Proc. of the 17th IEEE Int. Symposium on Robot and Human Interactive Communication, (RO-MAN 08), Munich, Germany, pp. 532-538, IEEE Omnipress

Whom to talk to? Estimating user interest from movement trajectories

Steffen Müller, Sven Hellbach, Erik Schaffernicht, Antje Ober, Andrea Scheidig, and Horst-Michael Gross

Abstract— **Correctly identifying people who are interested in an interaction with a mobile robot is an essential task for a smart Human-Robot Interaction.**

In this paper an approach is presented for selecting suitable trajectory features in a task specific manner from a huge amount of different forms of possible representations. Different sub-sampling techniques are proposed to generate trajectory sequences from which features are extracted. The trajectory data was generated in real world experiments that include extensive user interviews to acquire information about user behaviors and intentions. Using those feature vectors in a classification method enables the robot to estimate the user's interaction interest.

For generating low-dimensional feature vectors, a common method, the Principle Component Analysis, is applied. The selection and combination of useful features out of a set of possible features is carried out by an information theoretic approach based on the Mutual Information and Joint Mutual Information with respect to the user's interaction interest. The introduced procedure is evaluated with neural classifiers, which are trained with the extracted features of the trajectories and the user behavior gained by observation as well as user interviewing. The results achieved indicate that an estimation of the user's interaction interest using trajectory information is feasible.

I. INTRODUCTION

A future application of interactive service robots, like the robot platforms HOROS [1] or SCITOS [2], will be to provide information and other services to people in public environments and office buildings. A main question is how the robot should attract the attention of its interaction partners. It would be inconvenient to let the robot address every detectable person in the surrounding area, or to let him wait motionless for people to approach the robot themselves.

Intelligent robots should be able to show a natural human-like behavior. For instance, they should meet halfway with someone who is on the verge of starting an interaction and attract people who are not sure about the robot's abilities. In addition, they should avoid disturbing people who are unwilling to interact with them.

In order to do so, a robot must have the ability to estimate on its own, who is willing to interact and who is not. This estimation has to take place early enough before the person is reaching the robot, to enable the system to accomplish a user-adaptive welcoming. Human beings infer the intention of other people from their mimics, gestures, and body language. As a result of limited camera resolution, a robot is not able to specify those at far distances. A promising idea, followed in this paper, is to extract the required information from a person's movement trajectory which is available as soon as the person is detected by the robot.

Section II gives a short overview on recent research in this particular field of Human-Robot Interaction. This is followed by a description of the scenario in section III. Afterwards a systematization on different methods for trajectory representation is presented in section IV. How to select the right representation method is discussed in section V. After the results are presented in section VI, the paper concludes with section VII.

This work has partially received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 216487. All authors are with Neuroinformatics and Cognitive Robotics Lab, Ilmenau University of Technology, 98684 Ilmenau, Germany steffen.mueller@tu-ilmenau.de

II. RELATED WORK

The estimation of a user's interest to interact with a robot with the common goal of adapting the robot's behavior in starting an interaction became focus of attention in different works. Most approaches consider only the distance and certain zones to infer a user's intention since one-to-one interactions are typically carried out in small distances. Finke et al. [3] trained HMMs with data from sonar sensors to recognize people that were closer than 1 m to the robot. It was assumed that people which entered that zone are interested in an interaction. Nabe et al. [4] discovered in a field trial with a ROBOVIE-M robot that three zones exist in which a person shows either a watching, a talking or a physical interacting behavior depending on their distances to the robot. They announced to regulate the robot's behavior just by these distances. Michalowski et al. [5] subdivided the surrounding area of the roboceptionist VALERIE into certain zones, also depending on the distance to the reception desk. These zones were used together with head pose information from a camera to determine the degree of the person's engagement in a Human-Robot Interaction in order to control the rule-based behavior of Valerie in contacting people. An approach with the main focus set on the selection of the robot's actions can be found in Schulte et al. [6]. The tour-guide robot MINERVA measured the probability distibution of people in the surrounding area as well as their distance to the robot via laser. This information is used in a memory-based reinforcement learning approach to learn the best action of the robot for attracting people and involving them in an interaction.

A known problem in the use of distance and zones as criteria is that these measures might be applicable in some scenarios but it seems insufficient when people are moving quickly in the robot's surroundings. Preferably, the classification takes place as early as possible to let the robot react fast enough. The incorporation of movement trajectories enables this course of action. A first step into this direction was made by Holzapfel et al. [7]. They trained a neural network with data from a stereo camera tracker. At each classification step, some angle and distance features were used as input. People heading towards the robot were classified as being interested in an interaction, and people passing the robot as being not interested.

Within the discussed approaches the classification problem is always seen from the robot's view by using a robot centered coordinate system. In comparison, our approach interpretes the scene from the person's point of view which brings some advantages. Moreover, often only raw trajectory data is considered as classification input so far. Applying specific preprocessing and automatic evaluation steps might lead to more suitable classification results which is proposed in this work.

III. SCENARIO AND DATA ACQUISITION

To realize the estimation of the users' intention to interact with a robot, it is necessary to capture movement trajectory data. A spatio-temporal classifier, in particular a standard feedforward Neural Network, is trained with this data. This section discusses the data aquisition scenario.

We used our experimental robot plattform HOROS for this purpose. The robot is shown in Fig. 2, while the details of the applied multimodal people tracking system have been introduced in [8].

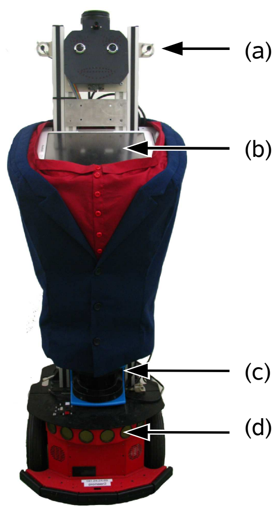

Fig. 2. The robot plattform HOROS. The marked devices are (a) microphones, (b) touch screen, (c) laser range finder, and (d) sonar sensors.

This approach is using different sensors of the robot to get a set of hypotheses of people's positions in the robot's surroundings. A local occupancy gridmap which is built up from the measurements of the robot's sonar sensors is compared to a static map of the environment, yielding hypotheses based on unexpected differences. Besides the hypotheses from this cue, a SICK laser range scanner is used to detect legs in greater distances. Using a difference approach, individual rays of the range scan are compared to a dynamic background model, which adapts slowly to variations e.g. due to slight movements of the robot. The disadvantage of the laser model, that people become part of the background if they don't move for a certain time, is compensated by the combination of the different sensors by means of probabilistic sensor fusion techniques. The system generates a list of trajectory points for people moving in a 6 m circle around the robot with an update frequency of 10 Hz.

For the purpose of recording meaningful trajectory data an appropriate location is needed. This comes with some strings attached. Besides having enough space to detect and track a person soon enough for accomplishing a robot action before the person leaves the robot's interactive operation area, the place should have a high frequency of single individuals passing by. One disadvantage of the used tracking system is that people who walk very close to each other are hard to separate.

However, two entrance halls at our university were chosen, which seemed appropriate for this task. The robot was placed in three different positions on 15 non-consecutive days for about 6 hours each in order to gain trajectories from different robot views. This allowed a position invariant estimation of the user's intention to interact.

Besides the choice of the place, different design decisions had to be made before the experiments could be carried out. First of all, a promising scenario for the target group, consisting of students, employees and visitors, had to be found. As the robot was supposed to be deployed for a longer period of time, we wanted to offer an application which lead to both, interactions with new people as well as repeated interactions with people passing the robot regularly in order to integrate the robot into their everyday life.

The resulting information terminal application provided the daily menus of different cafeterias, the ongoing cinema program, bus timetables, jokes, events and as a special feature, an audio-based gender estimation accomplished by the robot. An eye-catching GUI was created especially for the application which was presented on the robots touch display. It had a simple navigation structure to ensure that even people who are not familiar with using technical devices are able to interact with the robot intuitively. To let the robot differ sufficiently from a static information terminal, we focused on the integration of a broad spectrum of robot activities by varying the robots voice responses, changing its facial expressions and trying to produce a continuous dialogue process.

The robot behaved as follows during the experiments: Whenever a person was detected by HOROS, a randomly chosen voice output started which was intended to attract the user's attention and influence the user's decision to interact.

Each possible voice output had a different effect on the user. These actions had specifiable moods, they were meant to be funny, happy, or provocative. Several actions were tested to find the most suited to the people's expectations and wishes. After an action was initiated, the robot waited for the person to approach or to pass by. Via its sensors the robot took note of hesitating people, too. In that case, the robot started an additional animating voice output to convince the person to interact. During each trial hidden experimentators watched the scene and made notes about what happened. The experimentators waited until the current person had left the robot's surroundings. Afterwards they tracked the person and started an interview to gain new knowledge about Human-Robot Interactions besides just recording trajectory data.

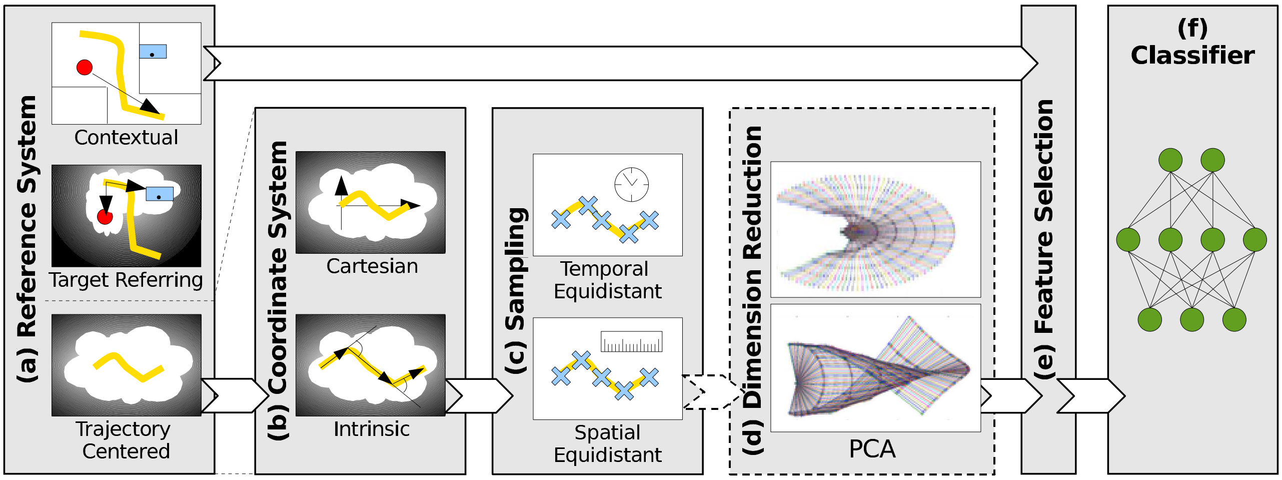

Fig. 1. Overview of different kinds of representation for trajectories. Choosing the right reference system (a), coordinate system (b) and sampling technique (c) is highly task dependent. Optionally, it is possible to apply dimension reduction (d). In (d) the influence for two of the Eigentrajectories on the recomposed trajectory are shown. Choosing informative representations via feature selection (e) and the neural network as final classifier (f) round out the system.

Especially the process of establishing dialogs with humans was our field of interest, concerning getting the attention of people, signalizing the robot's interest to interact and finally establishing the interaction. A bunch of questions arised in this context that we tried to answer by designing a questionnaire for user interviews. Details regarding the questions, the collected answers and conclusion are provided in section VI.

IV. TRAJECTORY REPRESENTATION

One important step after data acquisition using a tracker is to transform the trajectories into an appropriate representation. This helps to provide necessary information about the trajectory more explicitely, which will support the classifier. Similar as performing image pre-processing steps before classifying the image content, trajectory data needs to undergo an adequate transformation.

Being confronted with the fact that collecting real world data for training is time consuming, one has to deal with few training examples. Having only a small amount of data implicitely means that the number of input dimensions of the classifier is not supposed to exceed a certain amount.

The final classification algorithm, as well as all necessary preprocessing steps, have to work online on the robot. This means the method has to deal with the fact, that during the classification process only a part of the trajectory is known. Hence, the classifier is trained with trajectory fragments. Those fragments are extracted by sliding a window with a fixed size over the trajectory (highlighted end of the trajectory in Fig. 3(a)). For obtaining the right window size, methods from time series analysis could be used [9]. Since the system aims at online classification, a compromise has to be found. On one hand having an adequate length to provide enough information is crucial and on the other hand, being able to estimate the user's interest in a reasonable short time before passing by is important, too.

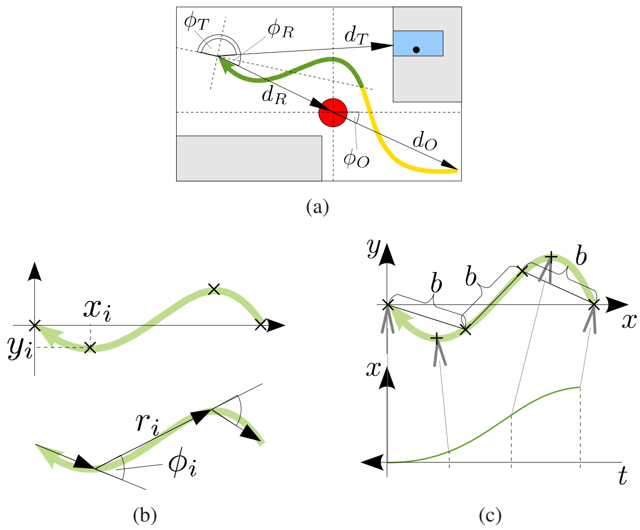

Fig. 3. (a) Reference systems: Contextual Information is provided by specifying the polar coordinates of the trajectory's origin (ϕ_O, d_O) in a stationary coordinate system originating at the robot's position (which is in this case facing to the right). The positions of the robot (ϕ_R, d_R) or other target objects (like a door) (ϕ_T, d_T) are provided in a trajectory centered coordinate system. (b) Trajectory centered coordinate system: The upper plot shows a point (x_i, y_i) on the trajectory represented in cartesian coordinates, while the lower one shows its intrinsic representation (α_i, l_i). (c) Sampling techniques: In the upper graph the trajectory is sampled with four points $(\times)$ being equally spaced with distance b. The lower one shows the cartesian x-axis ploted against time. Equidistant sampling concerning the t-axis leads to different points on the trajectory $(+)$.

Unfortunately, choosing the right representation is highly task dependent. Most recent publications either spend much effort in designing a very task specific solution using expert knowledge or skip it at all. This paper's intention is to present an evaluation procedure to support the designer in the decision-making process. For this purpose, Fig. 1 gives a short systematization of necessary steps to be considered. So far, this systematization is not complete, but it helps to get an idea of the necessary steps to follow.

The user's situation is described within the tracking data. To allow varying views on this situation, the data is split up into different forms of representations. These forms are achieved by using diverse reference systems (Fig. 1(a)), which may contain a different amount of information. In this paper three versions are discussed.

Choosing the right reference system has a high influence on the classification task. For example, if the robot would present general information, e.g. weather report, the direction from which a person approaches would play no decisive role. So a target referring coordinate system would be sufficient. But if the robot presents information about the place it is located in, e.g. in a museum, arriving people would probably be more interested than leaving ones. In this case a contextual view is essential.

First among the considered representation is a *contextual* reference system (Fig. 3(a)), modeling information about the surroundings. This paper suggests to provide information about the origin of the trajectory in a robot-centered way. This is done by specifying the first detected point of the trajectory (ϕ_O, d_O) in relation to the robot's fixed position. Most approaches are limited to such global types of representation. For the classification task presented in this paper using this technique is expected to lead to unsatisfying results.

The intention of the user to interact with the robot can be observed in a better way by centering the coordinate system to the user and specifying the relative position to objects of interest (Fig. 3(a)). Literarily spoken this helps the robot to put itself into the person's position. Such a *target referring* representation is done for the last point in time of each trajectory. Hence, the relative position of the nearest target (ϕ_T, d_T) with the smallest angle and the robot (ϕ_R, d_R) is stored. This is done by specifying the distance d and the angle ϕ relative to the person's motion direction.

Furthermore, to rid the representation of the surroundings a purely *trajectory centered* reference system is used. So, the characteristics of the trajectory is put into focus, which contains implicitly the velocity, acceleration and straightness.

Two different ways are suggested for describing the trajectory in a trajectory centered reference system (Fig. 1(b)). The first one (lower graph in Fig. 3(b)) can be compared with a route description. The trajectory is represented in an intrinsic way, which leads from one trajectory point to its subsequent one. This is done by specifying the distance l_i and the angle α_i with respect to the motion direction to the next point.

For the second one (upper graph in Fig. 3(b)) the origin of the cartesian (x, y)-coordinate system is set to the last trajectory point in time. The coordinate system's orientation is determined by the corresponding first trajectory point in the window. In our case the orientation is chosen in the way that this point lies on the x-axis.

Data originating from the tracker is sampled in a specific way. But what kind of sampling technique (Fig. 1(c) and Fig. 3(c)) helps to get most of the information? Sampling in a spatial manner helps to keep the number of data points constant for a certain trajectory length. For example, if a person is standing, only one data point is sampled instead of a point cloud. So, the shape of the trajectory is in the focus of this type of representation. This is similar to the way a HMM would handle the observations on the trajectory, compensating variations in speed and time.

On the other hand, sampling can be done in a temporal equidistant way. Hence, the speed of a person is coded implicitly in the trajectory.

Some of the information implied in each of the trajectory representations might be correlated. Finding out the correlated data helps to reduce the number of information that needs to be provided to the classifier. For this task Principal Component Analysis (PCA) is commonly used, which provides a set of Eigenvectors and Eigenvalues. The Eigenvectors, which are called Eigentrajectories (Fig. 1(d)) in this paper, span a new space, while the Eigenvalues indicate the variance of each dimension of the new space.

The systematization presented in this section is to be understood as a first proposal. Further efforts have to be made to achieve an exhaustive taxonomy. Still, it becomes clear, that a large number of representations can be generated. But which one is the correct choice for the given task of estimating the user's interest in an interaction with the robot? To make this determination, information theoretic measures like Mutual Information and Joint Mutual Information (Fig. 1(e)) are applied together with the class labels gained from the user observations and interviews. Both are discussed in the next section.

V. FEATURE SELECTION AND CLASSIFICATION

The decision whether an observed person is interested in an interaction with the robot is estimated by a neural classifier, more specifically a Multi Layer Perceptron (MLP). This classifier is trained with a part of the recorded data referred to as training set, and the evalution is done subsequently with the rest of the available data termed test set. Each variable or dimension of a specific trajectory representation is considered to be a feature.

Due to the abundance of possible input feature representations, the resulting neural classifier using all features is not only prone to overfitting and related problems like increased time demands for training, but we are interested in finding the best way to preprocess our data, too. Hence, we incorporated a feature selection step. These feature selection methods are applied to identify irrelevant and redundant features in the input space.

Feature selection methods are commonly distinguished into "Filter", "Wrapper" and "Embedded" approaches (see [10]). Both wrapper and embedded methods utilize a learning machine to evaluate the features, while the filter algorithms are independent of the used classifier. For this work a filter approach is chosen, because of this independence property. The relevance of each feature is assessed with a statistical measure, e.g. simple Pearson correlation coefficient, Fisher discriminant analysis or information theoretic measures.

In this case, Mutual Information (MI) is applied (see [10], chapter 6). MI expresses the correlation between a feature

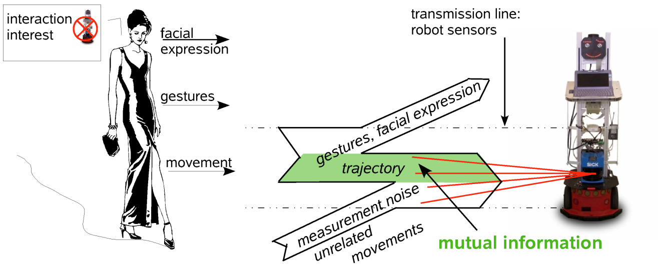

Fig. 4. Interpretation of the information-theoretic measure Mutual Information for our scenario. The approaching person sends signals of interaction interest via gestures, facial expression and movement trajectories. Up to now, the presented system only captures the movement trajectories including some measurement noise, while e.g. gestures are lost. Additionally, maybe some parts of the movement trajectory are not correlated to the interaction interest at all. The relevant signals observable by the robot are measured with MI.

x_i, e.g. the x-coordinate of the position at the first trajectory point, and the target Y, the class information representing the intention to interact or not to interact. High MI values indicate a strong correlation, while small values are hinting at irrelevant features. It can be determinded by the following equation, which is based on the well-known Kullback-Leibler divergence:

$$MI(x_i, Y) = \int_{x_i} \int_Y P(x_i, Y) \log \frac{P(x_i, Y)}{P(x_i)P(Y)} dx_i dY \quad (1)$$

A graphical interpretation of the meaning of Mutual Information in the context of the scenario is given in Fig. 4.

The computation of the probability densities is quite demanding, due to the necessary integration. In the practical implementation a histogram based approach was used to estimate the Probability Density Functions (PDF), simplifying the above equation by replacing the integrals with sums over the bins of the histogram. For more details on the histogram technique and other approaches for PDF estimation the interested reader is refered to [11].

The resulting MI value indicates the relevance of a feature with respect to the target, but it does not include information about possible redundancies between features. This drawback can be overcome by computing the pairwise Mutual Information $MI(x_i, x_j)$ between the features. In this case a high MI value indicates redundancy, low values independence.

A step further, the use of Joint Mutual Information (JMI) eliminates another possible problem, that occurs if two features are unimportant themselves, but provide information if they are used in combination (think of the classical XOR problem). JMI computes the relevance of a set of features $X = \{x_1, .., x_n\}$ with respect to the target in the following way:

$$JMI(X, Y) = \int_{X,Y} P(x_1, .., x_n, Y) \log \frac{P(x_1, .., x_n, Y)}{P(x_1, .., x_n)P(Y)} \quad (2)$$

Selecting feature sets from different trajectory representations with high relevance for the user's interest is the final step before the training of the classifier. The next section discusses which representations proved to be relevant.

VI. EXPERIMENTS & RESULTS

This section discusses the conducted experiments, the obtained results and what conclusions can be drawn.

After the trajectory recording sessions, we were left with a very unbalanced data set. The number of 43 interactions with the robot is quite low, compared to the number of 450 non-interaction cases. However, an interaction took place if and only if the user provided input to the system, e.g. pushed a button. Additionally, 34 individuals stated that they were undecided but finally did not interact with HOROS.

The analysis of the questionnaires revealed some of the reasons for this unexpected behavior. The main issues covered with our questions were:

- *How should a robot behave in public environments, especially in office environments?*
  As expected, the results showed that friendly and funny actions were preferred from the users and lead to most interactions. Sad or provocative robot voice outputs are not recommendable since only in a few cases successful interactions could be established by this behavior. Moreover those actions had a slightly negative influence on the users. Interestingly the results indicate that an informal way to talk to people is more suitable than being too formal.
- *Which robot actions are appropriate for robots in public environments, especially in office buildings, e.g. which voice output, robot movements and facial expressions should be used?*
  The choice of voice outputs was already discussed in the previous question. Additionally, people wanted to get as much information as possible about the robot's abilities during dialog initialization but with as little amount of words as possible. We found out that next to a voice output a movement of the robot is very important. Many people noticed that HOROS didn't move during the experiments at all and suggested to include movements in order to recognize the robot earlier while passing by the robot. Regarding the robot's facial expression, a friendly and smart smiling was the best choice.
- *Is a user-adaptive robot behavior necessary for such environments during dialog initiation and beneficial for future work?*
  In our long-term experiment we figured out that a user-adaptive robot behavior is necessary during dialog initiation. People who had to pass by the robot regularly and didn't have interest in an interaction tried to avoid the robot and partially changed their way through the building in order to be not recognized and contacted by the robot. Also we couldn't convince all undecided people just by using voice output. For those people, a more active robot behavior would be necessary. For people who had interest in an interaction a time-consuming greeting was needless. A short greeting and meeting a person half way is completely appropriate for these cases.

- *How good is* Horos *perceived in an environment by people who don't know about its presence and how much is the robot able to arouse interest in an interaction?*
  Since Horos was standing static during the experiments, it partially was perceived too late. The eye-catching appearance of Horos did not compensate the lack of movements.
- *Which robot dependent and independent variables influence the user's descisions to interact with the robot, e.g. the person's mood, haste and attitude towards robots?*
  In our questionnaires we tried to collect information about all possible facts that could have an effect on a user's intention to interact with the robot. Other variables like a persons mood, gender or a general interest in robots were apparently no factor for engaging in interactions.
- *Why did people dodge an interaction?*
  One motive are time constraints. Because of them, many people, students and employees alike, rushed through the hall in a hurry. Another reason was, that people avoided the robot because they were unsure if they were allowed to use the system or expressed fear to damage the robot. Those people who interacted with the robot once, refused to use him again, mostly because the application of the information terminal did not yield enough benefit for them.

The main insight gained during the data selection phase was, that people in a working environment need a strong reason to use a robot.

In the following, the recorded trajectories are examined in detail. The different forms of trajectory representation, see section IV, are compared to each other to identify a useful subset of trajectory features. At first, the time equidistant re-sampling method is compared to the spatial equidistant one, while intrinsic and cartesian trajectory centered coordinates, are analyzed in more detail. The following evaluations have been done once for both resampling methods, all leading to the conclusion that, contrary to expectations, the disadvantage of losing velocity information in the case of spatial resampling is not that relevant for the given classification task. For all further illustrations, we will concentrate on time equidistant resampling, knowing that spatial resampling will yield similar results.

The next point of interest is the amount of information contained in the raw data points considering the intrinsic and cartesian descriptions (see section IV).

Fig. 5 shows a plot of the Mutual Information between the individual features and the interest of people to interact, computed as described in section V. A look at the figure shows that in comparison cartesian coordinates (x_i, y_i) seem to contain more information about the target interest than intrinsic ones (α_i, l_i). Especially the x-coordinate of cartesian representation is coding the speed and thus seems to be the most relevant. By plotting the pairwise Mutual Information between the cartesian coordinates (see Fig. 6 (a)) it is observable that there is a high redundancy between

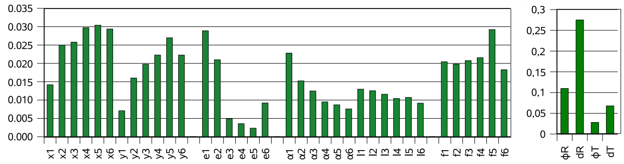

Fig. 5. This image shows the Mutual Information for the individual features of the datapoints to the interest in an interaction. Columns 1 to 12 show the resampled datapoints in cartesian coordinates (first x then y components), 13 to 18 describe the principal components for the cartesian representation, 19 to 30 show the data using intrinsic representation (angle α and length l of segments), 31 to 36 principal components of intrinsic representation. The right diagram shows the information of target refering data (relation to the robot d_R, ϕ_R and to the closest target point d_T, ϕ_T) on a different scale.

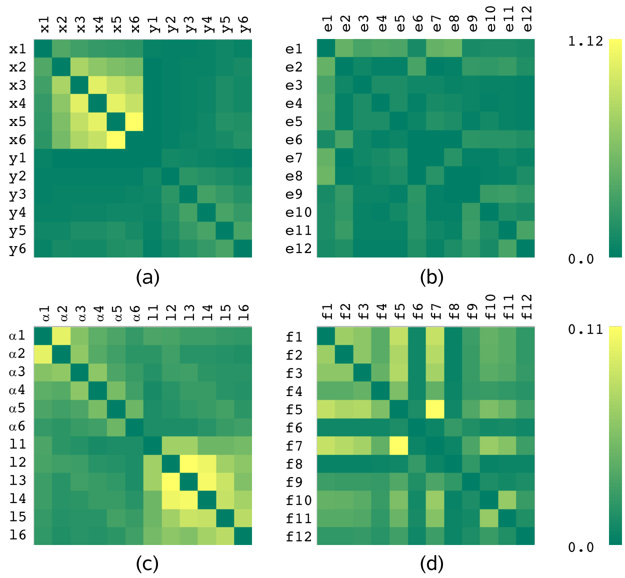

Fig. 6. Illustration of the pairwise Mutual Information matrices for the raw data in cartesian coordinates (a) the respective principal components (b), as well as the intrinsic representation (c) and corresponding principal components (d). Yellow/Brighter color represents high values and indicates that both variables share a lot of information, which means that they are redundant, at least partially. e.g. the MI between x_6 and x_5 is very high and thus one of the variables is redundant. Green/Darker ones imply lower values and an independence of the considered variables. (a) and (b) have the same scale which is different to the one of (c) and (d) which is lower at all.

the features in the trajectory window. For example, looking at matrix entry (x_5, x_6), which has a high pairwise Mutual Information leads to the conclusion that only one of the two features is necessary for the classification process. In comparison with entry (y_1, y_4) the low value between those leads to the conjecture that both features are not redundant. The x-coordinates have higher correlations because they result from translative motion, whereas the y-coordinates are coding the shape of the trajectory in the window including the curvature and therefore are less correlated among each other than the x's. An absolute quantitative comparison is fairly hard but the relation is observable.

In contrast to this, angular descriptors (Fig. 6 (c)) of the intrinsic representation are less redundant. But for this representation the length of the segments is highly correlated,

MLP inputs	Network complexity	error
raw data $(x_1, y_1, \ldots, x_6, y_6)$	input 12, hidden 16	0.2252
PCA data $(e_1, \ldots, e_{12})$	input 12, hidden 16	0.2454
raw data (x_4, y_4, x_5, y_5)	input 4, hidden 10	0.2057
PCA data $(e_1, \ldots, e_4)$	input 4, hidden 10	0.193
$(x_4, y_4, x_5, y_5), (\phi_R, d_R, \phi_T, d_T)$	input 8, hidden 16	0.1756

TABLE I

Classification error of MLPs using different subsets of features (first column) extracted from the windows of people's trajectories (see section IV). The binary output is compared to the class information and the errors are averaged over the test data set to compute the final error rate. The best result was achieved using the *target refering* information additionally.

because people usually do not vary their velocity that fast.

As introduced in section IV, the method used to reduce linear redundancies is PCA. In Fig. 5, the information of the most significant principal components e_i and f_j (the projection on Eigentrajectories) are given for both kinds of representation. In the cartesian case, the information is concentrated in the most significant components, which are nearly decorrelated, as a look on the pairwise Mutual Information matrix Fig. 6(b) shows clearly. These components describe the local shape of the trajectory, almost as well as the complete raw data does.

In contrast the intrinsic representation decorrelates the points on themselves. Considering the different scale, the Mutual Information matrix shows that the correlations between the principal components does not decrease at all. As a result we conclude, that the shape information of the *trajectory centered* data is expressed in the most compact way by projecting the cartesian raw data onto the most relevant Eigentrajectories (principal components).

The experiments with the classifier that has been trained once with the raw data and once with the data compressed with PCA, reached similar performances with a quite complex network. But as given in Table I, the compressed representations (line 2 and 4) allow to keep the network simple and, therefore, a better generalization over the test data set can be reached. During the learning phase, the target is set to one for interaction interest and zero for non-interacting trajectories. By means of this, the classification is done by thresholding the real-valued output in the interval $[0, 1]$ at 0.5. Hence, an expected interaction is coded as one and a non-interaction as zero. The classification error results from averaging the binary classification results over all trajectory windows in the test data set. This includes the ambiguous parts observed in the beginning of a trajectory.

A last experiment considering the obtainable classification accuracy was conducted. The expectation from Fig. 5 is that the classification rate could be increased once more by taking into account the target refering information for each trajectory window. Finally, without optimization of any network structure, a rate of about 82% correct classifications (see line 5 in Table I) promises a great benefit when selecting the right action to attract interested people or avoid incurious ones.

VII. CONCLUSION

This work concerns the task of classifying people in the surroundings of the robot as interested in an interaction or as incurious. The available hypotheses about the environment come from the people tracking system of the robot, which is used to create movement trajectories. After discussing different approaches to represent trajectories, a method was presented to prune these possibilites by means of information theoretic measures.

This method conveniently supports the design process and helps to break dependecies from expert knowledge. Hence, the presented method is applicable for a wider spectrum of applications than the one presented here. Projecting the cartesian raw data into a PCA subspace proved to be the best suited approach for the trajectory classification task.

In our future, work this knowledge about the user's current interest in an interaction will be used to automatically adapt the welcoming strategy of the robot.

Further on, the presented approach will be applied to estimate other user specific properties, like the degree of urgency. This will allow us to refine and personalize the dialog between human and robot even further.

REFERENCES

[1] H.-M. Gross, J. Richarz, S. Müller, A. Scheidig, and C. Martin, "Probabilistic multi-modal people tracker and monocular pointing pose estimator for visual instruction of mobile robot assistants," in *Proc. 2006 IEEE World Congress on Computational Intelligence (WCCI 2006), International Joint Conference on Neural Networks (IJCNN-06)*, 2006, pp. 8325–8333.

[2] H.-J. Böhme, A. Scheidig, T. Wilhelm, C. Schröter, C. Martin, A. König, S. Müller, and H.-M. Gross, "Progress in the development of an interactive mobile shopping assistant," in *Proc. of the Joint Conf. on Robotics: 37th International Symposium on Robotics (ISR 2006) and 4th German Conference on Robotics (Robotik 2006)*, 2006.

[3] M. Finke, K. L. Koay, K. Dautenhahn, C. Nehaniv, M. Walters, and J. Saunders, "Hey, i'm over here - how can a robot attract people's attention?" in *in Proc. 14th IEEE International Workshop on Robot and Human Interactive Communication (RO-MAN 2005)*, Nashville, USA, 2005, pp. 7–12.

[4] S. Nabe, T. Kanda, K. Hiraki, H. Ishiguro, K. Kogure, and N. Hagita, "Analysis of human behavior to a communication robot in an open field," in *Proceedings of the Conference on Human-Robot Interaction (HRI)*. Salt Lake City, USA: ACM, 2006, pp. 234–241.

[5] M. P. Michalowski, S. Sabanovic, and R. Simmons, "A spatial model of engagement for a social robot," in *Proceedings of the 9th International Workshop on Advanced Motion Control (AMC 2006)*. Pittsburgh, USA: IEEE Computer Society, 2006, pp. 762–767.

[6] J. Schulte, C. R. Rosenberg, and S. Thrun, "Spontaneous, short-term interaction with mobile robots," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*. Detroit, USA: IEEE Computer Society, 1999, pp. 658–663.

[7] H. Holzapfel, T. Schaaf, H. K. Ekenel, C. Schaa, and A. Waibel, "A robot learns to know people - first contacts of a robot," in *KI*, vol. 4314. Bremen, Germany: Springer, 2006, pp. 302–316.

[8] A. Scheidig, S. Müller, C. Martin, and H.-M. Gross, "Generating person's movement trajectories on a mobile robot," in *Proc. RO-MAN 2006 - 15th IEEE Int. Symposium on Robot and Human Interactive Communication*, 2006, pp. 747–752.

[9] B. Schelter, M. Winterhalder, and J. Timmer, Eds., *Handbook of Time Series Analysis*. Wiley-VCH, 2000.

[10] I. Guyon, S. Gunn, M. Nikravesh, and L. Zadeh, Eds., *Feature Extraction, Foundations and Applications*. Springer, 2006.

[11] D. W. Scott, *Multivariate density estimation: theory, practice, and visualization*. John Wiley & Sons, 1992.