A Card Playing Humanoid for Understanding Socio-emotional Interaction

Min-Gyu Kim* and Kenji Suzuki

University of Tsukuba, Dept. of System Information Engineering,
Tennodai, 1-1-1, Tsukuba, Japan
mingyu@ai.iit.tsukuba.ac.jp, kenji@ieee.org
http://www.ai.iit.tsukuba.ac.jp

Abstract. This paper describes the groundwork for designing a social and emotional interaction between a human and robot in game-playing. We considered that understanding deception in terms of mind reading plays a key role in realistic interactions for social robots. In order to understand the human mind, the humanoid robot observes nonverbal deception cues through multimodal perception during poker playing which is one of human social activities. Additionally, the humanoid manipulates the real environment which includes not only the game but also people to create a feeling of interacting with life-like machine and drive affective responses in determining the reaction.

Keywords: Human-robot social interaction, Mind reading, Deception, Humanoid playmate.

1 Introduction

In successful social interactions, humans understand and manipulate other people's behavior in terms of their mind. These capabilities are related to development of Theory of Mind (ToM) in developmental psychology. ToM is the ability to understand others' internal states and the relationship between human behaviors. Children with ToM understand that the others have mental states such as belief, motivation, emotion and intentions, and the mental states often cause behaviors. ToM is developed into the ability to infer the others' thought, desire and emotion, and the ability to predict the others' behavior based on their inference. A normally-developed children in 4-5 ages can understand the fact that belief plays a leading part in behavior and can distinguish the self desires from the other's belief. Also, the children can understand that the other's behavior can be determined from the belief and the knowledge based on perceptual experiences.

In this study, we focused on deception which is an important part of human social competence. According to ToM, the skills of deception in complex form

* This work is partially supported by Global COE Program on "Cybernics: fusion of human, machine, and information systems."

H.S. Yang et al. (Eds.): ICEC 2010, LNCS 6243, pp. 9–19, 2010.

are developed from false belief[1] which is a social cognitive ability to know that beliefs can be false and that they can be manipulated[2]. In the autistic children's case, the impaired false belief comprehension affects their social interaction and communication. From this point of view, forthcoming social robots should be able to cope with complicated social skills of human with the embedded cognitive ability to infer concealed intentions and emotions from social contexts and predict consequent behaviors.

Some of the preceding researches have made effort to develop social robots to understand human mind in a wide scope. The purpose of Leonardo underlies socially guided learning and social referencing[3],[4]. The embodiment makes it more capable of forming emotional attachment with humans. KASPAR is being investigated for cognitive development research[5]. In particular, it shows the possible use of therapeutic or educational robotic systems to encourage social interaction skills in autism children. Keepon is designed to study social development by communicating with children[6]. Its behaviors are intended to help children to understand its attentive and emotive actions. Most of the social robot related researches attain to mimic cognitive abilities of humans. On the other hands, there are many practical achievements which concentrate on game-playing agents capable of interacting with humans. In [7], Marquis *et al* have shown poker playing agents that embody emotions and communicate with people using multiple channels of communication. In [8], Kovács *et al* have built a virtual character to play chess on a physical board, which can express different emotions and produce speech.

In spite of many challenges from various angles mentioned above, because human social skills are extremely sophisticated, it is required to elaborate robots which can understand complex social skills and interact with people socio-emotionally in actual environments. So, in this study, we considered that understating deception in terms of mind reading plays a key role in complex and realistic interactions for social robots. Normally, when detecting deception through nonverbal contexts, the accuracy rate is usually about 50 percentages. However, raising stakes has an influence on the motivation to succeed in deceiving, and detecting the nonverbal deception in high stakes is easier than in low stake[9], [10]. In order to present how a humanoid robot can understand human deceptions in real social situations, we chose high-stakes poker since deception is inherent in game-playing. The humanoid plays poker, manipulating actual cards and observing human behaviors to understand how human feels and what human intents in the game. For mind-reading, our robot interprets nonverbal deception cues from multimodal stimuli such as facial expressions, eye blink, gestures and vocal stress. By observing human behaviors, it can realize what human reveals in the interaction.

In this paper, Section 2 will introduce the system overview for the social interaction. Section 3 will describe nonverbal deception cues that occur in daily social interaction and visual and auditory perceptions that have been implemented on the humanoid robot. Subsequently, the game environment will be briefly explained in Section 4 and the experiment will be demonstrated in Section 5.

2 System Overview

We formulated a component structure for the human-robot social interaction as shown in Figure 1. The humanoid perceives situational stimuli such as cards and their location. Also it observes emotional states of human player through visual and auditory sensors. The recognized information about current situation is memorized as experiences and considered when the humanoid selects next actions by decision-making based on present given situation and past experiences. The selected actions can have an effect on the surroundings including human and game, especially on the human player's circumstantial judgment and decision-making. This paper focuses on the groundwork for developing the ability to perceive and the ability to manipulate the surroundings.

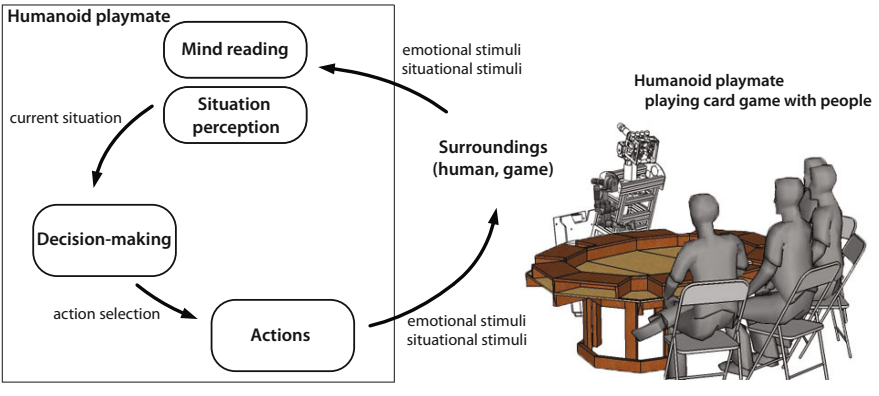

Fig. 1. A schematic diagram of human-robot social interaction

3 Nonverbal Cues for Mind Reading

3.1 Facial Feature Detection for Facial Expression Recognition

Much research on the body in cognitive sciences insists that temporal dynamics of human facial behavior such as the timing and duration of facial actions are a critical factor for interpretation of the observed behavior[11]. In particular, automatic facial expression recognition related research emphasizes the importance of facial expression's temporal dynamics for deception detection[12]. In this research, we adopted Active Appearance Models (AAMs) which are widely used for tracking and analyzing facial expressions.

AAM is a method to detect an object in 2D images using the statistical representation of shape and appearance variation[13]. AAMs align the pre-defined linear shape model with linear appearance variation to a previous unseen image containing the object. AAMs generally fit their shape and appearance components through a gradient decent. The appearance variation is linearly modeled by Principal Component Analysis (PCA) of shape $\mathbf{s}$ and texture $\mathbf{g}$.

$$\mathbf{s} = \mathbf{s}_m + \phi_s \mathbf{b}_s \qquad \mathbf{g} = \mathbf{g}_m + \phi_g \mathbf{b}_g \tag{1}$$

where $\mathbf{s}_m$, $\mathbf{g}_m$ are the mean shape and texture respectively, ϕ_s, ϕ_g are the eigenvectors of shape and texture covariance matrices. A third PCA is performed on a concatenated shape and texture parameters $\mathbf{b}$ to obtain a combined model vector $\mathbf{c}$.

$$\mathbf{b} = \phi_c \mathbf{c} \tag{2}$$

From the combined appearance model vector $\mathbf{c}$, a new instance of shape and texture can be generated.

$$\mathbf{s}_{model}(c) = \mathbf{s}_m + \mathbf{Q}_s \mathbf{c} \qquad \mathbf{g}_{model}(c) = \mathbf{g}_m + \mathbf{Q}_g \mathbf{c} \tag{3}$$

AAM fitting consists of minimizing the difference between the closest model instance and the target image by solving a nonlinear optimization problem. We employed Inverse compositional fitting methods.

3.2 Head and Hand Tracking for Gesture Recognition

According to Interpersonal Deception Theory, the use of gesture can lead to misinterpretations. Some researches have reported that when people lie, they display fewer of the gestures[15]. We considered that deceptive gestures naturally come out during the poker gaming, so that head and hand movement is one of influential factors in deception detection in the game.

For implementing head and hand movement tracking, it was assumed that in the captured images, humans move actively rather than backgrounds. Also, it was supposed that the humanoid's head movement to explore the environment will cause dynamic background changes. In order to extract foreground which indicates human's movements, we used a mixture of Gaussians to perform background subtraction in color images. A mixture of K Gaussian distributions adaptively models each pixel color. The probability density function of the kth Gaussian at pixel (i, j) at time t can be expressed as

$$N(x_{i,j}^t | m_{i,j}^{t,k}, \Sigma_{i,j}^{t,k}) = \frac{1}{(2\pi)^{\frac{n}{2}}} \tag{4}$$

Fig. 2. Facial expression tracking using Active Appearance Models

where $x_{i,j}^t$ is the color of pixel (i, j), $m_{i,j}^{t,k}$ and $\Sigma_{i,j}^{t,k}$ are the mean vector and the covariance matrix of the kth Gaussian in the mixture at time t respectively. Each Gaussian has an associated weight $w_{i,j}^{t,k}$ (where $0 < w_{i,j}^{t,k} < 1$) in the mixture. The covariance matrix is assumed to be diagonal to reduce the computational burden, that is, $\Sigma_{i,j}^{t,k} = diag((\sigma_{i,j}^{t,k,R})^2, (\sigma_{i,j}^{t,k,G})^2, (\sigma_{i,j}^{t,k,B})^2$ where R, G and B represent the three color components.

A K-means approximation of the EM algorithm is used to update the mixture model. Each new pixel color value, $x_{i,j}^t$, is checked against the existing K Gaussian distributions, until the pixel matches a distribution. The rest of our implementation followed the background subtraction technique in [16]. After the background subtraction, median filtering is performed to remove noise.

Next, the detected foreground pixels are processed further by skin color segmentation. Skin color segmentation is performed using the YCrCb color space. Color is represented by luma computed from nonlinear RGB, constructed as a weighted sum of the RGB values, and two color difference values Cr and Cb that are formed by subtracting luma from RGB red and blue components[17].

The output image extracted by the background subtraction and skin color detection includes potential candidates for the coordinates of head and hand. The face location is separated by haar-like feature based tracking to facilitate the segmentation between head and hand. Consequently, the hand in the image which the face is excluded can be easily detected by tracking blob features.

3.3 Eye Blink Detection

Some researches have shown that the eye blinking rate decreases when the cognitive load is increased. The measure of blink rate could provide another clue for the detection of deception[18].

At first, we took a template matching based approach to find the eye location. The basic idea of template matching is that one has the template of the sought feature region that has a high similarity with other images of that feature. The template and the similarity measurement are major parts within template matching. The right eye template is used in our system considering that the blinking of both eyes occurs simultaneously. A judgment of eye is made according to the similarity between the input image and the template. After seeking the most similar region with the right eye template, the eye blink is be detected.

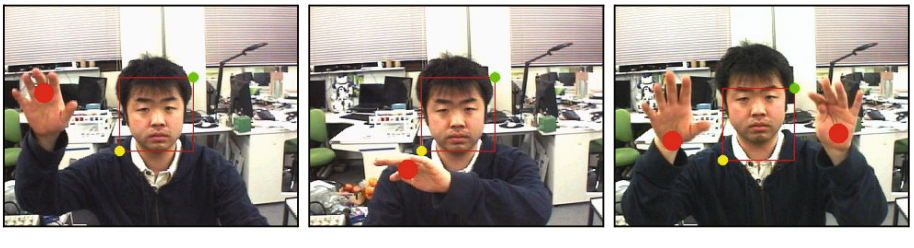

Fig. 3. Head and hand tracking for gesture recognition

Fig. 4. Blink detection

Each input image containing the closed eye and the opened eye has different similarity value as compared with the template image.

3.4 Voice

When we feel being deceived by people, we pay attention not only to their bodily expressions but also to vocal information such as voice tone and voice pitch. According to research on stress and lie detection[19], non-invasive physiological features like voice pitch variation is correlated to high stress situations. We have realized the basement to analyze the deceptive vocal information by MATLAB. In this paper, the spoken voice during the poker playing was recorded.

4 Humanoid Playmate

When a game is too easy, people feel bored and when it is too difficult, they feel frustrated. Balancing game with different levels is one of key issues in game design. From the psychological perspective of manipulating mind, one of our game strategies is to keep people to play the game for satisfaction and happiness. Game level adjustment plays an important role in satisfying this requirement. In poker game strategies, we would follow the flow model (see Figure 5), defined by M. Csikszentmihalyi as the mental state in which people are involved in an activity that nothing else seems to matter. Another game strategy is that the humanoid learns and imitates the human player's behaviors to create a sense of interacting with life-like machine and drive the human's affective responses in determining the behavioral reaction.

In order to accomplish the strategies, we built the humanoid as a playmate shown in Figure 6, named Genie developed by Artificial Intelligence Lab. at University of Tsukuba. The humanoid is composed of the upper torso with a wall-mounted 3DOF waist. It has totally 27DOF (8DOF for the head, 3DOF for the waist, 7DOF for the right arm, 4DOF for the right hand, and 5DOF for the left arm) in the body. The SSSA-Tsukuba artificial hand is comprised of three fingers and a tendon-driven mechanism. Additionally, the robot hand has a 1 DOF thumb for the adduction and abduction actuation of the thumb.

During the poker playing, the humanoid can manipulate actual poker cards. We attached an electromagnet on the fingertip of the robot hand because the

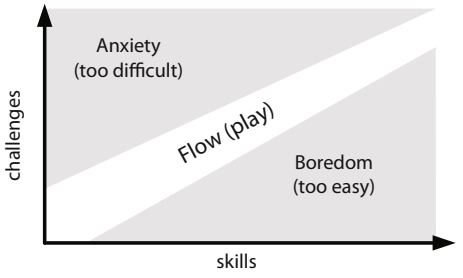

Fig. 5. Flow model with regard to challenge level and skill level

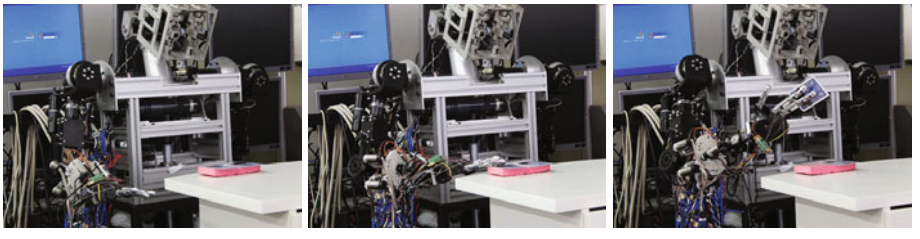

Fig. 6. Card manipulation (initial pose, reaching and picking up)

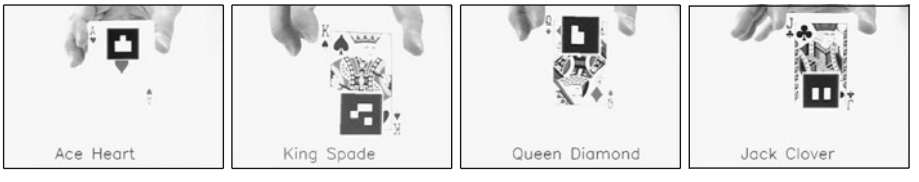

Fig. 7. Card suit recognition by AR toolkit marker

complex dynamic analysis of object manipulation by multi-fingered robot hand is out of our research scope. For physically manipulating cards, the humanoid has the kinematic profile of the movement. With the profiles, it can reach the target cards on the table as well as pick up to check its own card suits. Figure 6 illustrates the card manipulation by the right arm and hand which are initial pose, reaching and picking up, respectively. Perceiving card suits is achieved by tracking AR toolkit markers pasted on each card as shown in Figure 7.

5 Experiment

The purpose of this experiment is to evaluate the humanoid's observations on the natural scenes in a poker game regardless of deception detection and to get ideas for designing the robotic actions to measure. For the experiment, we asked two

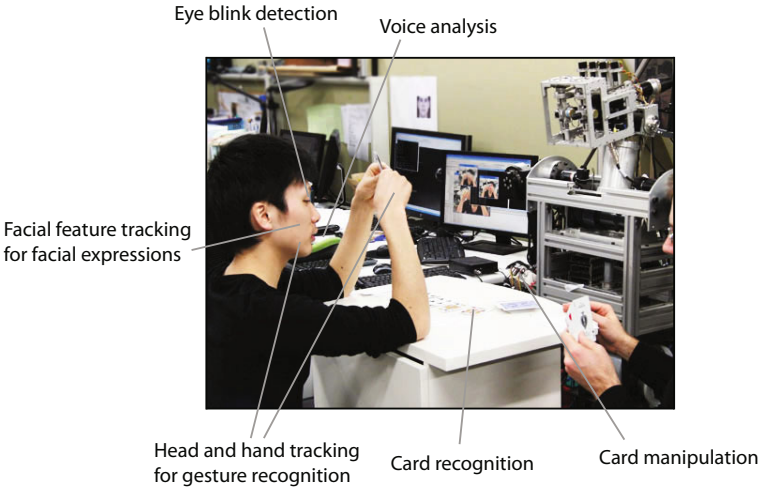

Fig. 8. Bodily expressions and voice observation during the poker game

people to play the Texas hold'em in front of the humanoid and to act naturally as shown in Figure 8. In order to encourage the participants to play for win, we provided a incentive. We then made the robot observe the behaviors of the subject who sat before it when two players were playing the game. It perceived his facial expressions, gestures, eye blink and voice without moving its body. The facial features and voice were recorded during the entire poker game. Because we expected that the player would not perform gestures and eye blinking a lot according to [18] and [20], they were analyzed by turns. We had implemented the card suit recognition and the card manipulation as described in Section 4. However, they were not applied to this experiment since the Texas hold'em game engine is not incorporated into the humanoid.

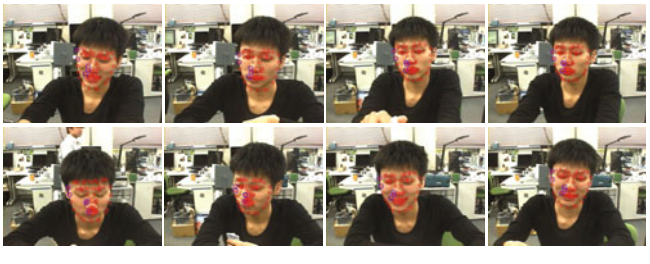

Fig. 9. Detection of facial features

Figure 9 demonstrates the facial features detected by using Active Appearance models. The 68 red landmarks fit the change of detected facial features. During the game, although the subject moved and turned the head a lot, the facial features were easily detected within the possible measurement range.

Fig. 10. Head and hand tracking

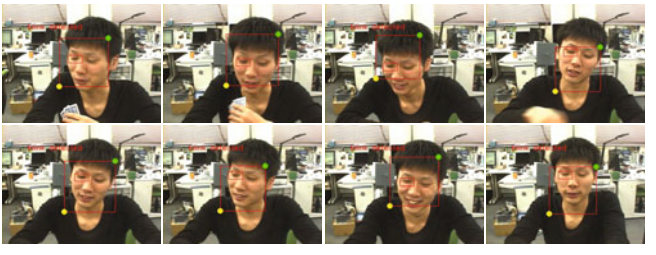

Fig. 11. Eye blink detection

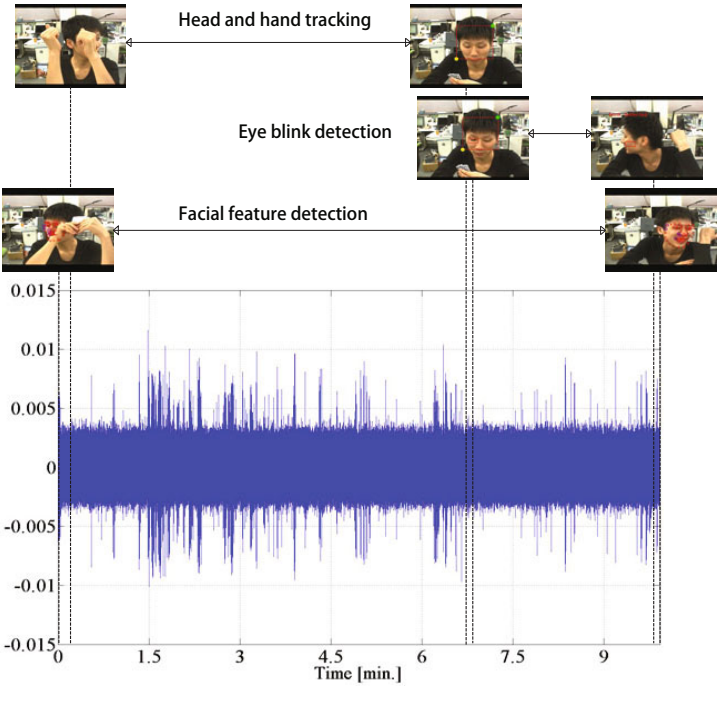

Fig. 12. The recorded sound wave in the time domain

In the game, the subject did not move his arm and hand much except for picking up of fake money for betting and shuffling the cards as a dealer. On the other hand, he often moved his head to comprehend the situation, and moved when he laughed out. The hand movements were traced excluding the cases that the subject put the hands on the table out of the range of vision. As shown in Figure 11, for the natural behaviors, the eye blink was not detected well because it was not easy to distinguish between the real eye blinking case and the other cases where the eyes were seemed to be closed when the subject looked down without hanging the head or really closed when he smiled broadly.

The sound which was occurred during the whole game is shown in Figure 12. The figure demonstrates the sound wave plotted in the time domain. The input raw sound was preprocessed by a band-pass filter within the voice frequency to reduce noise in the environment.

6 Conclusion

In this research, we have designed a humanoid playmate to understand human mind in terms of mind reading. Especially, a perceptual system has been implemented to acquire the circumstantial information such as nonverbal expressions and the poker cards. Also, simple card manipulation has been achieved simply to reach, pick up and put down the cards. In order to analyze the deceptive vocal information, we have realized a basic sound processing module. Finally in the experiment, the human player's natural behaviors were observed during a real poker game.

In the near future, we would improve the humanoid's actions to observe actively by moving the head, waist etc because the static observation has a limitation of analyzing the dynamic human behaviors. Additionally, in order to play with people interactively, we would integrate the Texas hold'em engine into the humanoid and develop the mind reading mechanism by combining the bodily expressions and voice to understand intentions and feelings.

References

1. Wimmer, H., Perner, J.: Beliefs about beliefs: Representation and Constraining Function of Wrong Beliefs in Young Children's Understanding of Deception. Cognition 13, 103–128 (1983)
2. Baron-Cohen, S., Tager-Flusberg, H., Cohen, D.: Understanding Other Minds: Perspectives from Developmental Cognitive Neuroscience, Oxford (1999)
3. Homaz, A.L., Berlin, M., Breazeal, C.: Robot Science Meets Social Science: An Embodied Computational Model of Social Referencing. In: COGSCI 2005 Workshop, pp. 7–17 (2005)
4. Thomaz, A.L., Breazeal, C.: Robot Learning via Socially Guided Exploration. In: 6th IEEE International Conference on Developmental Learning, pp. 82–87 (2007)
5. Dautenhahn, K., Nehaniv, C.L., Walters, M.L., Robins, B., Kose-Bagci, H., Assif Mirza, N., Blow, M.: KASPAR - A Minimally Expressive Humanoid Robot for Human-Robot Interaction Research. Special Issue on Humanoid Robots, Applied Bionics and Biomechanics, 369–397 (2009)

6. Kozima, H., Michalowski, M.P., Nakagawa, C.: Keepon: a Playful Robot for Research, Therapy, and Entertainment. Int. J. of Social Robotics 1, 3–18 (2008)
7. Marquis, S., Elliott, C.: Emotionally Responsive Poker Playing Agents. In: Notes for the 12th National Conference on Artificial Intelligence Workshop on Artificial Intelligence, Artificial Life, and Entertainment, pp. 11–15 (1994)
8. Kovács, G., Ruttkay, Z., Fazekas, A.: Virtual Chess Player with Emotions. In: 4th Hungarian Conference on Computer Graphics and Geometry (2007)
9. DePaulo, B.M., Stone, J.I., Lassiter, G.D.: Deceiving and Detecting Deceit. The Self and Social Life, 323–370 (1985)
10. DePaulo, B.M., LeMay, C.S., Epstein, J.A.: Effects of Importance of Success and Expectations for Success on Effectiveness at Deceiving. Personality and Social Psychology Bulletin, 14–24 (1991)
11. Ambadar, Z., Schooler, J.W., Cohn, J.F.: Deciphering the Enigmatic Face: the Importance of Facial Dynamics in Interpreting Subtle Facial Expressions. Psychol. Sci. 403–410 (2005)
12. Valstar, M.F., Pantic, M., Ambadar, Z., Cohn, J.F.: Spontaneous vs. Posed Facial Behavior: Automatic Analysis of Brow Actions. In: 8th International Conference on Multimodal Interfaces, pp. 162–170 (2006)
13. Matthews, I., Baker, S.: Active Appearance Models Revisited. Int. J. of Computer Vision 60, 135–164 (2004)
14. Cootes, T.F., Edwards, G.J., Taylor, C.J.: Active Appearance Models. In: Burkhardt, H., Neumann, B. (eds.) ECCV 1998. LNCS, vol. 1407, pp. 484–498. Springer, Heidelberg (1998)
15. Vrij, A., Edward, K., Roberts, K.P., Bull, R.: Detecting Deceit via Analysis of Verbal and Nonverbal Behavior. J. of Nonverbal Behavior, 239–263 (2004)
16. Javed, O., Shafique, K., Shah, M.: A Hierarchical Approach to Robust Background Subtraction using Color and Gradient Information. In: Workshop on Motion and Video Computing, pp. 22–27 (2002)
17. Vassili, V.V., Sazonov, V., Andreeva, A.: A Survey on Pixel-Based Skin Color Detection Techniques. In: Proc. Graphicon-2003, pp. 85–92 (2003)
18. Fukuda, K.: Eye Blinks: New Indices for the Detection of Deception. Int. J. of Psychophysiology, 239–245 (2001)
19. Sung, M., Pentland, A.: PokerMetrics: Stress and Line Detection through Non-Invasive Physiological Sensing. Ph.D. thesis, MIT Media Laboratory (2005)
20. Caso, L., Maricchiolo, F., Bonaiuto, M., Vrij, A., Mann, S.: The Impact of Deception and Suspicion on Different Hand Movements. J. of Nonverbal Behavior 30, 1–19 (2006)