Space Speaks – Towards Socially and Personality Aware Visual Surveillance

Gloria Zen, Bruno Lepri, Elisa Ricci, Oswald Lanz
Fondazione Bruno Kessler FBK-irst
Via Sommarive 18, 38123 Povo (TN), Italy
[gzen,lepri,eliricci,lanz]@fbk.eu

ABSTRACT

There is a complex unwritten code which regulates human interactions. In this paper we present a camera based monitoring system that explores the relationship between proxemics, visual attention and personality traits during interaction. People's relative positions an head poses are extracted with a multi-target tracking algorithm and used to (i) estimate, under the 'thin slices' hypothesis, the level of extroversion and neuroticism of a person, to (ii) learn a model of people interactive behavior that improves, once integrated in the algorithm, the accuracy of the tracking estimates, and to (iii) steer a set of active cameras to the subject found to exhibit the most peculiar interaction pattern. We report on experimental results in a natural scenario where people engage in a party.

Categories and Subject Descriptors

J.4 [**Computer Applications**]: Social and Behavioral Sciences—*Sociology*

General Terms

Algorithms, Experimentation, Human Factors, Measurement

Keywords

Human behavior modeling based on observations, Guided vision based on high-level reasoning, Multi-camera tracking

1. INTRODUCTION

In the last years we assisted to a growing interest in providing artificial systems with social intelligence [23]. It is well known that a relevant skill of human social intelligence is the non-verbal communication [3]: people communicate interest, power, affect, emotions, by means of cues such as facial expressions, gestures, postures, voice pitch, and so on. As pointed in [7], computer science (in particular, computer vision and speech processing), has developed various approaches and techniques for the automatic analysis of various non-verbal behaviors, such as vocal behavior (pitch, energy, and everything else than words in speech), facial expressions, gaze, head pose, gestures, body postures, and so on. In contrast, few efforts have been done in order to recognize automatically how people use, manage and share their physical space. Here we investigate on this, focusing on the specific type of non-verbal behavior which is the management of the spatial relationships and of the personal space, i.e. on proxemics [11].

In sum, our work addresses the following three points: (i) identify significant proxemics cues able to discriminate between two personality traits, Extraversion and Neuroticism; (ii) utilize the collected data to build up specific behavioral models for subjects with different individual characteristics, e.g. personality. These models are then used as 'a priori' knowledge for improving the performance of a particle filter-based tracking system; (iii) utilize the information about people spatial behaviors recognized online to control active cameras in order to collect near-field data of the subject of interest that can be used to extract detailed information on head pose and facial expression.

MPVA'10, October 29, 2010, Firenze, Italy.

1.1 Related Work

textbfProxemic Theory. The social meaning of the space use has been extensively studied in human sciences since the '60s. The term proxemics was introduced by the anthropologist Edward T. Hall [11]. This term describes four regions of interpersonal distance: public, social, personal and intimate. These regions are modulated by cultural, gender, social status differences, and by other individual characteristics, such as personality traits. Moreover, this theory postulate that different distances of approach convey different social meanings and interpersonal relationships, forming a set of implicit rules of distance. Therefore, when these rules are broken, a person might feel shocked, stressed, and so on.

Modeling Personality. Humans have the tendency to understand, explain and predict other humans' behavior in terms of stable properties that are variously assorted on the basis of the observation of everyday behavior. In folk-psychological practice, the personality is assessed along several dimensions: we are used to talk about an individual as being (non-)open-minded, (dis-)organized, too much/little focused on herself, etc. Several existing theories have formalized this folk-psychological practice to model personality by means of multi-factorial models, whereby an individual's personality is described in terms of a number of more fundamental dimensions known as traits. A well known example of a multi-factorial model is the Big Five [13] which owes its name to the five traits it takes as constitutive of people's personality: Extraversion, Neuroticism, Agreeableness, Conscientiousness, Openness to Experience. Recently, some works have started exploring the automatic recognition of personality [2, 17, 21, 16]. The general approach aims at isolating promising behavioral correlations of the targeted traits for personality classification or regression. We follow this trend.

Modeling Proxemics. Some studies over the spatial occupation and use have been conducted recently in the robotic and virtual environment communities. In robotics the goal is to increase the comfort level of humans building socially-aware interacting robots [6]. It is proven that people show a major confidence with a humanoid showing social awareness, e.g. a robot able to approach a group of people at appropriate distance instead of a robot endowed only with an avoidance function which simply permits to avoid the collision with people. Similar goals are followed by researchers who design avatars [10]. However few works in this fields have investigated explicitly the link among personality traits and proxemics [26].

In the computer vision community and in particular in the contest of multiple people tracking, few recent works [19] proposed to learn behavioral models in order to better describe the dynamics of people interactions. However, none of them focused on the aspect that different people may need different models depending on their personality and their specific way of acting.

1.2 Experimental Setup

We create a social scenario (Fig.1.a; an informal cocktail party) where a group of people (6-7 subjects) interact naturally. In particular, in our experiments we recorded two sessions of 20 and 30 min each for a total of 13 targets. We consider people moving freely rather than a meeting with subjects in a seating arrangement as in previous works [16]. Our choice is based on the hypothesis that the standing inter-personal distance is a more sensitive measure of proxemics behavior then seated interpersonal distance. Social psychology studies support our hypothesis [12].

The room where the party is held is equipped with a set of fixed camera (far field data) and a set of Pan Tilt Zoom (PTZ) camera for active tracking (near-field data). Figure 1.b depicts the map of our laboratory with the four fixed cameras (represented in a rectangular shape) set at each corner of the room and the three PTZ cameras (round shape) disposed in a way to cover the whole scene. The dimensions of the room are 4.8×6.0 meters.

The combination of fixed cameras with PTZ cameras working in a cooperative way is adopted in many surveillance systems, e.g. for biometric recognition tasks [4]. In our scenario the two types of camera are settled in a master-slave configuration: we process the streams of the four fixed cameras (master) to obtain in real-time an estimate of the positions of all the targets in the scene and steer the PTZ cameras (slave) to obtain high resolution images of target of interest (see Sec. 2.3).

2. AUTOMATIC PREDICTION OF EXTRAVERSION AND NEUROTICISM

In this work, we exploit proxemics behaviors enacted in a social interactive setting to predict two dimensions of the Big Five, Extraversion (sociable, playful vs reserved, shy) and Neuroticism (calm, unemotional vs insecure, anxious). The choice of these traits is due to the fact that Extraversion is the one that shows up more clearly in, and has the greater impact on social behavior. Moreover, some studies found some significant correlations among extraversion dimension and use of personal space: in particular, extrovert people seem requiring smaller interpersonal spacing [18, 25] Regarding the Neuroticism dimension, some social psychology studies showed that high neuroticism scorers tend to significantly prefer larger social distances than low neuroticism scorers [9] Extraversion and Neuroticism are measured by means of the sub-scales of the Italian version of the Big Five Marker Scale [20]. The scales are assigned to all the participants before the experimental sessions. In our subjects sample, the highest and the lowest Neuroticism scores are 57 and 34 respectively. The mean is 42 and the standard deviation is 6.52. For Extraversion trait, the highest score is 62 and the lowest score is 29. The mean and the standard deviation values are 43.54 and 9.84 respectively.

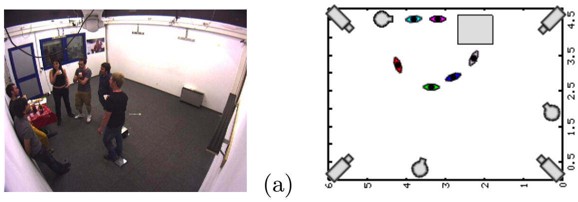
(a) (b)

Figure 1: The proposed scenario.

2.1 Feature Extraction

We process the streams of four fixed cameras with a person tracker to extract position and head pose of each target (see Sec. 3), and extract the following proxemic features:

Minimum distance. It is the distance that one target is keeping from the closest target. This feature is indicative of the most intimate type of relation that he/she is holding with respect to the group. This feature may be a cue for Extraversion, as introverts usually tend to stay at a more intimate distance to a few people (attachment) [5], while extroverts tend to stay at a less intimate distance but to interact with more people.

Velocity. It is computed considering the variation of the position of the target every 2 second. This feature may be indicative of the comfort degree of a person in the social situation. An introvert person may feel uncomfortable in a group interaction and as a reaction he/she will tend to isolate from the group or get closer to a person of confidence. An extrovert instead will tend mostly to move freely in the room to interact with different subjects [25].

Number of intimate, personal and social relationships. We estimate the number of persons with whom the target is holding different kind of relationships. In details, we adopt the following thresholds: 0.46 m for intimate, 0.76 m and 1.20 m for personal (close and far phase) and 2.10 m for public (close phase) relations [11]. We also take into account head orientation to decide if two subjects are interacting.

The features described are computed for each target at each frame. Taking inspiration from works on rapid cognition and first impression formation [1], we adopt a 'thin slice' approach for the personality detection, whereby people levels on given traits are discovered on the basis of the observation of short sequences of expressive behavior. In particular, we divide the acquisition session into temporal windows of 30 sec. duration. In this way for every feature we obtain vector data of 450 frames length. Then, for each vector data we compute the average and the standard deviation values. As a result, we obtain for each target N vectors of $2M$ dimension, where N is the number of thin slices into which the acquisition session has been divided, and M is the number of features considered.

2.2 Experimental Design and Results

We model the automatic recognition of Extraversion and Neuroticism traits as two classification problems. Extraversion and Neuroticism scores are dichotomized along the median value (41 for Extraversion and 41 for Neuroticism), yielding two classes: 6 people with Low Extraversion and 7 people with High Extraversion and 5 people with Low Neuroticism and 8 people with High Neuroticism. Those labels are added to the feature vectors described above for each considered time slice. The SVM classification algorithm with a second-order polynomial kernel is used [8]. The classifiers predicts personality traits by considering the behavior of a subject in a 30 seconds temporal window (similarly to a psychologist asked to recognize personality traits of a patient by observing 'thin slices' of behavior). We run several experiments using different feature vectors corresponding to different experimental conditions: all features (ALL); minimum distance (DIST); relationships-based features (REL), i.e. the combination of the number of intimate, personal-close, personal-far and social relationships; intimate (INT), personal-close (PERS-CLOSE), personal-far (PERS-FAR) and social (SOC) relationships; velocity features (VEL). A leave-one-user out procedure was used in the experiment at each fold, all the slices relative to twelve users are used for training, and those of the left-out user for testing.

Concerning the prediction of both Extraversion and Neuroticism traits (see Table 1), the best performance is obtained using a vector encompassing all the visual features. For Extroversion a relevant result is that the performance we get with only the number of different spatial relationships of a given subject is quite close to the one obtained with all the visual features. In this case, we reach an accuracy equal to 0.63 and a F-measure equal to 0.62. Also, observing spatial relationships features an interesting aspect is the good performance obtained only with the social relationship features. In predicting Neuroticism the highest accuracy (0.69) is obtained with the minimum distance features. Overall, the velocity-based features show the lowest predictive power.

2.3 Personality Aware Cameras

The automatic detection of personality traits can be used in many ways. For example in this paper we show how we used a 'social logic' to determine the target of interest which is followed by PTZ cameras. We employ the features SOC computed for the personality trait classification in the first half of the meeting to define a score of people extroversion and we decide to follow the target with the lowest score (the most introvert person). This could be used in several ways, for example to automatically monitor situations where this target is embarrassed and is not interacting with others.

Table 1: Performance on Personality Classification

	extraversion		neuroticism	
	Acc	F-meas	Acc	F-meas
ALL	**0.66**	**0.66**	**0.75**	**0.69**
DIST	0.54	0.38	0.69	0.52
REL	0.63	0.62	0.69	0.52
INT	0.53	0.37	0.67	0.50
PERS-CLOSE	0.56	0.46	0.56	0.51
PERS-FAR.	0.56	0.48	0.64	0.39
SOC	0.59	0.59	0.64	0.39
VEL	0.52	0.46	0.64	0.39

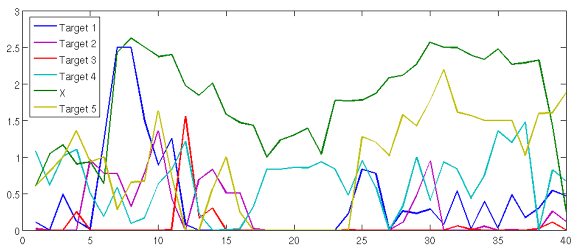

Figure 2: Average number of targets at social distance. Features associated to 40 'thin slices'.

While various kinds of analysis (*e.g.* gesture and emotion recognition) can be done having at disposal high resolution images of the face or the upper-body, here we focused on improving the head pose estimate of the target of interest. In fact estimating the tilt is very challenging from the low resolution (512×386 pixel, 90deg horizontal aperture) images of fixed far-field cameras while it can be done more easily form the output of a PTZ. For computing the head pose (pan and tilt) from images of PTZ cameras we used the approach proposed in [22], based on particle filters and Histograms of Oriented Gradient for pose classification. Since this approach is suited to monocular cameras we employed the images from a single PTZ, the one where the face of the target is more frontal (*i.e.* the one where the face is first found by a frontal face detector [24]). Every time the PTZ moves the particle filter algorithm is reinitialized. By using information about the calibration of PTZ cameras, the position of the targets and the relative pan and tilt estimated we can reconstruct the target focus of attention. An example of the output of the proposed system is shown in Fig. 6. In the sequence it is evident how with PTZ cameras we can detect when the target is looking down or when it is interacting with (looking towards) target 2 or target 3.

Fig. 2 depicts some of the SOC features used for Extroversion classification in case of the target depicted in Fig. 6 (target X). In particular, Fig. 2 shows that X tends to keep at least 2 targets at a social distance: in fact an introvert person feels more confident when she interacts with the people she is more close friend with. Note that while in this experiment we simply compute an Extroversion score based on SOC features, in the future we intend to collect a larger training set for personality classification and to use the margin of the SVM as an Extroversion score.

Table 2: **Classification scores and parameters of the interaction model in Sec. 3.1**

	extroversion	neuroticism	mode	spread	weight
t1	62	45	0.80	0.15	0.39
t2	56	44	0.66	0.09	0.29
t3	55	57	0.68	0.09	0.25
t4	40	41	0.82	0.08	0.25
t5	37	39	0.64	0.07	0.09
t6	31	43	0.62	0.07	0.25
t7	29	36	0.66	0.09	0.25

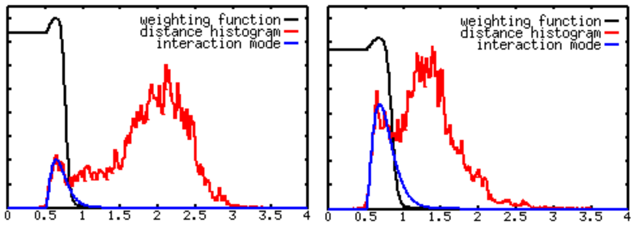

Figure 3: Distance-to-target histograms of two targets (t5 and t3 in Table 2)

3. LEARNING INTERACTION PATTERNS FOR IMPROVED VISUAL TRACKING

To extract physically observable patterns from multi camera recordings that are relevant to personality and social behavior studies (proximity and visual attention in this paper) we build on previous work on multi people tracking and head pose estimation [14, 15]. As our main contribution in this section we show how to embed a model of interactive behavior (Sec. 3.2) that emerges in the tracker's output over the two scenarios, and how to calibrate it to each individual (Sec. 3.1) to capture inter-personal variations that correlate with personality traits. We show that such model, once calibrated to each subject, improves tracking performance while encoding a number of plausible social behaviors.

We adopt a Bayesian state estimation framework to jointly track the spatial position and head orientation of people from image sequences (a state has thus a position and head pose component as in [15]). This involves the design of (i) a transition density $p(\mathbf{x}_t|\mathbf{x}_{t-1})$ that expresses how likely a target found to be in state $\mathbf{x}_{t-1}$ at time $t-1$ will end up in state $\mathbf{x}_t$ at time t, and (ii) a likelihood function $l(\mathbf{z}_t|\mathbf{x}_t)$ that expresses how likely hypothesis $\mathbf{x}_t$ would produce observation $\mathbf{z}_t$. Both models can be tailored to the application context (e.g. they can capture physical constraints of motion in the former, physics of signal formation in the latter) which may explain, at least partially, the popularity of the Particle Filter (a non parametric implementation over this framework) in the signal processing community.

In a multi target scenario $p(\cdot)$ and $l(\cdot)$ can be designed over the joint state space (where $\mathbf{x}$ has now one state component per target) to account for interactions among them. The challenge for practical applications is then to scale the estimation algorithm to the number of targets without incurring in the curse of dimensionality which is inherent in the formulation. In [14] it is shown how this is possible with $p(\cdot)$ modeling spatial exclusion among targets and $l(\cdot)$ implementing a model of the visual occlusion process. We used a Particle Filter implementation of it (see `tev.fbk.eu/smartrack`) to produce the features in Sec. 2, and elaborate on $p(\cdot)$ in Sec. 3.2.

3.1 Modeling the Interaction Space

To discover on an empirical basis whether there is emergent social behavior in the tracking space, we computed, for each tracked person in the two sequences, the histogram of its distance to the targets in its visual attention field: if $\mathbf{p}_t^k$ and $\mathbf{o}_t^k$ are the estimated position and head orientation of target k at time t we compute its Euclidean distance $\|\mathbf{p}_t^m - \mathbf{p}_t^k\|$ to each other target m and accumulate the value $\mathcal{N}(\mathbf{o}_t^m - \mathbf{o}_t^k|\sigma_o)$ in the corresponding histogram bin. Here $\mathcal{N}$ is a zero-mean Gaussian on the angular offset with variance σ_o (we used $\sigma_o = 40$deg in this paper) which accounts for the limited field of view and the uncertainty of the orientation estimates. The histograms of one representative actor per acquired session are shown in Fig. 3.1. All histograms show a prominent peak at a distance (<1m) on which people tend to interact, thus supporting related studies in human sciences [11]. As in [10] we describe the social mode in the histogram as a Gamma distribution and develop an algorithm to fit such model to the data at hand.

To do so we minimize a cost function over weight w, mode μ and spread ξ of $\Gamma_n(d|\mu,\xi) = (x^{n-1}e^{-x})/(n-1)!$ resembling a Gamma-like distribution of order n with $x = (d-\mu)/\xi + n - 1$ (we chose $n = 3$ for our purpose; b is the bin index, Δ the bin size, $H[b]$ is the histogram value in b):

$$\sum_b K(b,\mu,\xi) \cdot \|H[b] - w \cdot \Gamma_n(b\Delta,\mu,\xi)\|.$$

Here K is a normalized weighting function computed from Γ_n and its cumulative density c_n, which is introduced to inhibit the contributions of the non-interacting entries in the tail of H:

$$K(b,\mu,\xi) \propto \Gamma_n(b,\mu,\xi) + \mathcal{N}(c_n^\alpha(b,\mu,\xi)|\sigma_\xi).$$

Fig. 3.1 shows Γ_n with the parameters that minimize the proposed cost function on the two distance histograms for $\alpha = 4, \sigma_\xi = 0.1$, and their corresponding weighting functions. From the values in Tab. 2 is evident that those parameters correlate with personality traits: w encodes the amount of interaction, μ the usual distance to a partner during comunication, and ξ to the stability of μ with the various partners. In addition, the gap between μ and the histogram residual could be used as an index indicating whether a person is involved in group conversations.

3.2 Embedding Γ for Improved Tracking

In [14] it is shown how to effectively embed a pairwise Markov Random Field (MRF) potential in the multi target Bayesian tracker encoding the physical constraint that two targets cannot occupy the same part of the space at the same time, i.e. the spatial exclusion principle. Here we reformulate the MRF to account for the interaction prior learnt for target k with the method presented in the previous section as follows:

$$p(\mathbf{x}_t^k, \mathbf{x}_t) = \Gamma_3(\|\mathbf{p}_t^k - \mathbf{p}_t\| \,|\, \mu^k, \xi^k) \cdot \mathcal{N}(\mathbf{o}_t^k - \mathbf{o}_t|\sigma_o).$$

Fig. 4 shows the resulting MRFs generated by the five targets located at the green dots under various orientations $\mathbf{o}_t^k$, which are computed with the Belief Propagation algorithm

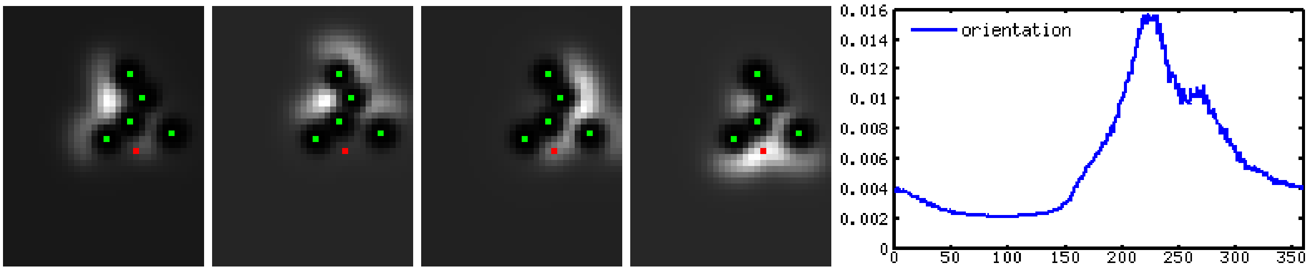

Figure 4: Interaction fields generated by five targets (the green dots) for various orientations $\mathbf{o}_t$ = $0, 90, 180, 270\,\text{deg}$ and field intensity perceived by a probe target (the red dot) as a function of $\mathbf{o}_t$ (right plot).

in [14]. The field perceived by the probe target (the red dot) shows a peak around $\mathbf{o}_t = 270$ where he sees a subject at the learnt social distance. This orientation is thus prioritized during tracking, and in situations where the observation likelihood l is ambiguous (e.g. due to noisy appearance model, clutter, occlusion) the MRF prior can compensate for the noise in the likelihood. This corrective behavior has been observed in the experiments reported in Sec. 3.3.

This MRF prior simulates a number of social behaviors that are observed also in natural interactions. If a target k is in the influence range of $p(\mathbf{x}_t^k, \mathbf{x}_t^m)$ generated by another target m at $\mathbf{x}_t^m$ (i.e. m is close and k looks at it) it perceives an attractive force emitted by m, thus simulating the intention of k to go closer to m in order to interact with it. Also, the MRF tends to prioritize orientation hypotheses towards targets that are located at the learnt social distance from k. On the other hand, if a person invades the intimate space of k the MRF acts as a repulsion force if k looks at it. This means that the tracker tends to either move k away from the invader or to prioritize orientations under which the invader is no longer seen, thus simulating intimidation. If a person is instead not seen by a target its behavior (i.e. the motion model of the tracker) is not influenced by it.

3.3 Experimental Results

To demonstrate the improvements that can be acheived with the interaction prior introduced in Sec. 3.2 we have reprocessed the second sequence and inspected the quality of the new tracker output on the head orientation estimates of the target with a noisy appearance model (the orange target in Fig. 5). The advantage of validating the improvements in terms of accuracy of head orientation rather than position is two-fold: (i) social interaction induces visual attention while proximity alone does not (if a person is close to another she may still don't care about) thus this output is more relevant to the scope of this paper, and (ii) the orientation estimates are more affected by noise in the target's appearance model than the position component and therefore it is easier to validate the impact of the prior.

Fig. 5 shows the real time tracker output with (top row) and without (bottom row) interaction prior on some frames of interest of the second sequence. In the first two images the presence of a target at a distance around the mode of Γ_3 (we calibrated it with the method in Sec. 3.1 on the data) biases its orientation estimate towards the target, thus compensating for the ambiguity in the visual likelihood (see Fig. 4 for the interaction fields generated on the output of the first image). In the third image the effect of the interaction field is low but still sufficient to correct the estimate. In the fourth image the subject is close to but not watching another target; the estimated orientation is still accurate despite the attraction by the other target. This shows that likelihood and interaction model are correctly balanced. The last image shows instead an example where the prior increases the error in the head orientation. Overall, however, we verified qualitatively that the accuracy of head orientation estimates of the orange target improves significantly over the major part of the sequence.

4. CONCLUSIONS

In this paper we have investigated on the relationship between proxemics, visual attention and personality traits during interaction: we designed an experiment to estimate the level of extroversion and neuroticism of people engaging in a party with a SVM classifier using related features extracted by means of video based people tracking and head orientation estimation. We also showed how to integrate, in form of a prior, a model of interactive behavior that emerges in the trackers output to boost its accuracy. Future work aims at better validating these contributions on more data collected in relevant scenarios, at a tighter integration between low-level (tracking and head pose estimation) and high-level tasks (personality trait classification), and at further elaborating on the self-adaptive behavior of the system introduced here by means of active camera steering policies.

5. ACKNOWLEDGMENTS

This work has been supported by PUMALAB, a FBK-irst Joint Research Project sponsored by the Autonomous Province of Trento. Bruno Lepri's research is also partially funded by Marie Curie - COFUND - 7th framework fellowship. We thank Francesco Tobia (FBK-irst) for providing software code for the control of active cameras.

6. REFERENCES

[1] N. Ambady and R. Rosenthal. Thin slices of expressive behaviors as predictors of interpersonal consequences: A meta-analysis. *Psychological Bulletin*, 111:256–274, 1992.

[2] S. Argamon, S. Dhawle, M. Koppel, and J. Pennbaker. Lexical predictors of personality type. *Interface and the Classification Society of North America*, 2005.

[3] M. Argyle. *Bodily Communication (2nd ed.)*. Madison: International Universities Press, 1988.

[4] A. D. Bimbo, F. Dini, G. Lisanti, and F. Pernici. Exploiting distinctive visual landmark maps in pan-tilt-zoom camera networks. *CVIU*, 114, 2010.

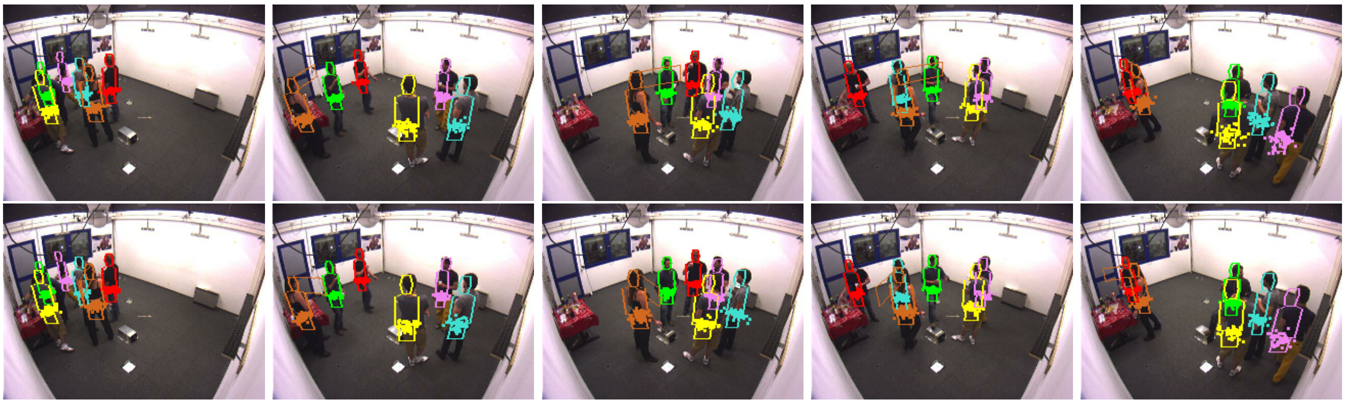

Figure 5: Real time estimates computed with (top row) and without (bottom row) interaction prior.

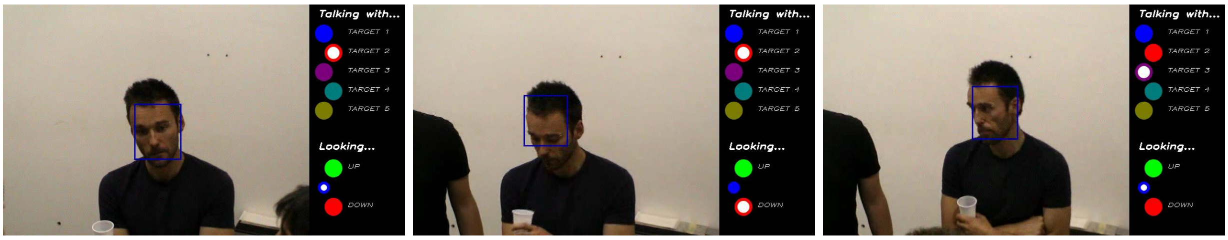

Figure 6: Example of the output obtained processing the streams of PTZ cameras.

[5] J. Bowlby. *Attachment and Loss*. Hogarth Press, 1969.

[6] A. G. Brooks and R. C. Arkin. Behavioral overlays for non-verbal communication expression on a humanoid robot. *Autonomous Robots*, 22(1):55–74, 2007.

[7] M. Cristani, V.Murino, and A. Vinciarelli. Socially intelligent surveillance and monitoring: Analysing social dimensions of physical space. *IEEE SISM*, 2010.

[8] N. Cristianini and J. Shawe-Taylor. *An Introduction to Support Vector Machines and other kernel-based learning methods*. Cambridge University Press, 2000.

[9] S. DeJulio and K. Duffy. Neuroticism and proxemic behavior. *Perceptual and Motor Skills*, 45:51–55, 1977.

[10] D. Friedman, A. Steed, and M. Slater. *Spatial Social Behavior in Second Life*. Springer, 2007.

[11] E. T. Hall. *The hidden dimension*. Anchor Books, 1963.

[12] W. Ickinger and S. Morris. Psychological characteristics and interpersonal distance. *www.sharktown.com/proxemics/intro.html*, 2001.

[13] O. P. John and S. Srivastava. The big five trait taxonomy: History, measurement, and theoretical perspectives. In L. A. Pervin and O. P. John, editors, *Handbook of Personality: Theory and Research*, pages 102–138. Guilford Press, New York, 1999.

[14] O. Lanz. Approximate bayesian multibody tracking. *IEEE TPAMI*, September 2006.

[15] O. Lanz and R. Brunelli. Dynamic head location and pose from video. *IEEE MFI*, September 2006.

[16] B. Lepri, N. Mana, A. Cappelletti, F. Pianesi, and M. Zancanaro. Modeling personality of participants during group interaction. *UMAP*, 2009.

[17] F. Mairesse and M. Walker. Automatic recognition of personality in conversation. *HLT-NAACL*, 2006.

[18] D. Pedersen. Correlates of behavioral personal space. *Psychological Reports*, 32(3):828–830, 1973.

[19] S. Pellegrini, A. Ess, K. Schindler, and L. van Gool. You'll never walk alone: Modeling social behavior for multi-target tracking. *ICCV*, 2009.

[20] M. Perugini and L. DiBlas. Analyzing personality-related adjectives from an eticemic perspective: the big five marker scale (bfms) and the italian ab5c taxonomy. *In De Raad, B. and Perugini, M. (Eds.), Big Five Assessment, Hogrefe und Huber Publishers. GĄöttingen, 281-304*, 2002.

[21] F. Pianesi, N. Mana, A. Cappelletti, B. Lepri, and M. Zancanaro. Multimodal recognition of personality traits in social interactions. *ICMI*, 2008.

[22] E. Ricci and J.-M. Odobez. Learning large margin likelihoods for realtime head pose tracking. *IEEE ICIP*, 2009.

[23] A. Vinciarelli, M. Pantic, and H. Bourlard. Social signal processing: Survey of an emerging domain. *Image and Vision Computing*, 2009.

[24] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. *CVPR*, 1, 2001.

[25] J. L. Williams. Personal space and its relation to extraversion introversion. *Canadian Journal of Behavioural Science*, 3(2):156–160, 1971.

[26] F. Yamaoka, T. Kanda, H. Ishiguro, and N. Hagita. A model of proximity control for information-presenting robots. *IEEE Trans. on Robotics*, 26:187–195, 2010.