Deducing evidence for social situations from dynamic geometric interaction data

Georg Groh* and Alexander Lehmann

Faculty of Informatics,
Technische Universität München,
Boltzmannstr 3, 85748 Garching, Germany
E-mail: grohg@in.tum.de
E-mail: lehmanna@in.tum.de
*Corresponding author

Abstract: We discuss how time-independent and time-dependent features of human social interaction geometry on small temporal and spatial scales may be used to extract evidence for or against the existence of social situations as a simple form of social context. Aside from providing a new method for quantitative investigation of human interaction behaviour, the ultimate vision motivating this research focuses on mobile devices autonomously measuring and processing data on interaction geometries in order to derive social situation context that can be used in mobile social networking scenarios. Our method is tested via an experiment using an IR tracking method already allowing for the precise determination of interpersonal distances and relative body orientation in a conversational setting. We investigate the performance of time-independent classifiers for the prediction of the involvement of pairs of persons in a social situation using relative distance and orientation. We then discuss results of using HMMs for exploiting the time-dependency of the interaction geometry.

Keywords: social signal processing; SSP; contextual social networking; geometry of social interaction, interpersonal distance and orientation; social situation.

Reference to this paper should be made as follows: Groh, G. and Lehmann, A. (xxxx) ‘Deducing evidence for social situations from dynamic geometric interaction data’, *Int. J. Social Computing and Cyber-Physical Systems*, Vol. X, No. Y, pp.000–000.

Biographical notes: Georg Groh obtained his diploma in Theoretical Physics and Informatics from German Universities Kaiserslautern, Hamburg and TU München. He then joined the Chair for Applied Informatics/Cooperative Systems at the Faculty for Informatics of TU München and received his PhD in Informatics with a thesis on ad-hoc groups in mobile communities. He is currently a Post-doc. pursuing his Habilitation degree in Informatics. His main scientific interests include contextual social networking and applied social signal processing.

Alexander Lehmann obtained his diploma in Informatics from TU Munchen. He is also a Commercial Pilot with several years of professional experience at Lufthansa. His diploma thesis at the Chair for Applied Informatics/Cooperative Systems dealt with algorithmic analysis of high-precision, infrastructure-free (indoor and outdoor) measurements of position and orientation in view of social situation detection. He is currently pursuing his PhD in Informatics in the field of applied social signal processing.

1 Introduction

Social networking as an important trend and paradigm in social computing with its growing emphasis on *explicified relations* between actors has been a very important stimulus for research on human social networks in recent years (see, e.g., Newman et al., 2006). While very interesting insights into the medium- and long-term social interaction patterns of human beings could be won using network data from the web (see, e.g., Lewis et al., 2008), algorithmic characterisation of *social interactions on small temporal and spatial scales* is still an emerging field of study. One field of study that is concerned with the algorithmic modelling and characterisation of social interaction on small temporal and spatial scales based on the analysis of social signals and behavioural cues, e.g., extracted from low level data sources such as audio- and video-streams, is *social signal processing* (SSP) (Vinciarelli et al., 2009). Here, time-scales of minutes or even seconds as opposed to days, weeks and months, as, e.g., deduced from social network data from online social networks, are considered. While SSP focuses on deriving models for *social* interaction and thus models of social context, the closely related field of reality mining (see, e.g., Pentland and Eagle, 2009) also regards individual behavioural patterns beyond immediate social interaction.

Among the data sources that have not been very extensively studied on a precise quantitative and algorithmic level are measurements of the *geometry of social interaction* such as interpersonal distances and body orientation. In this contribution, we investigate in how far the existence of *social situations* (as simple models of social interaction) can be deduced from *investigation of time-series of interpersonal distances and relative body orientation*. We mainly focus on using social situation models as a simple, useful form of social context for mobile social networking (MSN) services (see below).

In view of our research question we conducted a *quantitative experiment* that delivered an appropriate, precise dataset of interaction geometries. We then investigated suitable methods to process data of this kind in view of the detection of social situations. In doing so, we intend to make the following contributions:

On the one hand, as a first step, the results of annotating and suitably processing these dynamic data are intended to contribute to *a quantitative, general, algorithmically usable model of human behaviour with respect to interaction geometry* in order to render more precise the existing, often rather qualitative models (e.g., models of interpersonal zones, Hall, 1959; Mehrabian, 1959; Willis, 1966; Sundstrom and Altman, 1976; Amaoka et al., 2009). Furthermore, the developed experimental method as well as the methods for processing the data into time-dependent and time-independent interaction models can provide a tool-set for future sociological research to extend the resulting models of social interaction geometries to, e.g., other cultural frames and other types or levels of detail of social interaction.

On the other hand, as a second step, the possibility to measure these data with increasing precision using *mobile devices* [see Pentland and Eagle (2009) and Pentland et al. (2009) for a discussion of mobile devices as social sensors] makes it possible to measure and process geometrical interaction data in co-located MSN scenarios, using the aforementioned interaction models. That is, mobile devices of two users may measure their users' relative orientation and interpersonal distance in an preferably *infrastructure-independent* way and may then use the quantitative interaction model resulting from our experiment and its evaluation in order to deduce whether the two users are in a social situation or not. This simple form of social context may then be used for useful MSN services (Groh et al., 2010b). MSN can be viewed as broadening of the social networking paradigm by detecting and modelling the user's individual and social context via sensors in mobile (or wearable) devices and by using these context models and selectively opening them (as part of the individual information space) to other users from the network for new services or variants of existing services with increased usefulness.

In this article, we will start with a short discussion of our model for social situations and possible fields of application in MSN scenarios. After a subsequent brief examination of related research on geometry of social interactions, we describe our experiment that delivered precise real world data on interpersonal distances and orientation. We then briefly present previous classification results on the basis of a general time-independent model derived from this data. These results provide a baseline for the main part of the contribution, the investigation of whether and how this model may be enhanced by taking into account the dynamics of time-series of interpersonal distances and relative orientation data. We conclude with a short section on future research activities.

2 Social situation model and applications

As a special form of social context, a social situation is an instance of a simple indicator model for the existence of co-located social interaction of two or more persons. A social situation is defined by *full mutual awareness* of the existing social interaction among these persons (awareness, awareness of this mutual awareness, awareness of the awareness of mutual awareness and so forth). Social interaction encompasses all forms of verbal or non-verbal communication (the exchange of social signals such as gestures, posture, facial expressions, etc.). In other words, if two or more people are fully aware of their mutual social interaction, a social situation is established. For the purpose of this paper we exclude non-co-located social interaction, e.g., via textual chats or via phone conferences, since interaction geometry is not significant in these cases.

Our general model S of social situations consists of a four tuple:

$$S = \left(P, T \subset \mathbb{R}, X \subset \mathbb{R}^3, K\right).$$

P denotes the set of persons participating in a social situation, T denotes the time interval of the situation, X denotes the spatial reference of the social situation, which is usually given by the convex hull of all locations (centre of body mass) of all participants during the interaction, and K represents a set of keywords describing the semantics of the social situation. Naturally, social situations can be arbitrarily nested with respect to all tuple elements. The model assumes that at each point t in time, a social situation either does or does not exist, or, in other words, that the temporal reference T can be properly defined up to a reasonable degree of precision with a crisp indicator function μ_T: $\mathbb{R} \rightarrow \{0, 1\}$ for T:

$$\mu_{T(s)}(t) = \begin{cases} 1 \text{ if } t \in T_{(\text{situation } S \text{ is fully instantiated at time } t)} \\ 0 \text{ if } t \notin T_{(\text{situation } S \text{ is not instantiated at time } t)} \end{cases}.$$

without the necessity for a fuzzy indicator function μ_S: $\mathbb{R} \rightarrow [0, 1]$, which would imply that the situation is instantiated at time t only to a certain degree $\mu_S(t)$. This assumption is supported by considering the defining requirement of full mutual awareness from above: broadly established social reception and leave-taking protocols mirror the human tendency to explicitly establish full mutual awareness with respect to the existence status of a social situation. Doubts concerning the establishment status of a social situation are usually actively counteracted by actions that clarify this status by either signalling denial or acceptance of interaction.

Reasonable degrees of precision in this model are ultimately determined by human social cognition patterns and can be determined by the assumption that the time duration of the protocols for establishing or dissolving a social situation are small against the duration of the interaction. Furthermore, from the perspective of applied informatics, the required degree of precision is determined by the respective application from a whole range of useful applications of this model, which, e.g., include approaches for social life logging (Groh et al., 2010b) where a precision of ± 10 sec for T and ± 2 m for a boundary element of X can be assumed sufficient.

With respect to applications for social context data, especially instantiations of social situations, various possibilities exist. While general life logging applications record multimedia data, social life logging applications may restrict the logged data to social situations and/or use social situations as retrieval keys in life-logs. Other applications include approaches for using social situations (with T possibly in the past) as sources and targets of communication acts (Groh et al., 2010b). Here, social situations can, e.g., be used as message meta-data, for example as elements of automated explanation, if the respective message reaches a recipient indirectly. More generally speaking, they allow the recipient to efficiently establish the (social) context of the communication act. Other than that, social situations may also be used to establish privacy by restricting access to certain data to persons that, e.g., have recently been in a social situation with the owner of that data, etc. This allows for definitions of access rights which are socially, temporally, and spatially relating these three forms of locality in a meaningful way.

In general, social situations as instantiations of social networks or manifestations on *small temporal and spatial scales* may potentially be used for all services where groups and other sub-graphs of *long-term* social networks play a role such as information filtering, etc.

3 Detecting social situations from interaction geometry

Social situations may be detected using various social signals (Vinciarelli et al., 2009) derived from audio signals, video signals, measurements of position and orientation of persons, etc. In general, techniques from SSP may be used to derive higher abstraction level social contexts (such as social situations like discussed above) from time-series of lower abstraction level social signals.

It can be assumed that *combining evidence* from processing many such low-level signal sources will lead to a more complete view with respect to higher levels of social context. In this contribution, we restrict ourselves to deriving social situations from measurements of the *geometry* of social interaction, in particular interpersonal distances and relative body orientation by means of shoulder lines.

As already laid out in Section 1, our final concept is that mobile devices measure relative body orientation and interpersonal distances, preferably *without using any infrastructure* (such as IR beacons, GPS or the like) and thus just by using their own onboard sensors, and that our quantitative model, which we developed based on data acquired in an experiment *with* IR beacon and camera infrastructure, is used accordingly.

The results are intended to be combined with higher level social context model instances derived from other sources such as, e.g., audio signals. Especially in crowded contexts, where audio-based techniques may not be easily applicable due to background noise and Bluetooth encounters are too coarsely grained, using interaction geometry as social signals for deducing social context may be particularly valuable.

3.1 Related work on social situations and human geometry of interaction

With respect to the definition of the concept of social situations, several contributions may be identified especially in the socio-psychology literature. Eysenck, Arnold and Meili define the concept as follows:

> "A general term for the field of reference (stimuli, objects, fellow men, groups, values, etc.) […] of a person acting in society […] the Social Situation may be defined by three categories of the data and the manner in which they are linked: (a) the actual data which influences the acting person, (b) the attitudes which are brought into play at the time of the act, and (c) the degree of ego involvement or awareness of the actual data and attributes on the part of acting person." (Argyle et al., 1981)

Another insightful definition comes from Goffman: "By the term Social Situation I shall refer to the full spatial environment anywhere within which an entering becomes a member of the gathering that is (or does then become) present. Situations begin when mutual monitoring occurs and lapse when the next to last person has left" (Argyle et al., 1981). Eysenck et al. refer to higher levels of understanding social situations that go beyond our simple definition of a social situation from above (Section 2) such as attitudes and degrees of ego-involvement. These are subject to further analysis in the realm of SSP using, e.g., audio data so as to characterise and deduce the actors' attitudes. In this contribution, we will solely focus on the *existence* of a fully established social situation without further characterising it. Goffman's definition is more technical and pragmatic, closer to our definition, and uses mutual monitoring as a necessary starting criterion for a situation. He does not demand full mutual awareness of the social interaction although it can be assumed that this full mutual awareness is closely connected to mutual monitoring. He mainly focuses on the spatial connectedness or co-location of the actors, especially in regard of the definition of the end of a situation. In our definition, a situation may also end by loss of full mutual awareness of the existence of a social interaction (e.g., by deliberately turning away from somebody while still maintaining co-location). It starts, however, when this full mutual awareness is established.

Given awareness-based definitions of social situations, geometric signals (such as turning toward each other) can be of great significance for deducing their existence in cases of non-verbal social interaction (such as flirting, etc.) where audio signals would not work.

With respect to the relevance of interaction geometry for the evaluation of social situations, the model of interpersonal zones by Hall (1959) (see also, Sommer, 1959; Mehrabian, 1959; Willis, 1966; Sundstrom and Altman, 1976) distinguishes four characteristic interpersonal distance zones in which persons with different socio-emotional closeness are allowed to, and will typically interact with a person: the intimate zone (< 0.5 m) (intimate partners, close relatives), the casual-personal zone (0.5 m–1.2 m) (familiar persons), the socio-consultive zone (1.2 m–2.0 m) (interaction with persons related by formal and impersonal relationships), and the public zone (> 2.0 m) where typically no direct interactions take place anymore. Another study known to us that also uses tracking methods to quantitatively determine these zones in a more precise manner is (Amaoka et al., 2009) yet they use optical tracking methods without beacons and only distinguish a few basic classes of orientation. The model of concentric multivariate Gaussians being used in fact heavily relies on the original zone theory.

There are numerous studies that qualitatively study the influence of extroversion, cultural background, age, sex, and other profile parameters (see Russo, 1967; Heshka and Nelson, 1972; Wolfgang and Weiss, 2002; Sundstrom and Altman, 1976), as well as the influence of other contextual parameters such as the architecture in which the interaction takes place [excerting so-called sociopetal and sociofugal forces on the interacting persons (see Watson, 1968, Watson and Hall, 1969)]. In our study, we neglect these influence factors. It is a matter of additional experiments to quantitatively determine their influence and correspondingly augment our model.

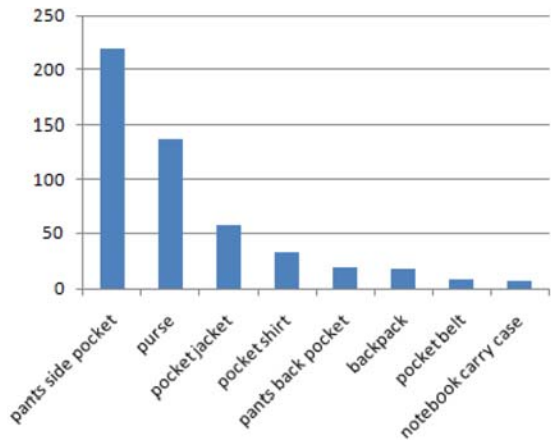
As far as we know, there are rather few models which relate body orientation and social interaction on a broader scale going beyond mere qualitative characterisations of body language. Most studies concentrate on special parameters such as Mehrabian (1959). Here, results show that male persons orient themselves directly opposite to each other in rather hostile situations.

3.2 Mobile devices as sensors for interaction geometry

Modern mobile devices like smart phones etc. usually carry a number of sensors that can be used for autonomous measurement of the parameters of interaction geometry. Measurements of interpersonal distances being socially meaningful in view of our model usually require an accuracy in the order of magnitude of < 10 cm which is not only very hard to achieve with the help of absolute positioning techniques (such as GPS or indoor positioning techniques based on WiFi and the like) but rather often also impossible due to the surrounding architecture, etc. However, cost-effective sensors for direct distance measurement (e.g., based on ultrasound) can reach accuracies of up to < 1 cm (see, e.g., Huang and Huang, 2009). If combined with suitable post-processing, these sensors can also be reliable enough for our purposes. It can be expected that such sensors will be integrated into mobile devices, maybe as part of an integrated multi-sensor component, since accurate measurements of distances to other devices (possibly interpreted as part of the social context) allow for a multitude of MSN applications, for example in social games etc. Aside from interpersonal distance, mobile device orientation may accurately and autonomously be measured using three-axis digital compasses, accelerometers, and/or gyroscopes together with appropriate filtering techniques. Several systems exist that use a combination of sensors to reach accuracies in the range of well below < 3° for each axis of rotation (see, e.g., Gebre-Egziabher et al., 2000; Woodman, 2007). Relative orientation may be easily determined by communicating the orientation measurements between devices. Integration of the required sensors into mobile devices in the near future is supported by various applications such as personal navigation (Wendlandt et al., 2006).

Hence, both interpersonal distance and relative orientation of two or more mobile devices with sufficient accuracy to fit our purposes can be autonomously determined by a large range of such devices either already existing today or in the very near future. In order to deduce the relative distance and orientation of the centres of mass of two persons carrying those devices, we conducted a survey among 500 persons in pedestrian areas of three German cities, investigating the modes of carrying mobile devices. Figure 1 shows the results. What is apparent is that the majority of users (219 users) carried their device in the front pockets of their pants.

Figure 1 Survey on carrying locations of mobile devices (see online version for colours)



It is possible to distinguish this case efficiently from those cases where the device is not carried at all with the help of acceleration sensors and from the cases where the device is not carried close to the body (e.g., in a bag, or purse, or backpack) with the help of temperature and humidity sensors. We assume that, at least applicable to the case when a device is carried in one of the front pockets of the pants, it is possible to deduce the orientation of the torso and correspondingly also calculate the necessary corrections for the determination of the distance between the bodies' centres of masses. The properties of this mapping are currently under investigation.

3.3 Our approach

So far we have discussed and covered the preliminaries of our approach. In view of our goal of the detection of social situations on the basis of data characterising the geometry of interaction we pursued the following detailed programme, the results of which we will discuss in the following chapters:

1. Conduct an experiment which delivers a *precise set* of data on the geometry of social interaction, especially interpersonal distances δd and relative body orientation $\delta\theta$.
2. Since no precise sociological model exists that quantitatively describes the relations between the probability of the existence of a social situation and (δd, $\delta\theta$) in an algorithmically usable way, propose a method to deduce such a generative, precise, and time-independent model from data acquired in experiments such as the one in 1.

(In our case we will use a standard expectation maximisation (EM) approach to arrive at Gaussian mixture models for $p(\delta d, \delta\theta | S^{\otimes/\ominus})$, the probability that a social situation exists or does not exist).

3 Investigate the discriminative power of this quantitative, time-*independent* model in comparison to other standard classifiers. If it turns out to be sufficiently good, suggest the model as means of an a priori model for use in MSN environments where (δd, $\delta\theta$) can be measured and the inference of the existence of social situations is useful as some form of social context.

4 Investigate and compare the discriminative power of an alternative time-*dependent* model [in our case we will use a hidden Markov model (HMM)].

4 Experiment and results of time-independent analysis

The aforementioned experiment measured the interaction geometries of nine persons who mostly did not know each other in advance, interacting in various social situations over a period of 30 minutes in an area of 3 × 3 m located in a larger room. The persons were asked to determine certain unknown biographical facts about the other participants in order to win a prize in a quiz after the 30 minutes interaction time. This task proved to be an excellent stimulus for rich, dynamical of changing and permutation of social situations. The location of each person's throat and orientation of their shoulder lines were measured with a precision of < 1 mm and < 1° using a system of four ceiling and four floor-mounted infrared cameras together with passive infrared beacons (*ARTtrack & DTrack User-Manual V1.24.3*, 2007) at a rate of 60 fps which were later computationally integrated to 6 fps during post-processing. For every time-frame and every pair of persons (i_1, i_2), the relative distances in the x, y-plane $\delta d_{(i_2,i_2)}(t) \in \left[-\sqrt{2}3, +\sqrt{2}3\right]$[meter], and the relative angle of the shoulder lines $\delta\theta_{(i_1,i_2)}(t) \in [-\pi, \pi]$ were computed from the recorded absolute position and orientation. The sign of the distance values was computed regarding the orientation of the first person i_1 in the (thus non-symmetrical) pair, regarding distances to persons i_2 in the back of this person i_1 as having a negative sign. Thus, for every time-frame t, we have $(9)_2 = 72$ pairs of ($\delta d(t)$, $\delta\theta(t)$).

All the same, the experiment was recorded via six ceiling-mounted video cameras. Based on the video-frames as well as an application visualising the absolute orientations and locations of each participant, each time-frame was annotated with respect to which persons were interacting with whom in social situations. Thus, it is known for each pair ($\delta d(t)$, $\delta\theta(t)$) of relative distances and orientations in each time-frame whether they belong to pairs of persons in a social situation $(\delta d(t), \delta\theta(t)) \in S^{\oplus}$, or pairs of persons not in a social situation $(\delta d(t), \delta\theta(t)) \in S^{\ominus}$. The experiment yielded $|S^{\oplus}| = 321{,}307$ (δd, $\delta\theta$) pairs corresponding to 'in a situation', and $|S^{\ominus}| = 398{,}335$ (δd, $\delta\theta$) pairs corresponding to 'not in a situation'.

Figure 2 Slices counting the number of observations relative to the total number of observations in the (δd, $\delta\theta$)-space for the set $S^{\oplus}$ (pairs of persons in a social situation) (see online version for colours)

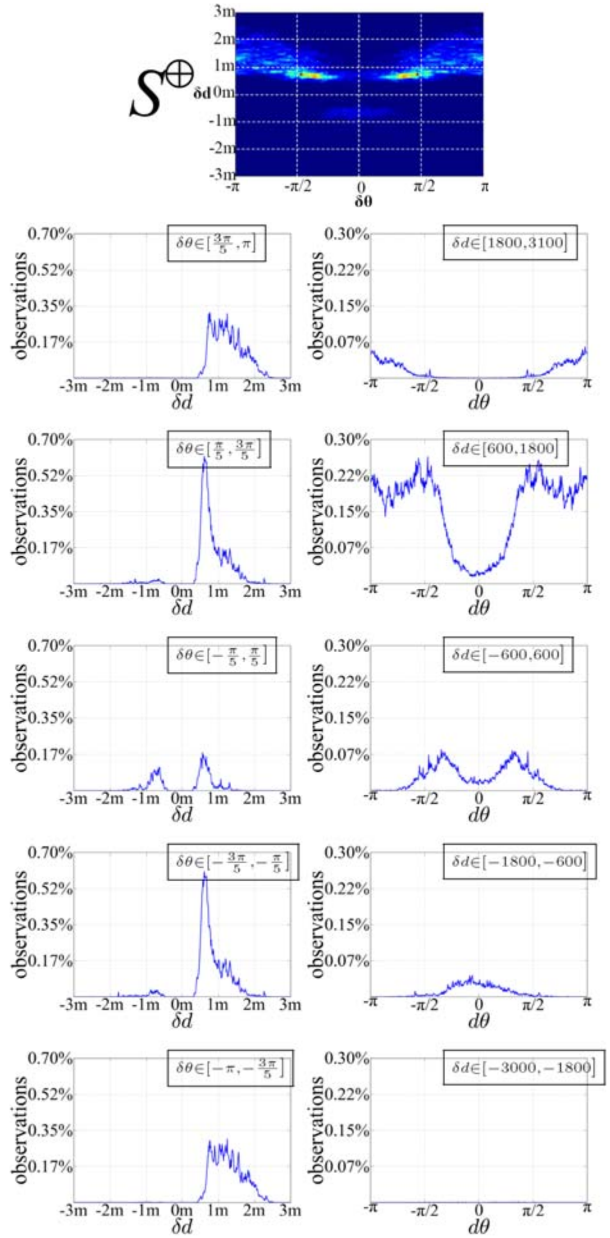
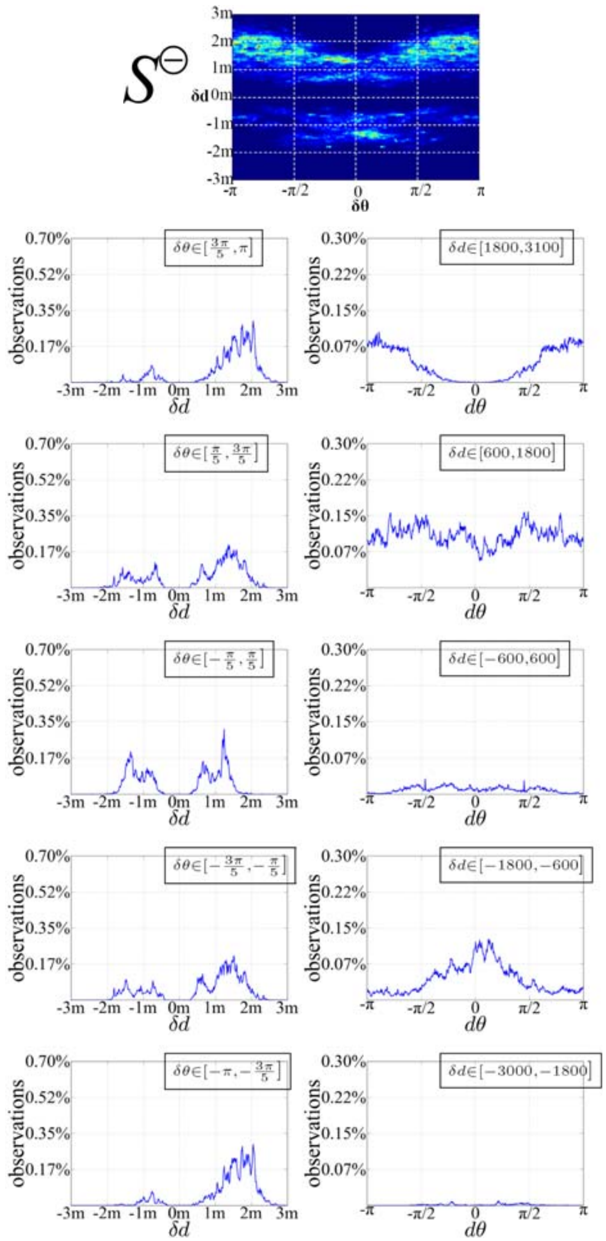


Figure 3 Slices counting the number of observations relative to the total number of observations in the (δd, $\delta \theta$)-space for the set $S^{\ominus}$ (pairs of persons not in a social situation) (see online version for colours)



Figures 2 and 3 show the acquired data from the experiment. For people participating in a social situation (Figure 2) we see that the distribution is sharply peaked at $\approx\left(75\text{ cm}, \pm\frac{\pi}{2}\right)$. This shows that comfortable conversation occurs at relative body angles of around or slightly less than 90 degrees. The distance corresponds to the casual-personal zone of hall. The sliced plots show that this distance is rather strictly used in social situations. Confrontational angles of ≈ 0 are rare. Some counts also occur in the back of people (negative δd).

These correspond to other people that 'apply' for the integration into, e.g., a circular social situation by placing themselves between two social situation members close to their backs. We also see that the intimate zone $\delta d < 50$ cm is strongly respected. This is the case for the pairs not in a social situation (Figure 3) as well. However, the distribution of body angles is much broader and does not show the dip around ≈ 0. Furthermore, the number of (δd, $\delta\theta$) pairs with negative δd (corresponding to situations where one person is facing the back of the other person) occur much more frequently than in the case of $S^{\oplus}$. The involved interpersonal distances are, on an average scale, larger than for $S^{\oplus}$. The absence of counts for larger negative distances can be attributed to the desire of not turning one's back to people, even if not socially interacting with them, at least in this rather confined space of the experiment.

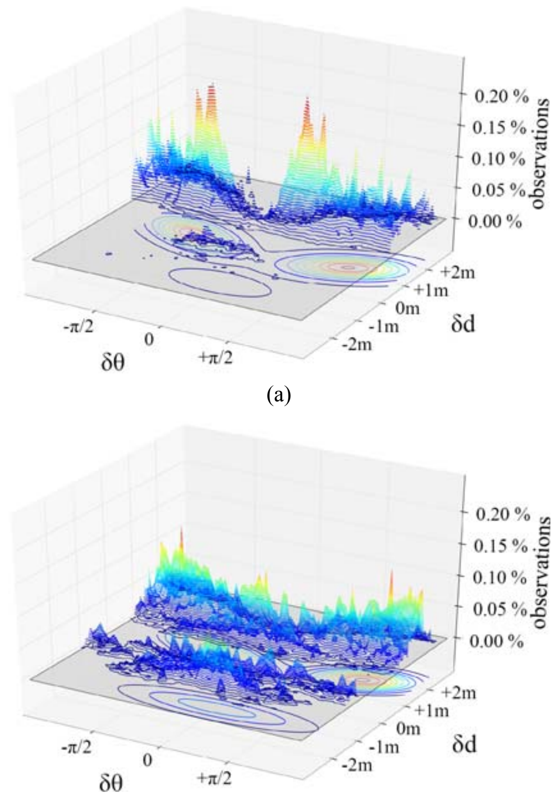
4.1 Time-independent modelling: $p(\delta d, \delta\theta|S^{\oplus/\ominus})$

Aside from the only semi-quantitative models for interpersonal distance (for example, Amaoka et al., 2009), no algorithmically usable model relating (δd, $\delta\theta$) to the probability of an established social situation is known to us. Thus, our first intended contribution was the construction of such a model where we concentrate solely on interpersonal distance δd and relative body orientation $\delta\theta$ of the individuals. We leave out all other meaningful elements of interaction geometry, like dynamic positions and orientations of limbs (from which, e.g., gestures may be inferred), because they are only available via full fledged wearable computing sensors worn on arms, fingers, etc., or via sophisticated vision-based methods. Both scenarios are not compliant with our vision of autonomously measuring interaction geometry in every possible situation without infrastructure (such as ceiling mounted cameras), preferably using standard mobile computing devices such as smart-phones. Relative quantities are used in the model since absolute position- and orientation-measurements are very difficult to realise (Lehmann, 2010). Furthermore, projections of relative distance vectors and relative rotations onto the local x, y-plane were used instead of the full distance vectors and relative rotations (see Groh et al., 2010a for details).

For the set $\tilde{P}$ of N possibly interacting persons, we are thus in principle left to derive a complicated probability distribution depending on the $n(n-1)/2$ interpersonal distances and relative orientations yielding probabilities for each subset of $\tilde{P}$ forming a social situation. This approach is clearly not practical. We therefore focused on a dyadic model $p(\delta d, \delta\theta|S^{\oplus/\ominus})$, specifying the probability of existence or non-existence of a social situation between two given actors. N-ary social situations can then be easily identified with graph clustering techniques using the derived probabilities as edge weights.

Using the experimental data (pairs of persons in a social situation $(\delta d(t), \delta\theta(t)) \in S^{\oplus}$) or pairs of persons not in a social situation $(\delta d(t), \delta\theta(t)) \in S^{\ominus}$), a Gaussian mixture model was inferred using a standard EM-algorithm with three, five, and seven Gaussians yielding time-independent probability distributions $p(\delta d, \delta\theta|S^{\oplus})$ and $p(\delta d, \delta\theta|S^{\ominus})$ for both of the two classes. Figure 4 shows examples of the derived GMMs.

Figure 4 The counts on the 100 × 100 grid for the set, (a) $S^{\oplus}$ and (b) $S^{\ominus}$ (see online version for colours)



(a)

(b)

Note: On the lower level, contour lines for the resulting GMM (three Gaussians) are drawn.

Source: Groh et al. (2010a)

We compared the discriminative power of the derived GMMs using them as classifiers and comparing them with other standard classifiers trained on our two-class ($S^{\oplus}$ and $S^{\ominus}$) dataset, yielding the results shown in Table 1. One can see that these classifiers can achieve approximately 75% accuracy on average on our dataset (neglecting the standard Naive Bayes approach, which performed less successful). The results of the respective classifiers with more than three Gaussians did not exceed those with three Gaussians and were in the same range.

Table 1 Time-independent classification results

Classifier	*Accuracy*	*Implementation*
Gaussian mixture model (three Gaussians)	74.50%	Own Impl. (EM) [see, e.g., Bishop, (2007), p.435]
Naive Bayes	65.45%	Weka 3.2.6 (Hall et al., 2009; John and Langley, 1995)
Naive Bayes (w. kernel estimator)	73.10%	Weka 3.2.6 (Hall et al., 2009; John and Langley, 1995)
Naive Bayes (w. supervised discretisation)	72.96%	Weka 3.2.6 (Hall et al., 2009; John and Langley, 1995; Fayyad and Irani, 1993)
Support vector machine	77.81%	Weka 3.2.6 + LIBSVM (Hall et al., 2009; Chang and Lin, 2001)

Note: Each result obtained with ten-fold cross-validation.

Source: Groh et al. (2010a)

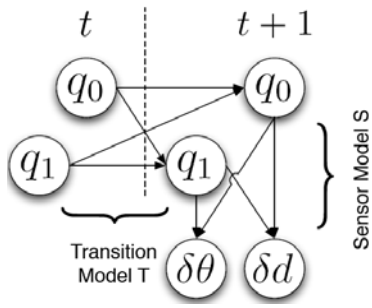
From the comparison and the absolute values of accuracy we can thus infer that our time-independent model for the prediction of dyadic social interaction based on interpersonal distance and relative body orientation is a reasonable model to be used by MSN applications as evidence for or against the existence of a social situation. Further experiments incorporating cultural aspects, gender, or body-height differences into the model may further improve its expressiveness. Another reasonably simple improvement would incorporate the overall interpersonal density $\overline{\delta d}$ of all persons in social reach [< 30 m (Bradner and Mark, 2002)] as a measure of crowdedness, resulting in a model $p\left(\delta d, \delta\theta, \overline{\delta d}\right)$. This is reasonable since the social expressiveness of δd, $\delta\theta$ is different in a crowded subway-setting from a large open-space setting.

5 Using time-dependency

The time-dependency of ($\delta d(t)$, $\delta\theta(t)$) for a given pair of persons can be exploited via modelling with dynamic Bayesian networks (DBN). One of the simplest cases of DBN is a first order HMM. We model our case with the HMM shown in Figure 5. In this simple model, the hidden states q_0 and q_1 correspond to being ($S^{\oplus}$) or not being ($S^{\ominus}$) in a in a social situation, respectively. The sensor model reflects the interpersonal distance and relative body orientation measurements. State and observation sequences are acquired for each pair of individuals starting from the first timeframe corresponding to a set of persons being within a social situation and ending with the last such frame. In order to get a qualitative overview of the state transition probabilities as well as the initial state probabilities to be expected during ten-fold cross validation, we also computed the quasi-optimal transition model $p_{ij} = p(q_i|q_j)$ by straightforward analysis of the whole set of sequences by means of counting the respective events belonging to either one of the four cases $q_0 \rightarrow q_1$, $q_0 \rightarrow q_0$, $q_1 \rightarrow q_1$, and $q_1 \rightarrow q_0$, as well as the initial state probabilities $\pi(0) = (p(q_0(0)), p(q_1(0)))$ by means of computation and normalisation of how many times each state occurred at the beginning of any sequence. From this qualitative overview of the whole dataset, we get $\pi(0) = (0.028, 0.972)$ and

$$p_{ij} = \begin{pmatrix} 0.998182 & 0.001818 \\ 0.001542 & 0.998458 \end{pmatrix}.$$

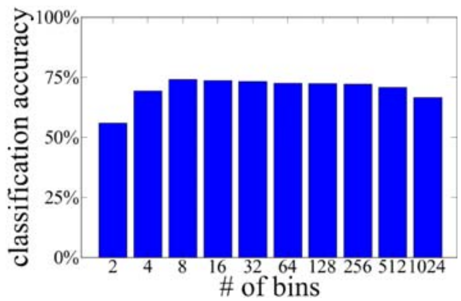
Figure 5 First order HMM



Using ten-fold cross validation we compared the performance of two HMMs, one of which made use of a GMM sensor model whereas the other one relied on discretised observations and/or probability densities. In both cases, we used the Viterbi algorithm so as to be able to determine the most likely state sequence according to Rabiner (1989) using numerical improvements suggested in Mann (2006). We then compared the estimated state sequence with our annotated (true) state sequence and counted the correctly classified states.

For the HMM/GMM approach, a three Gaussians GMM, obtained by neglecting time-dependency (see previous section), was used for $p(\delta\theta, \delta d|q)$. This DBN lead to a mean accuracy of 74.21% where, noticeably, varying initial probabilities of $\pi(0)$ have no significant influence on the classifier's overall accuracy. So, similar results (≈ 74.2%) occur for $\pi(0) = (j, 1 - j)$ where $j \in [0, 1]$. This, however, is in correspondence with expectations (Russell and Norvig, 2003). As stated above, we also used a discretised probability function in comparison to the GMM for the sensor model. For this, the space $\delta d(t) \in \left[-\sqrt{2}3, +\sqrt{2}3\right] \times \delta\theta(t) \in [-\pi, \pi]$ clearpage of relative distances and orientations was partitioned in a symmetric grid of $n \times n$ bins and the numbers of occurrences of observations corresponding to each bin were counted for $S^{\oplus}$ (or q_0, respectively) and $S^{\ominus}$ (or q_1, respectively), and ultimately normalised. From Figure 6, we see that in case of very noisy/coarse views (few bins) the accuracy drops to what we would expect by a random state choice (e.g., a coin flip). Surprisingly, however, we also see that at the bottom line only rather few bins suffice to reach reasonable accuracies. For $n = 8$, accuracy is maximal and only slightly drops for larger n. Summa summarum this effect apparently defines a characteristic level of granularity which distinguishes interaction geometries in social vs. non-social situations.

Figure 6 Varying the bin-size of the sensor model calculation (see online version for colours)



Lastly, we also implemented a standard expectation maximisation algorithm for HMMs (Rabiner, 1989) used for re-estimation respective learning of the transition model as well as the sensor model. Running this algorithm on our data, however, most of the times resulted in reaching only local extreme of the expectation value of the likelihood of the target function which the EM-algorithm is supposed to maximise. Correspondingly, computed accuracies did not suit the hitherto established general accuracy of the DBN.

6 Conclusions

We demonstrated a novel method for the quantitative, geometrical characterisation of social interactions on small spatial and temporal scales based on interpersonal distance and relative body orientation.

It was our goal to employ geometric behavioural cues to classify social interactions by means of a simple model of either existing or non-existing social situations. For this purpose, we conducted an experiment and manually annotated the social situations that occurred, and were thus able to classify the ($\delta\theta$, δd) pairs of interpersonal distance and relative body orientation of pairs of persons in each time-frame as either belonging to or not belonging to a social situation, respectively. The accurate measurements of ($\delta\theta$, δd) were won with the help of IR tracking which is able to reveal interesting sociological micro-phenomena.

As a quantitative, probabilistic, time-independent model of social interaction geometry that is easy to interpret, we inferred a Gaussian mixture model for the probability of an existing/non-existing social situation on ($\delta\theta$, δd) using EM-learning. We also trained and tested various other classifiers and showed that interpersonal distance in conjunction with relative body orientation can indeed be used to provide evidence for or against an existing social situation with sufficient accuracy. In this comparison, the GMM model's performance as a classifier was among the best.

This model may thus be used in MSN services as a quantitative a-priori model for inferring the existence of social situations as a simple form of social context, using ($\delta\theta$, δd) measurements from mobile devices. It is suggested to be combined with evidence from other social signal sources such as audio.

The resulting IR-based experimental method may be used by social scientists to further refine the resulting quantitative models, e.g., with respect to cultural differences, age, varying surroundings, etc.

Furthermore, we see that the exploitation of time-dependency does not substantially improve the predictive performance of time-independent classifiers in our case. Given a series of observations ($\delta d(t)$, $\delta\theta(t)$), the analysis of the most likely sequence of hidden states does not provide a better result than the classification of the observations independent of time. Note that we are not interested in predicting the next state at a given time t, but rather in the analysis of whole sequences of observations in order to determine or deliver evidence for or against the existence of a social situation as well as its temporal duration. Also, the standard HMM EM learning approach did not deliver any advantage over the time-independent classifiers. The trained GMMs for $p(\delta\theta, \delta d|S^{\oplus})$ and $p(\delta\theta, \delta d|S^{\ominus})$ allow for a sufficiently accurate analysis of (sequences of) observations of human interaction geometry in order to derive social situation models.

One of our next steps with respect to the further development of this method will be working on the relation between device orientation- and relative distance-measurements as well as the related orientation and distances of the bodies that the devices are attached to.

References

Amaoka, T., Laga, H., Saito, S. and Nakajima, M. (2009) 'Personal space modeling for human-computer interaction', *ICEC '09: Proceedings of the 8th International Conference on Entertainment Computing*, pp.60–72.

Argyle, M., Furnham, A. and Graham, J.A. (1981) *Social Situations*, Press Syndicate of the University of Cambridge.

ARTtrack & DTrack User-Manual V1.24.3 (2007) Advanced Realtime Tracking GmbH, Germany.

Bishop, C.M. (2007) *Pattern Recognition and Machine Learning (Information Science and Statistics)*, 1st ed. 2006, corr. 2nd printing ed., Springer.

Bradner, E. and Mark, G. (2002) 'Why distance matters: Effects on cooperation, persuation and deception', November 2002, *Proc. CSCW02*, New Orleans.

Chang, C-c. and Lin, C-J. (2001) *Libsvm: A Library for Support Vector Machines*.

Fayyad, U. and Irani, K. (1993) 'Multi-interval discretization of continuous-valued attributes for classification learning', *Proceedings of the International Joint Conference on Uncertainty in AI*, pp.1022–1027.

Gebre-Egziabher, D., Elkaim, G.H., Powell, J.D. and Parkinson, B.W. (2000) 'A gyro-free quaternion-based attitude determination system suitable for implementation using low cost sensors', *Position Location and Navigation Symposium, IEEE 2000*, pp.185–192.

Groh, G., Lehmann, A., Reimers, J., Friess, R. and Schwarz, L. (2010a) 'Detecting social situations from interaction geometry', *Proc. IEEE SocialCom 2010*.

Groh, G., Lehmann, A., Wang, T., Huber, S. and Hammerl, F. (2010b) 'Applications for social situation models', *Proc. Int'l Conf. Wireless Applications and Computing Conference*, Freiburg, Germany.

Hall, E.T. (1959) *The Silent Language*, Doubleday.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P. and Witten, I.H. (2009) 'The weka data mining software: an update', *SIGKDD Explorations*, Vol. 11, No. 1.

Heshka, S. and Nelson, Y. (1972) 'Interpersonal speaking distance as a function of age, sex and relationship', *Sociometry*, Vol. 35, No. 4, pp.491–498.

Huang, K-N. and Huang, Y-P. (2009) 'Multiple-frequency ultrasonic distance measurement using direct digital frequency synthesizers', *Sensors and Actuators A: Physical*, Vol. 149, No. 1, pp.42–50.

John, G.H. and Langley, P. (1995) 'Estimating continuous distributions in Bayesian classifiers', *Eleventh Conference on Uncertainty in Artificial Intelligence*, San Mateo, Morgan Kaufmann, pp.338–345.

Lehmann, A. (2010) 'Towards mobile location- and orientation-based detection of social situations', Diploma thesis, TU-München, WS 2009/2010; Supervisor: Georg Groh.

Lewis, K., Kaufman, J., Gonzalez, M., Wimmer, A. and Christakis, N. (2008) 'Tastes, ties, and time: a new social network dataset using facebook.com', *Social Networks*, Vol. 30, No. 4, pp.330–342.

Mann, T.P. (2006) 'Numerically stable hidden Markov model implementation', *An HMM Scaling Tutorial*.

Mehrabian, A. (1959) *Nonverbal Communication*, Aldine Transaction.

Newman, M., Barabasi, A-L. and Watts, D.J. (2006) *The Structure and Dynamics of Networks*, Princeton University Press.

Pentland, A. and Eagle, N. (2009) 'Eigen behaviors: identifying structure in routine', *Behavioral Ecology and Sociobiology*, Vol. 63, p.7.

Pentland, A., Eagle, N. and Lazer, D. (2009) 'Inferring social network structure using mobile phone data', *Proceedings of the National Academy of Sciences (PNAS)*, Vol. 106, No. 36, pp.15274–15278.

Rabiner, L.R. (1989) 'A tutorial on hidden Markov models and selected applications in speech recognition', *Proceedings of the IEEE*, Vol. 77, No. 2, pp.257–286.

Russell, S. and Norvig, P. (2003) *Artificial Intelligence: A Modern Approach*, 2nd ed., Prentice-Hall, Englewood Cliffs, NJ.

Russo, N. (1967) 'Connotation of seating arrangements', *Cornell Journal of Social Relations*, Vol. 2, No. 1, pp.37–44.

Sommer, R. (1959) *Personal Space*, Prentice-Hall, Inc.

Sundstrom, E. and Altman, I. (1976) 'Interpersonal relationships and personal space: research review and theoretical model', *Human Ecology*, Vol. 4, No. 1, pp.47–67.

Vinciarelli, A., Pantic, M. and Bourlard, H. (2009) 'Social signal processing: survey of an emerging domain', *Image and Vision Computing*, Vol. 27, No. 12, pp.1743–1759.

Watson, O.M. (1968) 'Proxemic behavior: a cross-cultural study', PhD dissertation, University of Colorado, Boulder, Colorado.

Watson, O.M. and Hall, E.T. (1969) 'On proxemic research', *Current Anthropology*, April–June, Vol. 10, Nos. 2/3, pp.222–224.

Wendlandt, K., Khider, M., Angermann, M. and Robertson, P. (2006) 'Continuous location and direction estimation with multiple sensors using particle filtering', *2006 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems*, pp.92–97.

Willis, F. (1966) 'Initial speaking distance as a function of the speaker's relationship', *Psychonomic Science*, Vol. 5, pp.221–222.

Wolfgang, A. and Weiss, D.S. (2002) 'A locus of control and social distance comparison of Canadian and West Indian-born students', *International Journal of Intercultural Relations*, Vol. 4, Nos. 3–4, 1980, pp.295–305.

Woodman, O.J. (2007) *An Introduction to Inertial Navigation*, Tech-Report 696, University of Cambridge, CL.