoning about situations [10]. The achieved situational awareness allows the distinction between different situation concepts. The human co-worker might be monitoring the production process, can be distracted by other co-workers or through his work load, or can be interacting with the robotic system. Performed gestures and actions have to be considered regarding this contextual information. Thus, the same gestures can either be directed towards the robot or might be addressed to an outside person. Consequently, the reasoning system concludes the robot's objective, which can be, e.g., to comply with gestures or proceed with a given task.

In order to achieve safety for the human co-worker the robotic motion is determined through an reactive, online path planning module [11]. Based on the full human body tracking it is possible to determine possibly impending collision during robot motion. If such a collision is predicted, the robot is stopped and a re-planning is invoked. During planning the risk for collision is determined and if it is too high the path re-planning will retry until the human co-worker clears the workspace.

V. Conclusion

We presented in this paper a robust unidirectional gesture approach, which is based on the trajectory of one hand only. Additional information about the finger setup is not needed for the robust gesture detection and classification.

The objective of our two-stage bayesian network approach was the robust classification of intended hand gestures, combined with an elimination strategy of unintended gesture executions. The here proposed heuristic driven restrictions showed in the evaluation, that their selection and stochastical modeling obtained the high classification quality of the HMM classifiers, while rejecting hand trajectories which did not represent a gesture execution of the user.

References

[1] S. Akyol. *Nicht-intrusive Erkennung isolierter Gesten und Gebrden*. Aachen, 2003.

[2] M. Ehrenmann, T. Ltticke, and R. Dillmann. Erkennung dynamischer Gesten zur Kommandierung mobiler Roboter. In *Autonome Mobile Systeme*, pages 20–26, 2000.

[3] S. Eicherler and G. Rigoll. Kontinuierliche Erkennung von spontan ausgefhrten Gesten mit neuen stochastischen Dekodierverfahren. In *Workshop Dynamische Perzeption*, 1998.

[4] M. Elmezain, A. Al-Hamadi, and B. Michaelis. Improving hand gesture recognition using 3d combined features. In *2nd International Conference on Machine Vision*, pages 128–132, 2009.

[5] G. Reddy et al. Assamese online handwritten digit recognition system using hidden markov models. DAR '12, pages 108–113, New York, NY, USA, 2012. ACM.

[6] J. Graf and F. Dittrich and H. Wörn. High Performance Optical Flow Serves Bayesian Filtering for Safe Human-Robot Cooperation. In *Joint 41th International Symposium on Robotics and 6th German Conference on Robotics*, pages 325–332, 2010.

[7] A. Kendon. *Gesture: Visible action as utterance.* Cambridge University Press, 2004.

[8] M. Liebhardt. Motion Retargeting. URL http://www.ros.org/wiki/reem_teleop.

[9] Karl Pearson. On lines and planes of closest fit to systems of points in space. *Phil. Mag.*, 6(2):559–572, 1901.

[10] Stephan Puls and Heinz Wörn. Combining HMM-Based Continuous Human Action Recognition and Spatio-Temporal Reasoning for Augmented Situation Awareness. In *IADIS*, pages 133–140, 2012.

[11] Stephan Puls, Patrick Betz, Max Wyden, and Heinz Wörn. Path Planning for Industrial Robots in Human-Robot Interaction. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), WS Robot Motion Planning: Online, Reactive and in Real-time*, 2012.

[12] L. Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286, 1989.