Adaptive Spacing in Human-Robot Interactions
Panagiotis Papadakis, Patrick Rives, Anne Spalanzani

To cite this version:
Panagiotis Papadakis, Patrick Rives, Anne Spalanzani. Adaptive Spacing in Human-Robot Interactions. IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, IROS’14, Sep 2014, Chicago, United States. hal-01010224

HAL Id: hal-01010224
https://inria.hal.science/hal-01010224
Submitted on 19 Jun 2014
Adaptive Spacing in Human-Robot Interactions

Panagiotis Papadakis¹, Patrick Rives¹ and Anne Spalanzani²

Abstract—Social spacing in human-robot interactions is among the main features that is useful when integrating human social intelligence into robotic perception and action skills. One of the main challenges, is to capture the transitions incurred by the human and further take into account robot constraints. Towards this goal, we introduce a novel methodology that can instantiate diverse social spacing models depending on the context and further as a function of uncertainty and robot perception capacity. Our method is based on the use of non-stationary, skew-normal probability density functions for the space of individuals and on treating multi-person space interactions through social mapping. We show the utility of our approach in practice using an indoor robot operating in the presence of humans, allowing it to exhibit socially intelligent responses.

I. INTRODUCTION

As technological and societal progress set forth the deployment of robots within human society, humans and robots are expected to share space on the premise of natural interaction. In this versatile problem, social intelligence can be inscribed into robotic skills at various dimensions of human-robot interaction. Among those, understanding and controlling social space [1], [2], [3] is a core element of human social behaviour and further pertinent to robots as it can be expressed through their motion.

Social space sensitivity as a static property does not convey information for an interaction, however, its amplitude changes are strong, implicit cues for the initiation, evolution and termination, or avoidance of an interaction. For this reason, a major part of human-robot interaction studies [4], [5], [6], [7], [8] has sought to reveal the factors influencing sensitivity in relation to social spacing. Such studies verify what is well known from human sociology and psychology studies, i.e. that theoretical models may vary considerably across human individuals but they further give new insights, occasionally via conflicting findings that arise due to the involvement of robots [9], [10].

It is thus becoming more and more evident that socially-compliant robot skills should first be able to accommodate an extended range of social spacing models (Fig. 1) and second, allow smooth transitions during human-robot interactions. Despite the noticeable progress at distinct steps, however, the overall problem still lacks a framework that is both sufficiently descriptive of human social behaviour and at the same time accounting for robotic constraints.

Fig. 1. Varying poses (top) and corresponding interpretation (bottom) using adaptive spacing: (a) frontal motion, (b) neutral and (c) dominant side.

Most earlier approaches have focused on the first aspect through the development of complex computational models of social spacing, for example, as functions of human posture [11], direction [12] or intention of interaction [13]. The most common route is to aggregate distinct functions which inevitably produces discontinuities within the synthesized social space and consequently, those are not suited for robots that should respond smoothly to transitions in social sensitivity. The function of an individual’s social sensitivity that we proposed in our earlier work [14] surpassed this limitation, nevertheless, its design did not allow a generalization to varying social behaviour models. Probably the most elaborate attempt to address most of the challenges can be found in the work of Lam et al. [15] that accounted for various models of sensitivity for humans and robots as well as uncertainty in perception. However, due to the usage of a finite-state machine for modelling interaction cases and hard distance thresholds, smooth transitions are not feasible.

Overall, we can assert that earlier works converge on a set of features having the strongest influence in social spacing, that we identify as the lateral and vertical dominance. Although these depend on a variety of factors such as cultural background, age, etc. they are statistically biased towards the right [9], [16], [17] and frontal area [18], [19], [20] respectively. Otherwise, in the absence of such bias the shape of social space is reduced to a kernel in the form of concentric spheres or ellipses. Literature extends to numerous other works in a similar spirit, hence we focus here onto those that we deem more relevant and refer to recent surveys [21], [22] for the interested reader.

In this work, we propose a method to express the dominant factors that bias an individual’s social spacing by a single model (example shown in Fig. 1), allowing smooth transitions and in relation to uncertainty and perception capacity.

¹P. Papadakis and P. Rives are with team LAGADIC at INRIA Méditerranée, Sophia-Antipolis, France{panagiotis.papadakis, patrick.rives}@inria.fr
²Anne Spalanzani is with UPMF - Grenoble 2 - INRIA - LIG, France anne.spalanzani@inria.fr
We pursue this by using non-stationary, skew-normal probability density functions and build upon our previous work on social mapping [14] to deal with interactions of multiple participants. The complete methodology allows detection and analysis of diverse human social behaviours while doing so in real-time using conventional computing power.

The remainder of this paper is organized as follows. In Section II we unfold in detail our approach and highlight its theoretical contribution. In the sequel, in Section III, we show the utility of our approach through experiments on an indoor robotic platform operating in the presence of humans and conclude by summarizing our findings in Section IV.

II. PROPOSED APPROACH

We organize our description in three parts. First, in section II-A, we describe how our new model uniquely expresses the prominent social spacing models for individuals while in section II-B we link it with uncertainty and robot perception capacity. Finally, in section II-C we review our approach for mapping social interactions of multiple participants in relation to the new model of individual social spacing.

A. Atomic social space function

For the construction of an individual’s social space we begin by collecting dominant social spacing cues. Here, we consider (i) the position of the human \( t = (t_x, t_y)^T \in \mathbb{R}^2 \), (ii) the orientation \( \theta \in [0, 2\pi) \) and (iii) the dominant side \( d \in \{-1, +1\} \) where \(-1,+1\) correspond to the left and right side of the human respectively.

These cues are collected from a module that performs human detection and tracking of the articulated human motion (see section IV) under known conditions. The position of the human is taken as the body centroid while the orientation from the torso orientation. The dominant side is inferred either implicitly by the minimum lateral distance of the human from obstacles, based on the hypothesis that people maintain a smaller distance from their dominant side and vice versa, or explicitly through a distinctive hand gesture. In this way, a person is finally denoted as \( p = (t^T, \theta, d)^T \).

On the basis of these atomic cues, we construct a function \( IS_p(\cdot) \) for the social space of an individual \( p \), whose parameters are controlled through certainty, robotic perception capacity and the way in which these influence common social spacing models. We begin by identifying the social spacing models that are prominent within the literature and adopted in the majority of human-robot interaction studies, as follows; (a) concentric circles [1], (b) egg-shape [18], (c) ellipse [23] and (d) dominant-side [24], as sketched at the top of Fig. 2.

Through an examination of these models and earlier studies (see Section I) we can distinguish a set of generative properties of atomic social spacing. First, all models suggest that social sensitivity is a concave function of interpersonal distance. In other words, it is a unimodal function where sensitivity is maximum at the centroid of the person, while it degrades in the remaining surrounding area. Second, sensitivity follows a progressive decay as a function of distance, which allows us to naturally assume the existence of a smooth function, where smoothness implies infinite differentiability. And finally, a person’s sensitivity may exhibit asymmetry along the vertical and lateral dimension.

To express these characteristics within a single representative function, we appoint to \( IS_p(\cdot) \) a bivariate Skew-Normal (SN) probability function denoted as \( \mathcal{N} \), which is by definition a smooth and concave function, satisfying the first two requirements. In particular, the family of skew-normal distributions [25], [26] constitutes a superset of the normal distribution \( \mathcal{N} \), being further parameterized by the so-called shape parameter that controls the amount of skewness, which allows us to accommodate the last requirement by introducing skewness along an arbitrary direction. Following [26], this is achieved by setting \( IS_p(\cdot) \) to be a probability density function distributed as \( \mathcal{N}(t, \Omega, \alpha) \) defined as:

\[
IS_p(u) = 2 \phi(u) \Phi(\alpha^T u)
\]

where \( u \in \mathbb{R}^2 \), \( \phi(.) \) denotes the normal probability density function \( \mathcal{N} \) with covariance matrix \( \Omega \). \( \Phi(.) \) is the respective cumulative distribution function of \( \phi \) and \( \alpha = (\alpha_1, \alpha_2)^T \) is the parameter vector that controls the skewness. Without loss of generality, we take the position of the human \( t \) to be centred at the coordinates origin \( O \) and with body direction \( \theta = \pi/2 \) aligned to the vertical axis. Furthermore, due to the inherent characteristic of the SN distribution that its mode does not coincide with that of the non-skewed version, we translate the skewed density function so that its mode coincides with the location of the human. In lack of a known closed-form solution, we compute the location of the mode numerically, e.g. through gradient descent. Finally, since we use \( IS_p(\cdot) \) to account for the social sensitivity of an individual human in real world coordinates, an isotropic scaling can be applied so that proxemics-based distances or other experimentally derived distances can be assigned to the isocontours of the function. The second row of Fig. 2 shows the resulting \( IS_p(\cdot) \) function that is attributed to a human and how its parameters are controlled to instantiate the main spacing models.

\[
\begin{align*}
\alpha &= \theta^T \\
\Omega &= \sigma^2 I \\
\alpha &= (0, \alpha_2)^T \\
\Omega &= \sigma^2 I \\
\alpha &= (\alpha_1, 0)^T \\
\Omega &= \sigma^2 I
\end{align*}
\]

Fig. 2. Top row: prominent social spacing models for individuals (a) concentric circles, (b) egg-shape, (c) ellipse and (d) smaller in dominant side. Bottom row: corresponding interpretation by controlling the parameters of the bivariate skew-normal probability density function.
B. Adapting to perception capacity and uncertainty

We proceed by linking the atomic sensitivity function with two distinctive dimensions in human-robot interaction, namely, robotic perception capacity and uncertainty. This is necessary for making our approach applicable to varying settings and allowing smooth transitions among models.

The first dimension refers to the ability of a robot to infer the information about the body orientation $\theta$ and dominant side $d$. This ability depends on the sensors and algorithms that a robot is equipped and in turn, determines the form of the $p$ vector for characterizing a human. To account for these differences we distinguish which social spacing models are applicable in each case as given in Table I where the first column lists possible perception modalities while the applicable social spacing models are given in the second.

<table>
<thead>
<tr>
<th>Perception capability</th>
<th>Applicable models</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p = (a)$</td>
<td>(a)</td>
</tr>
<tr>
<td>$p = (a', \theta)$</td>
<td>(a), (b), (c)</td>
</tr>
<tr>
<td>$p = (a', d)$</td>
<td>(a), (c), (d)</td>
</tr>
<tr>
<td>$p = (a', \theta, d)$</td>
<td>(a), (b), (c), (d)</td>
</tr>
</tbody>
</table>

Once perception capacity is given, we continue by accounting for the second distinguished dimension in human-robot interaction, i.e. uncertainty. This is necessary as switching between social spacing models may not only occur due to the social behaviour of the human, but it could further be the result of sensing uncertainty. Common factors that introduce uncertainty within the individual elements of the vector $p$ during human-robot interactions are self-occlusions, occlusions from other humans, occlusions due to the environment and range limits.

Giving an explicit model of uncertainty that accounts for all these factors is beyond our scope here as that depends on the hardware set-up and environment conditions. Instead, we focus here on the more general case wherein uncertainty is either the result of excessive or too short distance between the robot and the human that may prevent capturing of the complete human body. This can be alleviated by regulating transitions among the main social spacing models. Our hypothesis is that the concentric circles model (a) constitutes the most basic and simple social spacing model that is assigned at minimum certainty, while all the remaining models (b), (c) and (d) arise as certainty increases and completely outweigh model (a) when certainty is maximized.

Let us denote as $IS_{m,p}(u)$ a specific social space function for a person $p$ where $m \in \{a,b,c,d\}$ enumerates the corresponding model and as $p \in [0,1]$ the certainty factor, then the expected social sensitivity is obtained as:

$$IS_p(u) = IS_{a,p}(u) \cdot (1 - p) + IS_{m,p}(u) \cdot p \tag{2}$$

Eq. (2) implements the transitional mechanism between spacing models as a function of certainty $p$. In order to further guarantee smooth transitions certainty is in turn expressed as a function $\rho(.)$ that is $C^k$ (continuous up to the $k^{th}$ order).

A practical example is given in Fig. 3 that depicts frames where certainty regarding social spacing cues varies as a result of distance from the human, following a 1D Gaussian approximation and in line with related findings [27]. Initially, the human resides at the far limit of the perceptual range (frame #i) bearing minimum certainty $\rho_i$, then he resides at the center (frame #j) with maximum certainty $\rho_j$ and finally (frame #k) continues to approach which reduces certainty to $\rho_k$ as the body extends beyond the field of view. Finally, we note that while our approach favours smooth model transition it also allows hard switches that may naturally occur (e.g. through an intentional sudden posture change).

C. Mapping interactions of multiple participants

In the previous section we showed how we can effectively address the problem of adaptive social spacing for individuals through a general framework and by taking into account robot perception capacity and uncertainty. In the sequel, we build upon our previous work on social mapping [14] that provided a general framework for mapping the collective interactions among humans by extending the notion of personal space to interpersonal social spaces. Its formulation allows the usage of any type of atomic space function as long as it satisfies the conditions for being a probability density function. Here, we employ the social mapping paradigm on the new atomic social space function which uses the SN pdf.

At the first step, we take $m_i$ i.i.d. (independent and identically distributed) samples from the atomic social space function $IS_p(.)$ of each individual $i = 1,2,...,N$ where $N$ is the total number of individuals. For the determination of the number of samples $m_i$ for each human we follow the same principle as before, i.e. using the detection certainty to regulate the number of samples so that $m_i \in \{0,1,...,m_{max}\}$. We gather the total number $M = m_1 + m_2 + ... + m_N$ of sampled points from all individuals into a set $S = \{s_1,s_2,...,s_M\} \subseteq \mathbb{R}^2$, holding the data that we will use to reconstruct a global density function of social sensitivity.

Following [14], we formulate this as a one-class density estimation problem and we solve it by employing Kernel Principal Component Analysis for novelty detection [28]. This involves the computation of the centralized Gram matrix $K$ for the set $S$ whose $(i,j)^{th}$ entry is given by:
\[ \tilde{K}_{ij} = K_{ij} - \frac{1}{M} \sum_{r=1}^{M} K_{ir} - \frac{1}{M} \sum_{r=1}^{M} K_{rj} + \frac{1}{M^2} \sum_{r,s=1}^{M} K_{rs} \]  
(3)

where in our implementation we set \( K_{ij} = k(x,y) = \exp(-\gamma ||s_i - s_j||^2) \). This step is concluded by computing the eigen-decomposition of \( \tilde{K} \) which gives a set of eigenvectors \( e^l \), \( l = 1,2,...,M \) and corresponding eigenvalues \( \lambda^l \).

Finally, we use the reconstruction error within the subspace of the first \( 1 \leq l_{max} \leq M \) principal directions as the global sociality density estimate \( D(q) \) for an arbitrary point \( q \in \mathbb{R}^2 \). As was shown in [14], using Kernel PCA for novelty detection is superior to a Parzen window aggregation as the former concentrates the density within groups of humans and ensures that the density is maximized on the human position.

### III. Experiments

We have performed our experiments with an indoor robotic platform Neobotix MP-500 which employs a two-wheel differential drive mobility system (see Fig. 4 (a)) and using the Robot Operating System (ROS) [29]. Robot navigation was semi-supervised while detection and analysis of (inter)personal social spaces were entirely automated. To detect the articulated human motion we used the Asus XtionPROLive camera which captures synchronized RGB-D images in conjunction with the OpenNI library. An array of 8 cameras in total are mounted sideways on top the robotic platform (Fig. 4 (b)), by regular 45° increments that allow a panoramic sensing of the environment. Table II describes our experiments on the basis of the Human-Robot Interaction Taxonomy provided in [30]. The goal of these experiments is to demonstrate the utility of the proposed framework in diverse scenarios. Following up earlier work [31], [20] that developed socially-intelligent path planning strategies on the basis of a given set of interacting people our focus here is on the actual extraction of the higher-level spatial relationships within real human-human and human-robot encounters.

#### a) Experiment #1:
Two trials were first performed by each of three participants in a scenario where the human and the robot approach each other along a narrow corridor. An explicit or implicit interaction was required so that the robot can bypass the human safely and with minimal disturbance to the atomic social space. Fig. 5 shows snapshots from a specific trial and the respective stages of situation awareness for the robot. In detail, once the human is first detected at a distance where certainty is minimum he is attributed with the basic spacing model of concentric circles (Fig. 5 (a)). As he approaches (Fig. 5 (b)), certainty increases with respect to the body orientation which coincides with the motion direction and hence the egg-shaped model is assigned. Upon approach, the human signifies explicitly that

<table>
<thead>
<tr>
<th>Task type</th>
<th>Experiment #1</th>
<th>Experiment #2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Robot morphology</td>
<td>Socially compliant path-planning</td>
<td>Functional</td>
</tr>
<tr>
<td>Human roles</td>
<td>Bystander</td>
<td></td>
</tr>
<tr>
<td>Physical proximity</td>
<td>Passing</td>
<td>Synchronous</td>
</tr>
<tr>
<td>Time</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Space</td>
<td>Collocated</td>
<td></td>
</tr>
<tr>
<td>Human-Robot ratio</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Interaction</td>
<td>Robot &amp; Human</td>
<td>Robot &amp; Multiple humans</td>
</tr>
<tr>
<td>Autonomy/Intervention</td>
<td>50%/50%</td>
<td>100%/0%</td>
</tr>
</tbody>
</table>

Fig. 4. Illustration of (a) the robotic platform that was used for the indoor experiments and (b) the corresponding perception setup.

For a video illustration of the presented experiments please refer to the accompanying file provided as supplementary material.
he gives priority to the robot to pass from his right side through a distinctive hand gesture. This is interpreted as a cue that signifies the human’s dominant side, shifting the social sensitivity density of the human towards his left (Fig. 5 (c)) and allowing the robot to bypass the human from the right side (Fig. 5 (d)). This experiment demonstrates how our framework allows transitions among social spacing models by appropriate control of the non-stationary, skew-normal density function in a continuous mode of operation.

Within the total set of trials, the participants either completely gave all their space clearance to allow the robot to pass or performed a similar distinctive gesture to signify their dominant side. As our aim here is not to demonstrate a gesture recognition approach, in Fig. 5 we chose to illustrate the trial where the hand gesture was mostly distinctive and how it affected the adaptation of the atomic social space.

b) Experiment #2: In the sequel, multiple humans interact within a spacious area while the robot employs social mapping [14] in order to extract interpersonal social zones at various density levels. This is done by detecting and analysing the spatial interactions on a per-frame basis using the isocontours of different levels of sociality density. Fig. 6 (top row) depicts various such examples taken from a sequence wherein humans interact with each other, together with the corresponding social mapping output (bottom row). Three increasing levels of global social sensitivity have been set which we enumerate as I, II and III respectively, while the thickness of the contour of each zone is set proportional to the number of contained humans. The different frames illustrate how our proposed approach can extract the basic features that are useful for a robot to understand the spatial relations among humans within a highly dynamic scene.

At a second step, we have evaluated the utility of the complete proposed framework by showing different path trajectories which were planned by the robot when instructed to traverse the scene from a starting point A to a destination, oriented point B, using the baseline ROS trajectory planner that employs the Dynamic Window Approach (DWA) [32]. The trajectories are shown in blue color at the bottom row of Fig. 6 and are the outcome of a priority-based path planning scheme which operates on the 2D grid produced by social mapping and is conditioned on the I, II and III levels.

In short, within this scheme path planning is initially constrained to the free space of the exterior of the total I level. If this area is too small to allow a feasible path, path planning is performed on the next larger area as specified by the exterior of the II level and if necessary this continues until the III level. Following this rule, the extracted paths which are shown for each frame are those that maximally respect the global social sensitivity. This extends the notion of adaptive spacing proposed by our method to the notion of adaptive socially-compliant path planning. Here, we demonstrate this feature using DWA and believe that it could be equivalently integrated with other relevant strategies.

We finally note that despite the fact that our approach is transparent to the number of human participants, the previous examples involved a maximum of three collocated humans since additional members typically occluded those being previously detected. Therefore, to complement our experiments we applied our approach within a synthetic scenario that involves 5 collocated people and show the corresponding results within Fig. 7. To compensate for the lack of realism, we have used real human Motion Capture (MOCAP) sequences of dynamic pair-wise interactions taken from the Carnegie Mellon University dataset (http://mocap.cs.cmu.edu/). While such a scenario clearly surpasses the perception capabilities of contemporary robots in terms of detection and analysis of the articulated human motion, it is provided to show the applicability of the proposed methods in conditions that are far more complex that those explored in earlier works.

In terms of efficiency, execution rate mainly depends on the number of detected humans which prescribes the total number $M$ of points that are sampled to reconstruct the global sensitivity density. All other parameters are set constant according to the employed sensors, which in our set-up yielded an execution rate ranging from 40 to 8 frames/sec.
We introduced a novel probabilistic framework capable of instantiating diverse models of social spacing and accounting for distinctive dimensions in human-robot interaction, namely, perception capacity and certainty. We have concretely shown how our method allows smooth adaptation in the situation awareness of a robot within common human-robot interaction examples and further showed its utility at the level of path planning by adapting trajectories to social sensitivity levels. By coupling skills of human social behaviour perception with path planning within real interaction scenarios, we believe that our work makes a considerable forward step in the field of socially-embedded robots.

The description of our methods was focused on skills of social intelligence where robots should avoid interrupting human activities. As part of our ongoing work we are interested in integrating these skills with complementary behaviours where robots are able to initiate, maintain and terminate interactions with humans. As this involves the usage of additional behaviour cues and superior recognition skills, we plan to investigate the role of certainty and perception capacity in human-robot interactions at a finer level.

IV. CONCLUSIONS

REFERENCES