HMM-BASED DECEPTION RECOGNITION FROM VISUAL CUES

Gabriel Tsechpenakis[(1)], Dimitris Metaxas[(1)], Mark Adkins[(2)], John Kruse[(2)], Judee K. Burgoon[(2)], Matthew L. Jensen[(2)], Thomas Meservy[(2)], Douglas P. Twitchell[(2)], Amit Deokar[(2)], and Jay F. Nunamaker [(2)]

[(1)] Center for Computational Biomedicine, Imaging and Modeling (CBIM), Division of Computer and Information Sciences, Rutgers University, 110 Frelinghuysen Road, Piscataway, NJ 08854-8019, USA

[(2)] Center for the Management of Information (CMI), The University of Arizona, 1130 East Helen Str., Tucson, AZ 85721-0108, USA

ABSTRACT

Behavioral indicators of deception and behavioral state are extremely difficult for humans to analyze. This research effort attempts to leverage automated systems to augment humans in detecting deception by analyzing nonverbal behavior on video. By tracking faces and hands of an individual, it is anticipated that objective behavioral indicators of deception can be isolated, extracted and synthesized to create a more accurate means for detecting human deception. Blob analysis, a method for analyzing the movement of the head and hands based on the identification of skin color is presented. A proof-of-concept study is presented that uses blob analysis to extract visual cues and events, throughout the examined videos. The integration of these cues is done using a hierarchical Hidden Markov Model to explore behavioral state identification in the detection of deception, mainly involving the detection of agitated and over-controlled behaviors.

1. INTRODUCTION

Deception recognition has a wide range of applications, such as security in buildings, border crossing, airport screening and interrogations.

Behaviors associated with deception might be classified into two groups: agitation and over-control. In an effort to suppress deceptive cues and appear truthful [6], liars may overcompensate and dramatically reduce all behavior [4, 14]. Such tenseness and over-control can be seen in decreased head movements [1], leg movements [5] and hand and arm movements [15] which may accompany deceptive communication.

Two theories that guide the development of automated systems for detecting deception through identifying agitated and controlled behavior are Interpersonal Deception Theory (IDT) and Expectancy Violations Theory (EVT) [2, 8]. According to these theories, deception is a dynamic process and a comparison between expected and received messages may be more helpful in identifying deceit than searching for a group of deception indicators.

In our work, in order to recognize agitated and over-controlled behaviors, we extract the head and hands of the examined subject, using the method presented in [10] and we determine, extract and integrate movement features and events that can be used for the detection. According to our approach, there are two levels of analysis-recognition.

In the first level, we use the movement descriptors extracted directly from the blob analysis to recognize two kinds of movements: (a) the adaptors, which are movements indicating a low level of the subject's awareness, such as self touching, and (b) the illustrators, which are gestures performed while talking, illustrating and assisting the speech. These kinds of movements assist the recognition of deceptive behavior, since reduced illustrators and increased adaptors are identical in deception. In the second level of our analysis, we use the adaptors and illustrators, along with some movement descriptors directly extracted from the blob analysis, to recognize possible deception.

The two levels of our analysis are defined and implemented as a two-layer hierarchical HMM [13, 9], where the first layer concerns the learning of the adaptors/illustrators from the blob analysis, whereas the second layer corresponds to the deception recognition.

The paper is organized as follows: Section 2 explains our approach in identifying deception based on observed behavior. Subsections 2.1 and 2.2 explore the steps that

0-7803-9332-5/05/$20.00 ©2005 IEEE

are involved in the visual cues extraction, i.e. the head and hands blob analysis and the movement descriptors that we use, whereas subsection 2.3 explains how our movement descriptors describe the behavioral states. In Section 3, our hierarchical HMM-based recognition module is described, and in Section 4 our experimental results are presented and discussed. Finally, Section 5 addresses our conclusions and future steps.

2. OUR APPROACH

In our system, we follow three main steps. First, we detect and track the regions of interest in the examined video, i.e. the head and hands, using the skin color-based method of [10], briefly described in the following subsection. Then, from the extracted blob features, i.e. positions and orientations, we extract the movement descriptors used in the recognition. Finally, we use the proposed HMM-based approach to detect and recognize two possible behavioral states, namely the agitation and over-control, which indicate possible deception.

2.1. Blob Analysis

The first important step of our approach is to detect and track the body parts of our interest, i.e. the head and hands. Although research efforts have investigated this issue [7, 16, 11], accurate tracking of people and their body parts is still an open topic.

According to [10], using color analysis, eigenspace-based shape segmentation, and Kalman filters, we have been able to track the position, size, and angle of different body parts with great accuracy. Fig. 1 shows a single frame of four sample videos, which have been subjected to blob analysis. The ellipses in the figure represent the body parts' position, size, and angle.

Blob analysis extracts hand and face regions using the color distribution from an image sequence. A Look-Up-Table (LUT) with three color components (red, green, blue) is created based on the color distribution of the face and hands. This three-color LUT, called a 3-D LUT, is built in advance of any analysis and is formed using skin color samples. After extracting the hand and face regions from an image sequence, the system computes elliptical "blobs" identifying candidates for the face and hands. The 3-D LUT may incorrectly identify candidate regions which are similar to skin color, however these candidates are disregarded through fine segmentation and comparing the subspaces of the face and hand candidates. Thus, the most face-like and hand-like regions in a video sequence are identified. From the blobs, the left hand, right hand and face are tracked continuously, i.e. using the motion information over time. Also, from positions and movements of the hands and face we can make further inferences about the torso and the relation of each body part to other people and objects.

Figure 1. Head and hands blob extraction in a single frame of four different sequences.

2.2. Movement Descriptors

Our main goal in tracking the head and hands is to identify a movement signature from which we could roughly estimate the subject's behavioral state. The term "signature" is used to describe how smooth or abrupt the movements are, how large the displacements of the head and the hands are, how often the hands touch the face and how often the hands come together. What we actually extracted and investigated is the motion trajectory, i.e. the projection of the three-dimensional motion on the image plane. For this purpose, after extracting the blobs, we recorded the successive positions of their centers and their change through time.

When the two hands come together or when a hand touches the face, the two corresponding blobs are merged into one and then we obtain results for only two blobs. When two blobs are merged into one, the blob centers' positions change rapidly, and this is the indication we use to detect such merging blobs.

The next step is to estimate (i) the position and velocity variances, which indicate how smooth the movements are, (ii) the number of times that hands come together, (iii) the number of times that a hand (or both hands) touches the face, and (iv) the duration of a hand touching the face. The events of hands touching the face and hands coming together are crucial for two reasons: (a) they may partially indicate a behavioral state, and (b) they constitute time segments in which the blob movements (positions, velocities and their variances) should not be taken into consideration; thus, we examine blob movements only when such events do not occur.

2.3. Behavioral State Indications based on Movement Observations.

After examining the videos of our training set, we found that the two behavioral states we want to recognize, i.e. "over-controlled" and "agitated" can be determined by the parameters described above. When a subject is over-controlled, there are only small blob displacements and the hands do not touch the face often. When a subject is agitated, the hands move often and more abruptly, and they touch the face more often and with short duration. Finally, when agitation or over-control is not detected, we assume that the examined subject's state is *normal* and what we observe is smooth hand movements, large displacements, and even hands touching the face often. These observations are verified by our state recognition scheme described in the next Section.

3. BEHAVIORAL STATE RECOGNITION

We chose to use an HMM-based approach to recognize the two main behavioral states of our interest, i.e. over-control and agitation. HMMs are widely used in gesture, gait and sign languages recognition [13], since they are ideal for activity recognition, involving continuous visual cues and their variations over time.

Generally, an HMM $\boldsymbol{\lambda}$ consists of a set of n states S_1, S_2, …, S_n. At regularly spaced discrete time intervals, the transition probability from state S_i to state S_j is α_{ij}, and the initial (starting) probability in the state S_i is π_i. Each S_i generates output $O \in \boldsymbol{\Omega}$, which is distributed according to a probability density function $\beta_i(O) = P(O \mid S_i)$, i.e. the probability of having the observation O in the system state S_i. In most recognition problems, $\beta_i(O)$ is a mixture of gaussian densities.

In our application, we need to recognize the aforementioned behavioral states when they are observed, as well as the changes in the subjects' behavior over time. An intermediate step of our approach is to recognize the identical movements, i.e. the adaptors and the illustrators. The recognition of such movements can be then used as observations, along with the visual cues extracted from the video, to estimate the system's state, i.e. the examined subject's behavioral state. Thus, our approach can be seen as two different HMMs, where the results (states) of the first are used as observations to the second one. To integrate this procedure and also include in the observation vectors of the second HMM the observations (or part of the observations) of the first one, we used the hierarchical structure illustrated in Fig. 2.

In the first layer, each observation vector X_i consists of the visual cues described in subsection 2.2, where each parameter is estimated in a reasonable time interval, which is 3-5secs or 90-150 frames in a 30fps video. The states Y_i represent either an adaptor or an illustrator, i.e. Y_i = {*adaptor*, *illustrator*}. In the next layer of our system, for each state Z_i we use as observation vectors the corresponding X_i and Y_i. The result of our recognition system is "agitation", "over-control" or "normal", i.e. Z_i = {*agitation*, *over-control*, *normal*}, where "*normal*" corresponds to the decision when no agitated or over-controlled behavior is observed. The Baum-Welch algorithm [12] is used to train separately the two HMMs (layers), and the Viterbi algorithm [12] is used to find the most likely state sequence during the recognition phase.

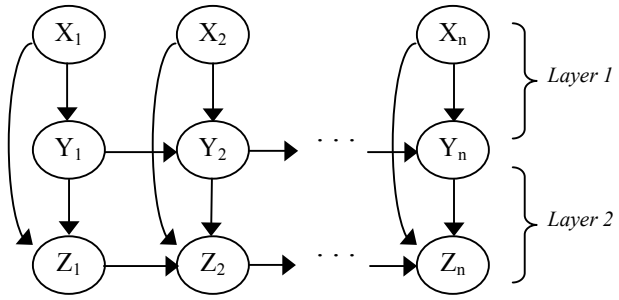

Figure 2. Hierarchical HMM structure for the behavioral state recognition

4. EXPERIMENTAL RESULTS

We used four sources for acquiring the training set for our hierarchical HMM: (a) seven CMI employees performed twenty gestures involving fingers, hands, arms, trunk and head, and each gesture was repeated 10-12 times by each participant, (b) four actors were hired for a proof-of-concept study to determine the feasibility of identifying behavioral states from gestures and body movement, simulating airport screening scenarios, (c) the Mock Theft Experiment [3] (38 videos), where some participants played the role of a thief while others where simply present during the theft, and they were all interviewed by untrained and trained interviewers, and (d) 26 real bank theft interviews. Totally we recorded and used as training set over 200 hours of video.

Tables 1 (a) and (b) show the parameters extracted for five interviews. The first three columns of Table 1(a) show the case examined, the behavioral state (a priori known), and the respective total video duration in seconds. Columns 4-6 show the variance of the blob's position for the head and the hands. Continuing the results for these five cases, in Table 1(b) the variance of the respective velocities are shown, and Table 1(c) shows the number of times the events "hands on face" and "hands together" occur; the last column shows the maximum duration of any two blobs being merged.

Applying our method in 27 testing videos, where the interviewed subjects were actors, we found out that in deceptive behaviors, i.e. when a subject is either over-

controlled or agitated, there is an increase of the adaptors performed, there is less head movement, and significantly fewer illustrators. The accuracy of recognizing deception was 92%, while the accuracy of recognizing a normal state was 90.9%.

Also, in 9 videos of real interviews (non-actors), the accuracy of recognizing deceptive behavior was 87.5%.

Case	State	Duration	Position change variance		
			Head	Hand (left)	Hand (right)
1	agitated	115	276.18	516.80	492.26
2	agitated	29	114.83	69.89	282.67
3	over-controlled	92	24.89	6.61	11.85
4	relaxed	68	260.13	303.05	104.83
5	relaxed	29	86.76	492.26	276.75

Table 1(a). Blobs' position variances.

Case	Velocity variance		
	Head	Hand (left)	Hand (right)
1	0.58	8.37	6.88
2	0.61	51.71	92.70
3	0.14	0.32	0.82
4	6.08	5.80	0.57
5	0.35	4.06	8.37

Table 1(b). Respective blobs' velocity variances.

Case	Hand on face (times / duration)	Hands together (times / duration)	Maximum duration (frames / duration)
1	9.57	0.3478	1.57
2	13.79	0	3.78
3	0	0.0217	3.11
4	2.94	0	13.31
5	27.59	0.1034	21.11

Table 1(c). Respective frequencies and durations of the examined events recognized from blob analysis.

5. CONCLUSIONS AND FUTURE WORK

We have developed a system for automated behavioral analysis, in terms of recognizing possible deception. We used an existing method for head and hands tracking, extracting blobs for the regions of interest. Our HMM-based approach for deception detection, allows the observation of behavioral changes over time, and is robust to gesture variations. We utilized movement descriptors that can reliably indicate the behavioral states, i.e. visual cues from the examined video, and descriptors learnt from these visual cues.

Our future steps involve the enrichment of our training set with real scenarios, such as airport screening and border crossing interviews. We are currently extending the visual cues, including torso and shoulders position, and relative positions between the blobs. Finally, we aim at a deception detection system close to real time.

6. REFERENCES

[1] D. Buller, J. Burgoon, C. White, and A. Ebesu, "Interpersonal Deception: VII. Behavioral Profiles of Falsification, Equivocation and Concealment," *Journal of Language and Social Psychology*, vol. 13, pp. 366-395, 1994.

[2] D. Buller, and J. Burgoon, "Interpersonal deception theory," Communication Theory, vol. 6, pp. 203-242, 1996.

[3] J. K. Burgoon, J. P. Blair, and E. Moyer, "Effects of Communication Modality on Arousal, Cognitive Complexity, Behavioral Control and Deception Detection During Deceptive Episodes," *Annual Meeting of the National Communication Association*, Miami Beach, Florida, 2003.

[4] B. DePaulo, J. Lindsay, B. Malone, L. Muhlenbruck, K. Charlton, and H. Cooper, "Cues to deception," *Psychological Bulletin*, vol. 129, pp. 74-118, 2003.

[5] P. Ekman, "Lying and Nonverbal Behavior: Theoretical Issues and New Findings," *Journal of Nonverbal Behavior*, vol. 12, pp. 163-176, 1988.

[6] P. Ekman, *Telling lies: Clues to deceit in the marketplace, politics, and marriage, vol. 2*, WW Norton and Company, New York, 1992.

[7] D.M. Gavrila, "The Visual Analysis of Human Movement: A Survey", *Computer Vision and Image Understanding*, Vol. 73(1), pp.82-98, 1999.

[8] J. George, D. P. Biros, J. K. Burgoon, and J. Nunamaker, "Training Professionals to Detect Deception," *NSF/NIJ Symposium on Intelligence and Security Informatics*, Tucson, AZ, 2003.

[9] N. Liu, B. C. Lovell, and P. J. Kootsookos, "Evaluation of HMM Training Algorithms for Letter Hand Gesture Recognition," *IEEE International Symposium on Signal Processing and Information Technology*, Darmstadt, Germany, December, 2003.

[10] S. Lu, G. Tsechpenakis, D. Metaxas, M. L. Jensen, and J. Kruse, "Blob Analysis of the Head and Hands: A Method for Deception Detection and Emotional State Identification," *Hawaii International Conference on System Sciences,* Big Island, Hawaii, January 2005.

[11] T. B. Moels, and E. Granum, "A Survey of Computer Vision-Based Human Motion Capture", *Computer Vision and Image Understanding*, Vol. 81(3), pp.231-268, 2001.

[12] L. R. Rabiner, "A tutorial on Hidden Markov Models and selected applications in speech recognition," *Proceedings of the IEEE*, 77(2), pp. 257–286, 1989.

[13] C. Vogler, H. Sun, and D. Metaxas, "A Framework for Motion Recognition with Applications to American Sign Language and Gait Recognition," *Workshop on Human Motion*, Austin, TX, December, 2000.

[14] A. Vrij, *Detecting lies and deceit: The psychology of lying and its implications for professional practice*, Wiley, Chichester, UK, 2000.

[15] A. Vrij, K. Edward, K. Roberts, and R. Bull, "Detecting deceit via analysis of verbal and nonverbal behavior," *Journal of Nonverbal Behavior*, vol. 24, pp. 239-263, 2000.

[16] Y. Wu and T. S. Huang, "Vision-Based Gesture Recognition: A Review," *International Gesture Workshop (GW'99)*, Gif-sur-Yvette, France, March 1999.