Challenges of Human Behavior Understanding

Albert Ali Salah[1], Theo Gevers[1], Nicu Sebe[2], and Alessandro Vinciarelli[3]

[1] Institute of Informatics, University of Amsterdam,
Amsterdam, The Netherlands
{a.a.salah,th.gevers}@uva.nl
[2] Dept. of Information Engineering and Computer Science
University of Trento
Trento, Italy
sebe@disi.unitn.it
[3] Department of Computing Science
University of Glasgow
Glasgow, Scotland
vincia@dcs.gla.ac.uk

Abstract. Recent advances in pattern recognition has allowed computer scientists and psychologists to jointly address automatic analysis of of human behavior via computers. The Workshop on Human Behavior Understanding at the International Conference on Pattern Recognition explores a number of different aspects and open questions in this field, and demonstrates the multi-disciplinary nature of this research area. In this brief summary, we give an overview of the Workshop and discuss the main research challenges.

1 Introduction

Domains where human behavior understanding is a crucial need (e.g., human-computer interaction, affective computing and social signal processing) rely on advanced pattern recognition techniques to automatically interpret complex behavioral patterns generated when humans interact with machines or with others. This is a difficult problem where many issues are still open, including the joint modeling of behavioral cues taking place at different time scales, the inherent uncertainty of machine detectable evidences of human behavior, the mutual influence of people involved in interactions, the presence of long term dependencies in observations extracted from human behavior, and the important role of dynamics in human behavior understanding.

The target topics of the Human Behavior Understanding (HBU) Workshop reflect some of the old and new questions in this domain:

- Social behavior analysis & modeling, multimodal behavior patterns
- Temporal patterns
- Facial, gestural and voice-based affect recognition
- Sign-language recognition
- Human motion analysis

- Pattern recognition applied to novel sensors
- Pattern discovery in personal sensor networks, reality mining
- Smart environments
- Human-computer interaction
- Benchmarking studies on novel databases
- New feature selection and extraction methods
- Mathematical description and integration of contextual information
- Behavioral biometrics

Some of these topics have been actively researched for a long time, like the analysis of face, voice, and bodily signals, yet these are taken up to new levels of difficulty by relaxing some of the simplifying constraints. Research focuses now on more natural settings with uncontrolled conditions, real-time operation requirements and interaction dynamics. Furthermore, domain-specific semantic information is drawn into the picture as we move from generic techniques to specific applications. This re-focusing is partly done by introducing richer taxonomies and increasing volumes of multi-modal data.

This chapter is meant as a summary of the issues covered in the Workshop, and subsequently, it is neither a balanced treatment of the domain, nor an extensive survey of all open questions. In Section 2 we distinguish between different spatio-temporal scales of human behavior. In its largest scale, patterns are discovered in the collective behavior of masses. Section 3 deals with the most heavily researched area of behavior analysis, pertaining to visual sensors. The visual patterns are usually shorter in their temporal extent, but we see that the temporal aspects are gaining importance in this modality as well. Section 4 focuses on social signal processing, which adds social semantics to signal processing. Finally, we conclude in Section 5 and give pointers to further reading material on some of the key issues.

2 Temporal Levels of Behaviors

It is possible to look at behaviors at different temporal levels. The microscopic behaviors happen in a short time frame, and have to be analyzed as such. A blink of the eye, a rapid hand gesture, a yawn can all be seen as microscopic behaviors. On the other hand, the movement of masses over longer temporal and spatial scales also contains recognizable patterns, and these can be said to exist in a macroscopical scale. The continuous range that stretches between these extremes contains many problems that are approached with a host of pattern recognition methods.

One of the areas in which different temporal scales come together is ambient intelligence. In [1], daily activities of people living in a sensed environment (like eating, using a computer, reading, watching television, etc.) are analyzed. The smaller time frame in which the activity is actually performed and the larger time frame which is composed of longer segments prone to contain the activity are

combined in a hierarchical framework. Here, SVM classifiers predict locally ongoing activities, and Conditional Random Fields are used to refine the prediction by estimating time segments of global activities.

While most of the papers submitted to the HBU Workshop dealt with behavior dynamics on a microscopical scale, we admitted work on both types of patterns. In [2], daily activity patterns of individuals are analyzed using large collections of mobile phone data. These patterns reveal that activity patterns within a given area of work strongly resemble each other. As more data are available from the population, it becomes possible to create *reality mining* applications and discover behavior patterns [3].

The diversity in the behavior-related patterns suggests the possibility of using diverse sensors in their assessment. Modern mobile phones are equipped with a host of sensors, thus allowing unprecedented opportunities of personal data collection. In [4], body-worn miniature inertial and magnetic sensors were used the collect data for activity classification. Each sensor unit used in the study comprises a triaxial gyroscope, a triaxial accelerometer, and a triaxial magnetometer. Using multiple types of sensors potentially increases the cost of a given system, but offers great increase in robustness. Especially in the context of ambient intelligence, multimodal analysis of behavior opens up new venues of applications such as behavioral biometrics and automated care for the elderly. In [1], infrared and object motion sensors are used in conjunction to classify daily activities in a sensor-equipped home setting.

The human behavior is not restricted to physical actions and behaviors. Many people now have a presence on the Web, and exhibit social networking behavior that is becoming ever more relevant. In the keynote talk of Ramesh Jain [5], the *macroscopic behavior* of masses on the Web is investigated.

3 Visual Action Recognition

Vision is currently the most heavily used sensory modality in the analysis of human behavior. Visual human action recognition concerns the detection and tracking of people, and more generally, the understanding of human behaviors from image sequences involving humans [6, 7]. Automated vision-based analysis of human actions finds many applications in surveillance, ambient assisted living, concept-based video retrieval, automatic sports video annotation and summarization, customer behavior analysis for marketing and gaming. So far a scalable and widely applicable system for this purpose remains elusive.

3.1 Tracking the Body

Tracking of humans and human behavior inherently involves estimation of body pose, locations and movements of body parts, interaction with objects, and sometimes also gaze estimation. While estimating the pose means determining the location and the orientation for an object, humans manifest more complex pose aspects. Pose estimation can be a post processing step in a tracking algorithm, or

it can be an active part of the tracking process. Recent approaches to tracking favor particle filtering based methods, as these can maintain multiple probabilistic hypotheses with respect to a parametrized body posture at any given time [8].

Pose estimation can be approached with different methods. A model of the human shape can be used in constraining the interpretation of the pose. In model-free pose estimation, the pose can be represented as a set of feature points, as a combination of simple shapes, or with stick-figures, which connect points with lines. [9] introduced the motion history image to represent human body movement. [10] recently extended this paradigm to propose a spatio-temporal silhouette representation, called silhouette energy image to characterize motion and shape properties. The challenges in this problem are dealing with both indoor and outdoor conditions, real-time operation, low level features extraction, motion analysis, and saliency computation, multi-camera fusion, among others.

While some methods aim at tracking and labelling the body parts in 2D, others try to map 2D sequences of image observations into 3D pose representations. In some cases, equipment is available to obtain depth information from the scene. Two papers in this collection describe such systems. In [11], a trinocular camera system is used for this purpose. In [12] input from multiple cameras are fused to determine the gesture trajectories of humans performing signs. In both approaches, hidden Markov model (HMM) is the classifier of choice to effect temporal classification.

The use of an explicit model of a person's kinematics, shape, and appearance in an analysis-by-synthesis framework is a widely investigated approach to human pose estimation from video [13]. In these approaches the model is used to synthesize an appearance from the current parametrization of the model, which is compared to the actual appearance. The discrepancy is minimized by changing the parameters appropriately, to a point where the model is able to synthesize a close match of the appearance. At that point, the converged parameters can be directly used to represent the pose of the person. The direct model pose estimation can de subdivided in multiple view 3D pose estimation [14] and monocular 3D pose estimation. For a detailed overview on methods of pose estimation, see [15].

Once the people are tracked and spatio-temporal features are extracted, action classification can take place. In recent work, static SIFT features were shown to perform well for many detection tasks, while histogram of oriented gradient (HoG) and histogram of optical flow (HoF) features were successful for action recognition [16]. Yet the temporal dimension is taken into account in only a few low-level descriptors: A 3D Harris operator that describes spatio-temporal interest points was described in [17]. In [18] human actions were modelled as three-dimensional shapes induced by the silhouettes in the space-time volume. [19] presented a spatio-temporal interest point detector, and analyzed a number of cuboid descriptors for action recognition. Most approaches opt for temporal integration of spatial descriptors, using different forms of dynamic Bayesian networks [20].

3.2 The Context

Automatic detection of an action may involve complex spatiotemporal and semantic reasoning. To constrain this problem, contextual cues are used. For the integration of contextual information, we need to define the context properly. It can mean several things:

1. Geometrical scene properties, i.e. 3D composition of the scene. The scene can be classified by using visual cues into one of the possible scene classes.
2. Scene type, i.e. the content and context of the scene. This property can be derived via texture analysis, and indicate indoor vs. outdoor, or common locations like street, football field, etc.
3. Objects in the scene: The objects of interaction provide valuable contextual cues.
4. Persons in the scene: The number of persons in the scene, and their visual features can provide contextual cues.
5. Temporal context: Detection of other prior actions will influence the detection of related actions. Interaction semantics would be in this category, and hence this probably the richest source of contextual information.

The application setting mostly determines what kind of contextual cue will be used in each setting. There are marked differences between content-based retrieval of actions from movies and detection of actions from one or more surveillance cameras. In the latter the cameras are mostly static, providing poorer context, as opposed to constructed narratives of films. There are also lots of self-occlusions, and the scale of action is typically much smaller. In a surveillance setting real-time operation is usually essential. On the other hand, videos can be processed in an offline fashion, has often higher resolution and scale for people performing the actions, engineered camera perspectives, moving and zooming camera angles. Furthermore, multimedia solutions often come to rescue where vision-based processing fails; subtitles (text), speech (transcript), social tags and file name associations are used to label videos. In one of the keynotes of the HBU Workshop, Ivan Laptev discusses several supervised and weakly-supervised approaches for action recognition in movies [21].

3.3 Benchmarking

Increased interest in human action/activity recognition in recent years resulted in several benchmarking and database annotation efforts. In this domain, we observe that the application focus shifts from the recognition of simple, generic human actions to the analysis of activities in a context and/or interactions between humans. The CAVIAR[4] Dataset for instance contains video recordings of settings like city center surveillance and analysis of customer behavior in a shopping mall. Activities include people walking alone, meeting with others, window shopping, entering and exiting shops, and leaving a package in a public

[4] http://homepages. inf.ed.ac.uk/rbf/CAVIAR/

place. The latest editions of Performance Evaluation of Tracking and Surveillance (PETS)[5] propose similar challenges such as crowd analysis and tracking of individuals in a crowd.

Recognition of unusual or dangerous actions is important for public infrastructure surveillance, and the existence of CCTV cameras and surveillance by security staff makes automated activity recognition a natural extension in these settings. The automated action recognition technology can support the existing personnel, and reduce the burden of inspection by making potentially interesting actions salient, as well as blocking obviously irrelevant information. The TRECVID[6] challenge for 2010 states that "Detecting human behaviors efficiently in vast amounts surveillance video, both retrospectively and in realtime, is fundamental technology for a variety of higher-level applications of critical Importance to public safety and security." For this purpose, a large dataset collected from Gatwick airport is made available in the 2010 challenge.

The recent SDHA (Semantic Descriptions of Human Actions) Challenge[7], organized as a satellite event to ICPR'2010, provides three public databases for various action recognition settings [22]. Some of the labelled actions have complicated semantic associations (e.g., stalking, flirting), which makes the dataset challenging. Further datasets for this type of research are detailed in [23]. In the present collection, [24] provides a detailed survey of evaluation protocols on the KTH action database, which is one of the most studied among these [25].

Apart from individual efforts, a number of previous projects tackled human action recognition from different perspectives. To give a few illustrative examples, the ADVISOR EC-IST project (Annotated Digital Video for Intelligent Surveillance and Optimised Retrieval) aimed at using computer vision algorithms to detect unusual human behavior and to use the developed technologies to improve the effectiveness of existing security operators[8]. The tackled behaviors were blocking, fighting, jumping over barriers, vandalism and overcrowding, all in a public transport scenario. The CAVIAR EC-IST project (Context Aware Vision using Image-based Active Recognition) targeted local image descriptors combined with task, scene, function and object contextual knowledge to improve image-based recognition processes.

3.4 Challenges

The primary challenge in this area is the great range of actions and gestures produced by humans even in relatively restricted domains. Humans use contextual cues extensively to recognize small but discriminative differences. Consider for instance the gestures of an orchestra conductor, which simultaneously specify the rhythm, the style and conductor's interpretation of the piece. A subtle facial expression or posture can convey that the players should play more legato, or

[5] http://www.cvg.rdg.ac.uk/PETS2009/
[6] http://trecvid.nist.gov/
[7] http://cvrc.ece.utexas.edu/SDHA2010/index.html
[8] http://www-sop.inria.fr/orion/ADVISOR/default.html

the energy in the overall composure of the conductor may suggest forte. The expression of the rhythm can temporarily shift from one hand to the other, as the conductor overlaps the expression of several cue. The gestures will be highly idiosyncratic, yet the orchestra generally knows how to adapt to the conductor. It is the representation of assumed knowledge (i.e. priors) in combination with real-time, adaptive and multi-modal information processing on both sides that makes the problem really difficult for a computer. This problem is investigated under the rubric of social signals.

4 Social Signals

A great class of human behaviors pertain to expressing and recognizing social signals, which have communicative and interactive aspects. Even in the absence of other people, socially formed habits manifest themselves in different ways like facial expressions and idiosyncratic gestures. Studying social interactions and developing automated ways of classifying human social behavior from all kinds of sensors is becoming important not only for natural human-computer interaction, but also for all kinds of applications we have mentioned in the previous section.

4.1 Taxonomies

In [26], a taxonomy is introduced for the analysis of social signals. The verbal signals that are usually direct manifestations of communicative intent are accompanied by *behavioral cues* that serve to convey information about the emotion, personality, status, dominance, regulation and rapport in a given social context. These cues reside in different modalities, like the physical appearance, gesture, posture, facial expression, focus of attention, vocal behavior (e.g., prosody and silences), and even the spatial arrangement of participants during an interaction.

In one of the major efforts directed for social signal processing, the SSPNet[9] project focuses on the analysis of political debates as a rich source of behavioral signals. The project defines the core questions of social signal processing as follows:

1. Is it possible to detect automatically nonverbal behavioral cues in data captured with sensors like microphones and cameras?
2. Is it possible to automatically infer attitudes from nonverbal behavioral cues detected through sensors like microphones and cameras?
3. Is it possible to synthesize nonverbal behavioral cues conveying desired relational attitudes for embodiment of social behaviors in artificial agents, robots or other manufacts?

These questions are generic, in the sense that they apply to many domains of behaviors equally. However, the computational aspects (for both analysis and synthesis) depend largely on the application domain, making the problem difficult or very difficult in each case.

[9] http://sspnet.eu/

4.2 Domains for Analysis

The automatic analysis of behavioral cues in a particular domain requires and fosters a decomposition of all activities in that domain. Each such domain has its own challenges and rewards. In the present collection, a number of such domains are investigated. Poggi and D'Errico present a scheme for the annotation of signals of dominance in political debates [27], which are behaviorally rich and interactive settings. In [28] a taxonomy of communicative and non-communicative behaviors of teachers towards their pupils is proposed, which can be used for guiding the development of an automatic analysis tool for a classroom. Such a tool would be a very valuable teaching aid.

In [29], Lepri et al. investigate prediction of personality traits from behavioral cues. A well-known taxonomy proposes five traits as constitutive of people's personality: Extraversion, Emotional Stability, Agreeableness, Conscientiousness, Openness to Experience [30]. In [29], the extraversion-introversion dimension is analyzed using four acoustic features (Conversational Activity, Emphasis, Influence and Mimicry) and one visual feature. In [31], the emotion content of the speech is analyzed and the resulting system is usable as a virtual speech coach for improving public speaking skills. The authors use a discriminative approach, and train SVM classifiers for each type of emotion.

The idiosyncratic variations constitute a major challenge of social signals. In successful dyadic interactions, human subjects exhibit a remarkable adaptivity to these variations. In the study of Özkan and Morency, backchannel feedback in dyadic interactions is analyzed [32]. A feature selection approach is proposed to automatically discover the subset of features relevant to this specific application.

4.3 Face Analysis

Affect-related signals constitute a large portion of nonverbal behavioral cues, and facial expressions are among the most extensively studied signals in this category. These result from movements of the facial muscles as the face changes in response to a person's internal emotional states, intentions, or social communications. Psychological studies suggest that facial expressions, as the main mode for nonverbal communication, play a vital role in human face-to-face communication [33, 34]. Computer recognition of facial expressions has many important applications in intelligent human-computer interaction, computer animation, surveillance and security, medical diagnosis, law enforcement, and awareness systems. Therefore, automatic facial expression analysis (from video or images) has received much attention in last two decades [35, 36]. Face analysis in conjunction with body and head pose orientation can reveal the attention focus of a person, which can also be a very useful cue in putting a behavior in its proper context [37].

The challenges of face analysis in the present context are finding the correct level of description, feature extraction and representation, spontaneous and posed expression classification, head pose and gaze direction estimation. The Workshop has received a number of submissions on these areas. In [38] spatiotemporal DCT features are used in a boosted classification framework for the

classification of face and head gestures. Face detection, tracking and analysis are much more difficult in real-life settings, as the resolution of the face area and the pose show great variations. In [39], a probabilistic approach is proposed for a multi-camera setup to track and recognize faces across difficult conditions.

5 Concluding Remarks

Understanding affective and social behavior of humans with computational tools is receiving increased interest. The present volume demonstrates that pattern recognition is an essential component of research in this area. Researchers seek to analyze patterns emanating from interactions between humans, as well as between humans and computers or smart systems, with the goal of designing more responsive and natural interfaces and applications.

Automatic classification of human behavior involves understanding of bodily motion [7, 15, 23], gestures and signs [40], analysis of facial expressions [36], and interpretation of affective signals [35]. On a higher level, these signals are integrated with the contextual properties of an application domain. Social signal processing deals with interactions between humans [26]. It integrates verbal cues with rich sets of non-verbal behavioral cues to deeply analyze social interactions. Ambient intelligence deals with smarter environments [41]. In ambient environments, the living space is equipped with many sensors that observe the behavior of humans and with many actuators to make the space responsive to changes in these behaviors. As a more focused application, perceptual user interfaces are concerned with more responsive human-computer interfaces [42, 43]. In this domain the computer is given the capacity to detect behavioral changes of its user. The analysis of spatio-temporal dynamics of human actions, observed through different sensory modalities, allows inference and customization on many levels [44, 45].

The submissions for the HBU Workshop demonstrate that the range of behaviors in the proposed applications is rapidly expanding. The set of behaviors under study includes concepts that are hard to describe precisely and mathematically. However, recent pattern recognition approaches developed for multimedia retrieval have shown us that a precise description is sometimes not necessary for the recognition of a concept. If an informative feature extraction step is combined with a powerful pattern classifier and a training set with sufficiently rich variation, it may be possible to learn appropriate descriptors for even the most challenging concepts. Subsequently, it is obvious that human behavior understanding will continue to be a very active research area in the near future, and will be instrumental in providing the tools for building more interactive systems.

References

1. Nicolini, C., Lepri, B., Teso, S., Passerini, A.: From on-going to complete activity recognition exploiting related activities. In Salah, A.A., Gevers, T., Sebe, N., Vinciarelli, A., eds.: HBU 2010. LNCS, vol. 6219, Springer, Heidelberg (2010) 26–37

2. Phithakkitnukoon, S., Horanont, T., Di Lorenzo, G., Shibasaki, R., Ratti, C.: Activity-aware map: Identifying human daily activity pattern using mobile phone data. In Salah, A.A., Gevers, T., Sebe, N., Vinciarelli, A., eds.: HBU 2010. LNCS, vol. 6219, Springer, Heidelberg (2010) 14–25
3. Eagle, N., Pentland, A.: Reality mining: sensing complex social systems. Personal and Ubiquitous Computing **10**(4) (2006) 255–268
4. Altun, K., Barshan, B.: Human activity recognition using inertial/magnetic sensor units. In Salah, A.A., Gevers, T., Sebe, N., Vinciarelli, A., eds.: HBU 2010. LNCS, vol. 6219, Springer, Heidelberg (2010) 38–51
5. Jain, R.: Understanding macroscopic human behavior. In Salah, A.A., Gevers, T., Sebe, N., Vinciarelli, A., eds.: HBU 2010. LNCS, vol. 6219, Springer, Heidelberg (2010) 13
6. Gavrila, D.: The Visual Analysis of Human Movement: A Survey. Computer Vision and Image Understanding **73**(1) (1999) 82–98
7. Wang, L., Hu, W., Tan, T.: Recent developments in human motion analysis. Pattern recognition **36**(3) (2003) 585–601
8. Isard, M., Blake, A.: Condensationconditional density propagation for visual tracking. International Journal of Computer Vision **29**(1) (1998) 5–28
9. Bobick, A., Davis, J.: The recognition of human movement using temporal templates. IEEE Transactions on Pattern Analysis and Machine Intelligence **23**(3) (2001) 257–267
10. Ahmad, M., Lee, S.W.: Variable silhouette energy image representations for recognizing human actions. Image and Vision Computing **28**(5) (2010) 814 – 824
11. Hahn, M., Quronfuleh, F., Woehler, C., Kummert, F.: 3d mean-shift tracking and recognition of working actions. In Salah, A.A., Gevers, T., Sebe, N., Vinciarelli, A., eds.: HBU 2010. LNCS, vol. 6219, Springer, Heidelberg (2010) 99–110
12. Richarz, J., Fink, G.A.: Feature representations for the recognition of 3D emblematic gestures. In Salah, A.A., Gevers, T., Sebe, N., Vinciarelli, A., eds.: HBU 2010. LNCS, vol. 6219, Springer, Heidelberg (2010) 111–122
13. Ali, S., Shah, M.: Human action recognition in videos using kinematic features and multiple instance learning. IEEE Transactions on Pattern Analysis and Machine Intelligence **32**(2) (2010) 288–303
14. Kehl, R., Gool, L.: Markerless tracking of complex human motions from multiple views. Computer Vision and Image Understanding **104**(2-3) (2006) 190–209
15. Moeslund, T., Hilton, A., Krüger, V.: A survey of advances in vision-based human motion capture and analysis. Computer vision and image understanding **104**(2-3) (2006) 90–126
16. Laptev, I., Marszalek, M., Schmid, C., Rozenfeld, B.: Learning realistic human actions from movies. In: IEEE Conference on Computer Vision and Pattern Recognition. (2008) 1–8
17. Laptev, I.: On space-time interest points. International Journal of Computer Vision **64**(2) (2005) 107–123
18. Gorelick, L., Blank, M., Shechtman, E., Irani, M., Basri, R.: Actions as space-time shapes. IEEE Transactions on Pattern Analysis and Machine Intelligence **29**(12) (2007) 2247–2253
19. Dollar, P., Rabaud, V., Cottrell, G., Belongie, S.: Behavior recognition via sparse spatio-temporal features. In: 2nd Joint IEEE Int. Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance. (2005) 65–72
20. Rius, I., González, J., Varona, J., Roca, X.: Action-specific motion prior for efficient Bayesian 3D human body tracking. Pattern Recognition **42**(11) (2009) 2907–2921

21. Laptev, I.: Recognizing human action in the wild. In Salah, A.A., Gevers, T., Sebe, N., Vinciarelli, A., eds.: HBU 2010. LNCS, vol. 6219, Springer, Heidelberg (2010) 87
22. Ryoo, M., Aggarwal, J.: Hierarchical recognition of human activities interacting with objects. In: IEEE Conference on Computer Vision and Pattern Recognition. (2007) 1–8
23. Poppe, R.: A survey on vision-based human action recognition. Image and Vision Computing (to appear)
24. Gao, Z., Chen, M.Y., Hauptmann, A., Cai, A.: Comparing evaluation protocols on the KTH dataset. In Salah, A.A., Gevers, T., Sebe, N., Vinciarelli, A., eds.: HBU 2010. LNCS, vol. 6219, Springer, Heidelberg (2010) 88–98
25. Schuldt, C., Laptev, I., Caputo, B.: Recognizing human actions: A local SVM approach. In: International Conference on Pattern Recognition. Volume 3., IEEE Computer Society (2004) 32–36
26. Vinciarelli, A., Pantic, M., Bourlard, H.: Social signal processing: Survey of an emerging domain. Image and Vision Computing **27**(12) (2009) 1743–1759
27. Poggi, I., D'Errico, F.: Dominance signals in debates. In Salah, A.A., Gevers, T., Sebe, N., Vinciarelli, A., eds.: HBU 2010. LNCS, vol. 6219, Springer, Heidelberg (2010) 161–172
28. D'Errico, F., Leone, G., Poggi, I.: Types of help in the teacher's multimodal behavior. In Salah, A.A., Gevers, T., Sebe, N., Vinciarelli, A., eds.: HBU 2010. LNCS, vol. 6219, Springer, Heidelberg (2010) 123–136
29. Lepri, B., Kalimeri, K., Pianesi, F.: Honest signals and their contribution to the automatic analysis of personality traits - a comparative study. In Salah, A.A., Gevers, T., Sebe, N., Vinciarelli, A., eds.: HBU 2010. LNCS, vol. 6219, Springer, Heidelberg (2010) 137–148
30. John, O., Srivastava, S.: The Big Five Trait Taxonomy: History, Measurement, and Theoretical Perspectives. In Pervian, L., John, O., eds.: Handbook of personality: theory and research, The Guilford Press (1999)
31. Pfister, T., Robinson, P.: Speech emotion classification and public speaking skill assessment. In Salah, A.A., Gevers, T., Sebe, N., Vinciarelli, A., eds.: HBU 2010. LNCS, vol. 6219, Springer, Heidelberg (2010) 149–160
32. Ozkan, D., Morency, L.P.: Concensus of self-features for nonverbal behavior analysis. In Salah, A.A., Gevers, T., Sebe, N., Vinciarelli, A., eds.: HBU 2010. LNCS, vol. 6219, Springer, Heidelberg (2010) 75–86
33. Ekman, P., Rosenberg, E.: What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS). Oxford University Press, USA (2005)
34. Mehrabian, A.: Nonverbal communication. Aldine (2007)
35. Zeng, Z., Pantic, M., Roisman, G., Huang, T.: A survey of affect recognition methods: Audio, visual, and spontaneous expressions. IEEE Transactions on Pattern Analysis and Machine Intelligence **31**(1) (2009) 39–58
36. Salah, A., Sebe, N., Gevers, T.: Communication and automatic interpretation of affect from facial expressions. In: Affective Computing and Interaction: Psychological, Cognitive and Neuroscientific Perspectives, IGI Global (to appear)
37. Yücel, Z., Salah, A.: Head pose and neural network based gaze direction estimation for joint attention modeling in embodied agents. In: Proc. 31st Annual Conference of Cognitive Science Society. (2009)
38. Cinar Akakin, H., Sankur, B.: Spatiotemporal-Boosted DCT Features for Head and Face Gesture Analysis. In Salah, A.A., Gevers, T., Sebe, N., Vinciarelli, A., eds.: HBU 2010. LNCS, vol. 6219, Springer, Heidelberg (2010) 64–74

39. Utsumi, Y., Iwai, Y., Ishiguro, H.: Face tracking and recognition considering the camera's field of view. In Salah, A.A., Gevers, T., Sebe, N., Vinciarelli, A., eds.: HBU 2010. LNCS, vol. 6219, Springer, Heidelberg (2010) 52–63
40. Ong, S., Ranganath, S.: Automatic sign language analysis: A survey and the future beyond lexical meaning. IEEE Transactions on Pattern Analysis and Machine Intelligence **27**(6) (2005) 873–891
41. Aarts, E., Encarnação, J.: True visions: The emergence of ambient intelligence. Springer (2006)
42. Crowley, J., Coutaz, J., Bérard, F.: Perceptual user interfaces: things that see. Communications of the ACM **43**(3) (2000) 54–64
43. Crowley, J.: Context driven observation of human activity. Proc. EUSAI (2003) 101–118
44. Guesgen, H., Marsland, S.: Spatio-temporal reasoning and context awareness. Handbook of Ambient Intelligence and Smart Environments (2010) 609–634
45. Pentland, A.: Looking at people: Sensing for ubiquitous and wearable computing. IEEE Transactions on Pattern Analysis and Machine Intelligence **22**(1) (2000) 107–119