Blob Analysis of the Head and Hands: A Method for Deception Detection

Shan Lu, Gabriel Tsechpenakis and Dimitris N. Metaxas
Center of Computational Biomedicine Imaging and Modeling
Rutgers University
shanlu@cs.rutgers.edu

Matthew L. Jensen and John Kruse
Center for the Management of Information
University of Arizona
mjensen@cmi.arizona.edu

Abstract

Behavioral indicators of deception and behavioral state are extremely difficult for humans to analyze. Blob analysis, a method for analyzing the movement of the head and hands based on the identification of skin color is presented. This method is validated with numerous skin tones. A proof-of-concept study is presented that uses blob analysis to explore behavioral state identification in the detection of deception.

1. Introduction

People have long sought to divine lies and truths in the world around them. For actions and decisions, both large and small, humans have to rely on others. Realistically, people have divergent needs and motivations that are the root of deception and betrayal. In response, we take the natural step of trying to find out what is and is not truthful.

These efforts range from simple fact checking, to polygraphs, to pseudoscientific physiognomic approaches. Some of these lines of attack are grounded in reality, while others actually decrease detection accuracy. Typically, most people fall victim to the Truth Bias where they tend to believe that others are truthful even when this trust is misplaced. On the other hand, law enforcement and intelligence professionals often hold the Othello Bias – a belief that people are generally deceptive. In either case, people are not effective at detecting deception [1].

There is, nonetheless, much scientific basis for the belief that deception can be identified through careful analyses of human action and response. There are several complimentary theories as to why these approaches may be effective. Perhaps the most well known is the physiological approach whereby people become aroused by deceit and may be betrayed by changes in such things as respiration or heart rate. The polygraph is based on these uncontrolled responses. Other theories rest on reflections of the differences between cognitive processes involved with recall of memories versus fabrication, or changes in gestures, language or mannerisms.

The problem with assessing behavioral indicators of deception is that they (1) are difficult to objectively discern, (2) require a great deal of attention, and (3) are easily thwarted by a range of biases.

In response, 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.

In this paper, we present our current research efforts in the direction of developing automated tools to identify deception and behavioral state. The paper is organized as follows: Section 2 explains our theory based approach in identifying deception based on observed behavior. Section 3 explores the steps that are involved in blob analysis and offers validation of blob analysis with multiple skin colors. Section 4 shares a proof-of-concept study that uses blob analysis in the identification of behavioral state and Section 5 addresses future steps.

2. Theoretical foundation

Some behaviors associated with deception might be classified into two groups: agitation and over-control. Related to agitation are manifestations of nervousness and fear [1]. Behaviors that have been linked with nervousness and fear include faster and louder speech [1] and fidgeting [2]. However, the link between fidgeting and deception is still debated. A large meta-analysis reviewing numerous studies on deception found a significant relationship between undirected fidgeting and deception, although it questions the role of self-touches and object touches in predicting deception [3].

Liars may be aware of behavioral cues, such as fidgeting, which might reveal their deception. In an effort to suppress deceptive cues and appear truthful, liars may overcompensate and dramatically reduce all behavior [3, 6]. Such tenseness and over-control can be seen in decreased head movements [7], leg movements [8] and hand and arm movements [9] 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). Buller and Burgoon's Interpersonal Deception Theory [4, 5] states that deception is a dynamic process. Deception is portrayed as a game of moves and countermoves where the deceiver adjusts the message in response to the perceived trust or suspicion of the receiver.

Observing what appears to be agitation or over-control in a person's communication does not necessarily mean that person is being deceptive. The person's normal behavior and the context in which the communication takes place should also be considered. Burgoon's Expectancy Violations Theory (EVT), is concerned with what nonverbal and verbal behavior patterns are considered normal or expected, what behaviors constitute violations of expectations, and what consequences violations create [10]. When applied to deception, this theory suggests that a comparison between expected and received messages may be more helpful in identifying deceit than searching for a group of deception indicators. If there is significant deviation between what is received and what is expected, suspicion is aroused. For example, suspicion could be raised if a suspect is relaxed during most of an interview, but suddenly becomes rigid and tense during questioning about the details of a crime.

3. Tracking the hands and face with blob analysis

Central to the recognition of changes in behavioral signals is the ability to recognize and track body parts such as the head and hands. Although research efforts have investigated this issue [11, 12, 13], accurate tracking of people and their body parts is still an open topic.

The Computational Biomedicine Imaging and Modeling Center at Rutgers University provides a foundation on which it is possible to track human body parts [14, 15]. 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. Figure 1 shows a single frame of a video which has been subjected to blob analysis. The ellipses in the figure represent the body parts' position, size, and angle.

Figure 1. Blobs capture the location of the head and hand

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, and 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 can be tracked continuously. 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. This allows the identification of gestures, posture and other body expressions. This process is described below.

3.1 Skin color segmentation

The skin color identification algorithm extracts hand and face regions using the color distribution from the image sequence. We prepare a 3-D LUT which is used for setting the color distribution. This 3-D LUT is trained using color sample images and is based on histogram back-projection [16]. Then the system extracts the hand and face regions using the 3-D LUT.

We set the 3-D LUT with skin color samples extracted from color images in advance such as those shown in Figure 2. Due to the skin color segmentation, all pixels in

the color samples can be classified into either the skin color region *M* or the background region *B*. According to the multi-dimensional color histogram of both regions, a combined histogram *C* is defined as

$$C_j = \max(\frac{w_m M_j - w_B B_j}{C_{\max}}, 0) \times D,$$

where *j* represents the index of each histogram bin, $C_{\max}$ is the maximum value of the combined histogram, *D* is the range of the output image, e.g. 255 for an 8-bit image, and w_m and w_b are weighted values representing the sensitivity for *C* in each histogram.

All pixels in the images are converted using the 3-D LUT as follows. If the color value of the pixel is the same as the mean color value of a unit cube, the value of the pixel is converted to the value of the unit cube. If the color pixel value is a value other than the mean color value of the unit cube, the pixel value is converted to a value that is interpolated by the PRISM algorithm [17] using the values of six neighboring cubes.

Identification of skin color is made difficult because the skin color varies based on a number of factors. First, skin color varies largely by ethnicity. However, for several ethnicities such as Caucasian, African American, Latin American, and Asian, the distribution of skin color in a color space concentrates in a small region [18]. 3-D LUTs for these ethnic groups are created by using correspondent color images from each ethnicity.

Other factors affecting skin-color detection are the lighting and view direction of the camera. Completely removing these physical influences is impossible in the real world. However, research has shown the normalization of color is effective in alleviating these influences [19]. In our system, prior to creating the color 3-D LUT, we convert the original (R, G, B) to a normalized color space (r, g, b) as

$$r = \frac{R}{R+G+B}, \quad g = \frac{G}{R+G+B}, \quad b = \frac{B}{R+G+B}$$

The new color value will be within the range of 0 to 1. For example, a color value (R=255, G=255, B=0) will become (0.5, 0.5, 0) by the normalization.

3.2 Detection of hand and face shape

After extracting the hand and face regions from an image sequence using a color 3-D LUT, the system computes blobs using the extracted image sequence and then tracks the left hand, right hand, and face. However, this color segmentation process may classify the wrong area as a hand or face because it has a color distribution similar to skin color. Rough searching and fine segmentation are used to avoid misclassification.

In rough searching, we first fit the detected area using a simple geometric shape, such as ellipse. Only those areas that meet specific standards remain as candidates of the face and hands for the fine segmentation. During fine segmentation, we define the most face-like and hand-like areas as the actual face and hand by comparing with subspaces of candidate face and hands.

With rough searching, we consider a connected skin color area as a blob. While approximating a blob shape by use of an ellipse, we calculate its shape and motion parameters using image moments [20] as following.

First, the central position (x_c, y_c) and the velocity *v(i)* of this blob are estimated as following.

$$x_c = \frac{M_{10}}{M_{00}} \quad y_c = \frac{M_{01}}{M_{00}}$$

$$v(i) = \sqrt{(x_c(i) - x_c(i-1))^2 + (y_c(i) - y_c(i-1))^2} / \nabla T$$

where M_{00}, M_{10}, and M_{01} are the first order moment, which is calculated from the image intensity $I(x, y)$ as following.

$$M_{00} = \sum_x \sum_y I(x,y) \quad M_{10} = \sum_x \sum_y xI(x,y) \quad M_{01} = \sum_x \sum_y yI(x,y)$$

Likely, the second order moment is calculated by the following equations.

$$M_{20} = \sum_x \sum_y x^2 I(x,y) \quad M_{02} = \sum_x \sum_y y^2 I(x,y)$$

By calculating the second order moment of the area, we obtain the long-axis a and short-axis b of the ellipse with following equation.

$$a = \sqrt{6(p + r + \sqrt{q^2 + (p-r)^2})}$$

$$b = \sqrt{6(p + r - \sqrt{q^2 + (p-r)^2})}$$

Where,

$$p = \frac{M_{20}}{M_{00}} - x_c^2 \quad q = 2\left(\frac{M_{20}}{M_{00}} - x_c y_c\right) \quad r = \frac{M_{02}}{M_{00}} - y_c^2$$

The parameters of the ellipse such as shape, aspect ratio between short and long axes, and the size of area are used to decide the areas of hand and face.

Following the rough search of hand and face shape, we designed a feature classifier to perform fine segmentation. This feature classifier aims to find reliable

hand and face areas based on an eigenspace [21] where the hand and face shape information are used to train this classifier. This eigenspace includes subspaces such as face subspace, one-hand subspace, and two-hand subspace. Each of the subspaces is created by using hand and face images from a variety of sample images as shown in Figure 2.

Figure 2. Image samples of faces and hands used for eigenspace-based hand segmentation

Lower resolution images are used to train the subspaces because we only need to determine if an area is that of face or hand. For each subspace, *k* eigenvectors with the largest eigenvalues are computed by the Karhunen-Loeve transform [22] from image samples. The detected skin-color area is then projected onto these subspaces, and one cluster from these subspaces is determined based on the maximum likelihood algorithm. If the likelihood of a cluster in a skin color area with regard to one of subspaces exceeds a threshold, this skin-color area will be classified as a hand or face area.

To validate the use of blob analysis of the head and hands across multiple skin tones we accessed two video repositories: the Mock Theft Experiment [23, 24] and the Airport Scenarios. In the Mock Theft Experiment, some participants played the role of a thief who stole a wallet while others where simply present during the theft. All participants were then interviewed to discover who had taken the wallet and these interviews were recorded.

In the Airport Scenarios four actors were hired to assist in performing a proof-of-concept study to determine the feasibility of identifying behavioral states from gestures and body movement. They participated in scenarios designed to simulate airport screening procedures. Within each scenario, each actor was asked to demonstrate three states: relaxed, agitated, and over-controlled.

Both repositories contain interview style conversations and these interviews were subjected to blob analysis. Table 1 shows the summary of the two data sets, and the number of color sample images.

In the validation using these two data sets, we trained the 3-D color table using a relatively small number of samples (376 and 85 frames respectively) of the face and hand images. The sample images are selected and cropped from video frames manually, as shown in Figure 2. Using these samples, a total of 15,954 and 52,809 video frames from the two data sets were processed successfully. Moreover, these results also showed that the method can detect a variety of skin colors of different ethnicities successfully. Figure 3 shows four sample frames on which blob analysis was successful.

Table 1. Summary of experimental data sets

	Mock Theft Exp.	Airport Scenarios
Total length of video clips	527 seconds	1,850 seconds
Ethnicity	Caucasian (13), Latin American(2), African American (2), Asian (1)	Caucasian (2), Latin American(1)
Number of subjects	18	3
Number of samples	376	85
Number of frames	15,954	52,809

Figure 3. Sample frames from blob analysis validation

3.3 Tracking hands and face as blobs

After obtaining hand and face regions from an image sequence using color a 3-D LUT and shape detection, the approach computes blobs using the extracted areas from image sequence and then tracks the left hand, right hand,

and face continuously. In order to separate hands when they overlap the face, blobs are computed not only for a single frame image but also for a time differential image between continuous frames [14]. First, the single frame image and the time differential image are binarized. Next, connected regions that are identified from the shape detection process are labeled as either a static or motion blob based on the blob's movement. The assignment is completed by predicting the location of each hand using a Kalman filter [25]. Then, the nearest blob to the predicted location is assigned to either the left or right hand, and is labeled as "hand blob" and another blob as "face blob". After the hand blobs and face blob have been labeled, the system updates the observed location of each of the blobs. If a hand blob exists in the current frame, the observed location is updated using the current hand blob. However, if the hand stops in front of the face, no hand blobs originating from the hand appear. Here, the observed location is updated from the location of the last hand blob because this situation is caused by the self occlusion of hand and face.

By using a Kalman filter, a dynamic process can be used to track the location of the hand or face blob on the image plane with the state vector $\boldsymbol{x}$ which includes its position and velocity. The dynamic process is defined as,

$$x_{k+1} = F_k x_k + G_k w_k ,$$

where,

$$F_k = \begin{pmatrix} 1 & \nabla t \\ 0 & 1 \end{pmatrix}, \quad G_k = \begin{pmatrix} \nabla t^2 / 2 \\ \nabla t \end{pmatrix}$$

The system noise is modeled via w_k, an unknown scalar acceleration. An observation model is given by

$$z_k = H x_{k+1} + v_k ,$$

where $H = [1,0]$, x_{k+1} is the actual state vector at time *k+1, v* is measurement noise, and z_{k+1} is the observed location at time *k+1*. The noise covariances can be determined by experiments so the system can perform optimal tracking.

Based on the continuous blob tracking, one can easily determine the hand-touch-face gesture when two or three blobs join together and become one bigger blob. However, to track the movements of hands continuously, the self occlusion problem caused when the hand enters the range of face needs to be addressed. Because the hands and face have similar values in the 3-D LUT, using only color will not differentiate the hands from the face. When the hand is moving and the face is still, the motion blob (hand) and static blob (face) can be separated based on their motion state [25]. However, when the hands and face have similar motion, the method fails to separate them. Here, the eigen-based shape detection described above will help to distinguish the hand from face region. At the predecessor frame at which the hand and face are separate, we calculate its eigen values *E(p)* using K-L transformation from the extracted hand region. When detecting the hand touching the face, i.e. the hand and face blob merging into one blob (hand-face blob), we search for the hand blob within the region of hand-face blob. We calculate the eigen value *E(i)* of a small area through the region of hand-face blob, and compare *E(i)* and *E(p)*. The location having the maximum likelihood between *E(i)* and *E(p)* is labeled as the position of hand blob.

In this paper, we only used the displacement and velocity of blobs for behavioral state estimation. In addition to these features, blobs can be used to analyze the motion of a human gesture. For instance, for a movement like a head lean is shown in Figure 4 (a). One can recognize this gesture by estimating the sequence of orientation of the head blob. The orientation of blobs is calculated as the degree between the Y-axis and long-axis of the head blob. Therefore, the head-leaning gesture can be described as a sequence, [a_0, a_1, a_2, a_3, ….]. Similarly, a gesture of nodding may be described as a sequence of blob shape and size, [(s_0, r_0), (s_1, r_1), (s_2, r_2), (s_3, r_3), ….] (Figure 4 (b)), where s_i is the blob size and r_i is the ratio of blob's short axis *(a)* and long axis *(b)*.

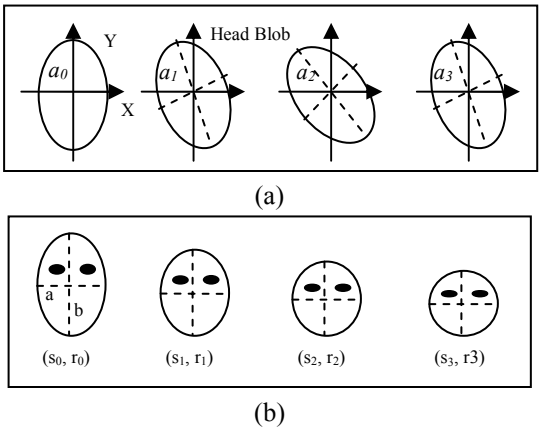

Figure 4. (a) A sequence of orientation of a head blob in a head-leaning gesture. (b) A sequence of blob size and blob shape in a nodding gesture

4. Estimating behavioral state from gestures

People use countless combinations of facial expressions and object manipulations in order to covey meaning. Although facial expressions have received much attention in the investigation of nonverbal communication, there is clear evidence that people

express and interpret others' behavioral and interpersonal states from such nonverbal cues as gestures, posture, gait and other physical movement [26]. The focus in this study is on those aspects of bodily expression covered primarily under kinesics, or movement of the body. This proof-of-concept study investigated the association between gestures and three behavioral states: relaxed, agitated, and over-controlled.

We used five interviews from the Mock Theft Experiment [23, 24] and Airport Scenarios to test our method for behavioral state estimation. Of the three interviews from the Airport Scenarios, two were deceptive and one was truthful. From the two interviews from the Mock Theft Experiment, one was truthful and one was deceptive.

4.1 Methods

Using the principles of IDT and EVT, a baseline of behavior was established for agitated, relaxed and over-controlled behavioral states using the Airport Scenarios. The video clips were manually segmented according to agitated, relaxed, and over-controlled displays, where the agitated and over-controlled displays were in the deceptive condition. The clips were subjected to blob analysis and resultant data from each video frame as well as the velocity of the hands' movements, the frequency of the hands touching the face, and the frequency of the hands coming together were recorded.

Our main goal in tracking the head and hands was 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.

Figure 5 illustrates two movement signatures, wherein (a) the subject is agitated as the result of deception and in (b) the subject is relaxed and telling the truth.

The blobs are indexed with the numbers 0 (for the head), 1 and 2 (for the two hands). 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 (blob 0 and 1).

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. Figures 6 and 7 illustrate two cases of blob merging (in two successive frames), for the events "hand on face" and "hands together," respectively, and can be easily seen how blob centers change positions rapidly on the image plane.

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.

After examining the videos of the training set, we found that the three behavioral states we want to recognize, i.e. "relaxed," "controlled" and "agitated" can be determined by the parameters described above. Thus, when a subject is controlled, there are only small blob displacements and the hands do not touch the face often. When a subject is relaxed, there are smooth hand movements and large displacements, and hands may touch the face often. Finally, when a subject is agitated, the hands move often and more abruptly, and they touch the face more often and with short duration. The *state* is formulated according to the following equation:

$$state = (W_1 F_1 + W_2 (F_2 + F_3)) F_0 \quad (1)$$

where F_1 is the variance of the head velocity V_{head}, i.e. $F_1 = var(V_{head})$, and $F_i = var(V_{hand(i)})/\ var(P_{hand(i)})$, $i = 1,2$, with $V_{hand(i)}$ and $P_{hand(i)}$ indicating a hand's velocity and position respectively.

Also, W_1 and W_2 are the weights with which head and hand parameters participate in the decision, and they are defined as:

$$W_1 = \frac{1}{\mathrm{var}^2(P_{head})} \quad (2)$$

where P_{head} is the position of the head, and

$$W_2 = \frac{1}{(f_{hand-face} + f_{hand-hand})} \quad (3)$$

where $f_{hand\text{-}face}$ is the frequency of a hand touching the face and $f_{hand\text{-}hand}$ is the frequency of the hands touching each other.

The weights defined in Equations (2) and (3) have the following meaning. From our observations, we could not tell whether a subject is agitated or relaxed from the head movements, which are usually rapid in these cases. Thus, when the head moves abruptly and often, we do not take it into consideration for our results. Also, the more often two blobs are merged into one, i.e. the more often the hands touch each other or a hand touches the face, the less

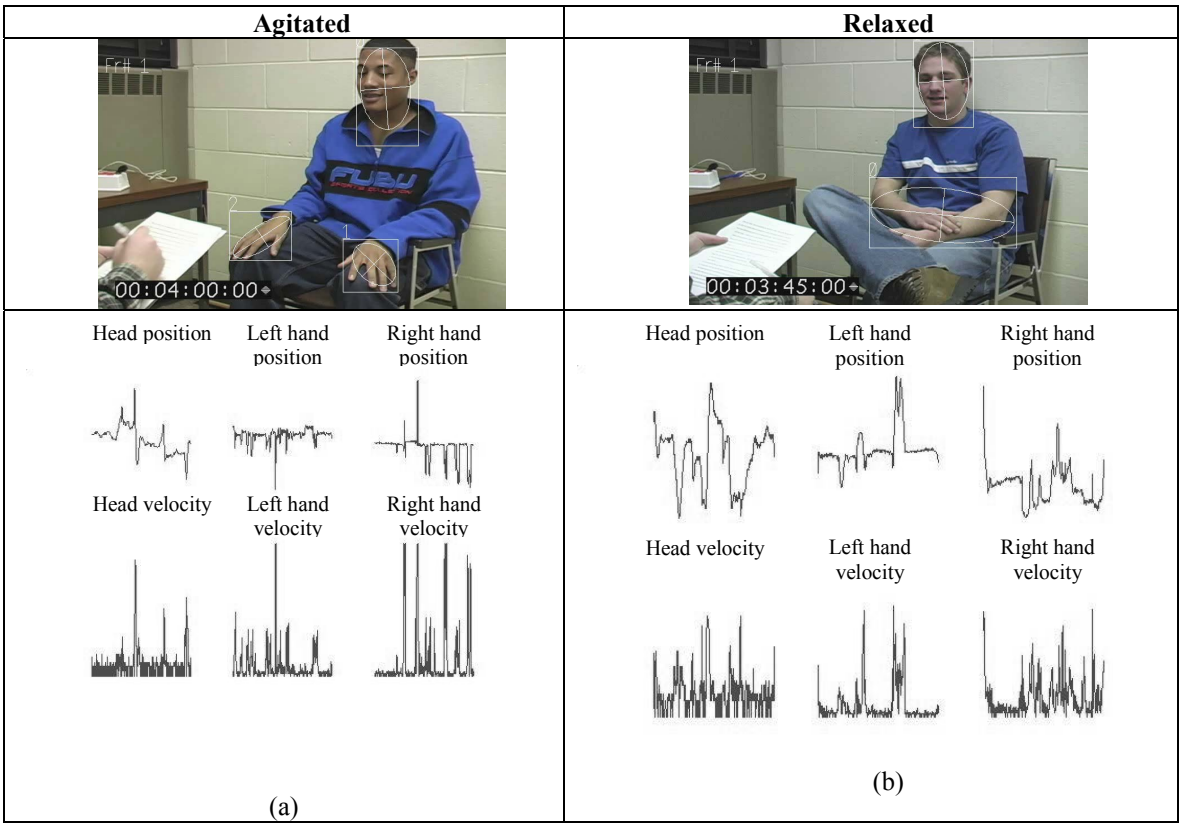

Figure 5. Example movement signatures based on blob position and velocity

Figure 6. Hand and head blobs merging

Figure 7. Hand blobs merging

information we have about the hand movements outside these events (time segments), and thus the respective positions and velocities are less useful.

Finally, the parameter F_0 is used as a normalization factor,

$$F_0 = \frac{f_{hand-face}}{D_{hand-face}} \quad (4)$$

where $D_{hand\text{-}face}$ is the duration (number of frames) of the event "hand on face". After normalization in the range between *0.0* and *1.0*, we can obtain the rough estimation of the state as shown in Table 2.

Table 2. Behavioral States of Actors

State	State Values
Controlled	*0.0<State < 0.2*
Relaxed	*0.2<State<0.7*
Agitated	*0.7<State <1.0*

4.2 Findings

The thresholds illustrated in Table 2 were then tested on subjects from the Mock Theft Experiment where

interviewees displayed relaxed and over-controlled behavior. The agitated interviewee was being deceptive while the relaxed interviewee was telling the truth. State was accurately determined for the interviewees using the state equation.

Table 3 shows the parameters extracted for the five interviews. The first two columns show the case examined and the respective total video duration in seconds. Columns 3-5 show the variance of the blob's position for the head and the hands, whereas columns 6-8 show the variance of the respective velocities. The next two columns give the number of times the events "hands on face" and "hands together" occur. The last column shows the maximum duration of any two merged blobs.

The final results for the subjects' states are shown in Table 4. The numbers in the third column directly indicate the behavioral state, using the thresholds illustrated in Table 2.

Table 3.
Blob features used for state recognition

Case	Total duration	Position change variance			Velocity variance			Hand on face (times / total duration)	Hands together (times / total duration)	Maximum duration (frames / total num of frames)
		Head	Hand (left)	Hand (right)	Head	Hand (left)	Hand (right)			
Agitated	115 sec	276.18	516.80	492.26	0.58	8.37	6.88	9.57	0.3478	1.57
Controlled	92sec	24.89	6.61	11.85	0.14	0.32	0.82	0	0.0217	3.11
Relaxed	68sec	260.13	303.05	104.83	6.08	5.80	0.57	2.94	0	13.31
Agitated	29sec	114.83	69.89	282.67	0.61	51.71	92.70	13.79	0	3.78
Relaxed	29sec	86.76	492.26	276.75	0.35	4.06	8.37	27.59	0.1034	21.11

Table 4.
Calculated behavioral states

A priori (state)	Subject	Result (state)
Agitated	Actor	0.84
Controlled	Actor	0.02
Relaxed	Actor	0.43
Agitated	Mock Theft Interview	0.76
Relaxed	Mock Theft Interview	0.53

4.3 Discussion

Clearly, automatically judging behavioral states from hand and head movement is very difficult. While this proof-of-concept study is simplistic in its approach to calculating behavioral states from video, it does show that such an approach may be possible. Five interviewees were automatically classified into over-controlled, relaxed, and agitated states based on their behavior.

The correct classification of five interviewees supports the belief that blob analysis is a useful and effective method for investigating nonverbal behavior related to deception. It offers a method for precise measurement of movement that is not easily measured by human observers. It offers flexibility in tracking a variety of skin colors and provides the ability to use automation in analyzing observed behavior.

Use of blobs may also provide the base for analysis that identifies transitions in behavioral state. In accordance with Interpersonal Deception Theory and Expectancy Violations Theory, the identification of such transitions may be a significant step forward in deception detection.

5. Future steps

Future efforts to expand our understanding and ability to detect deception will include the combination of multiple cues in a more robust model of behavioral state and a more comprehensive data set for establishing deceptive and truthful behavior.

We are currently working on using the data from the head and hands blobs in a more sophisticated scheme via

Hidden Markov Models (HMMs) [27, 28, 29]. Instead of using the "state" Equation (1), the weights defined in Equations (2), (3) and the factor defined in Equation (4), we use the blob parameters and the events ('hand on face' and 'hands together') as observations, and train the system for each one of the three states. Following this approach, we have three HMMs, one for each state. The more the training data we obtain, the more robust our system can become. This method provides us flexibility regarding the aforementioned thresholds and the system can adjust to variations in people's behaviors under similar behavioral states.

While blob analysis may be a useful approach in determining behavioral states, large hurdles exist for actual deployment of such a system. In order to screen and detect the behavioral state of people, a near real-time, automated system is necessary. In building a near real-time system, we face some serious challenges such as video-rate processing and response, minimum operator-interrupt, and automatic detection and recovery from failures. In our current experiments, the processing time of blob analysis reaches about 15 frames per second at a 320x240 resolution. Considering the improvement of computer technology, a faster processing rate may be expected along with higher image resolution.

Another issue confounding the creation of a near real-time system is the considerable effort required to create the training skin samples. This task becomes even more onerous when dealing with large numbers of people that would be present in a public area. We are currently exploring this problem and we believe that training using a combination of natural images and computer-synthesized samples may be a possible solution to this issue.

6. Conclusion

To ease some of the problems associated with assessing behavioral indicators of deception and behavioral state, the development of an automated system based on blob analysis of the head and hands is proposed. Automatically analyzing behavior can be accomplished when grounded firmly in accepted theory and empirical evidence. The proposed foundation for identifying behavior that is associated with deception is a combination of two theories from the human communication field and the approach has been initially explored in the proof-of-concept study investigating behavioral state. Although the proof-of-concept study presented here is a small first step, our approach shows promise in addressing the challenge of behavior analysis in deception detection and behavioral state identification.

7. References

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

[2] M. Zuckerman, B. DePaulo, and R. Rosenthal, "Verbal and nonverbal communication of deception," in Advances in experimental social psychology, vol. 14, L. Berkowitz, Ed. New York: Academic Press, 1981, pp. 1-59.

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

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

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

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

[7] 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.

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

[9] 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.

[10] J. K. Burgoon, "A communication model of personal space violations: Explication and an initial test," Human Communication Research, vol. 4, pp. 129-142, 1978.

[11] D.M. Gavrila, "The Visual Analysis of Human Movement: A Survey", Computer Vision and Image Understanding, Vol.73, No.1, pp.82-98, 1999.

[12] Ying Wu and Thomas S. Huang, "Vision-Based Gesture Recognition: A Review", International Gesture Workshop, GW'99, pp.103-115, Gif-sur-Yvette, France, March 1999.

[13] Thomas B. Moels and Erik Granum, "A Survey of Computer Vision-Based Human Motion Capture", Computer Vision and Image Understanding, Vol.81, No.3, pp.231-268, 2001.

[14] K. Imagawa, S. Lu, and S. Igi, "Color-Based Hands Tracking System for Sign Language Recognition," presented at Proceedings of 3rd International Conference on Automatic Face and Gesture Recognition, 1998.

[15] S. Lu, D. Metaxas, D. Samaras, and J. Oliensis, "Using Multiple Cues for Hand Tracking and Model Refinement," presented at IEEE CVPR 2003, Madison, Wisconsin, 2003.
[16] M.J. Swin and D.H. Ballard, "Color Indexing", International Journal of Computer vision, Vol.7, No.1, pp11-32, 1991.
[17] K. Kanamori, H. Kotera, O. Yamada, H. Motomura, R. Iikawa, and T. Fumoto, "Fast Color Processor with Programmable interpolation by Small Memory (PRISM)", Journal of Electronic Imaging, Vol.2, No.3, pp.213-224, 1993.
[18] T. Gevers, A. W. Smeulders, "Color-based Object recognition", Pattern Recognition, pp.453-464, 1999.
[19] M. Jones and J. Rehg, "Statistical Color Models with Application to Skin Detection," Tech-Rep. CRL 98/11, Compaq Cambridge Research Lab, 1998.
[20] Berthold Klaus Paul Horn, "Robot Vision", The MIT Press, 1986.
[21] M. Turk and A. Pentland, "Eigenfaces for Recognition," Journal of Cognitive Neuroscience, Vol.3, No.1, 1991.
[22] M.A. Turk and A.P. Pentland, "Face Recognition using eigenfaces", Proc. Of IEEE Conference on Computer Vision and Pattern Recognition, pp.585-591, June, 1991.
[23] 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," presented at Annual Meeting of the National Communication Association, Miami Beach, Florida, 2003.
[24] J. K. Burgoon, J. P. Blair, T. Qin, and J. F. Nunamaker, "Detecting Deception Through Linguistic Analysis," presented at NSF/NIJ Symposium on Intelligence and Security Informatics, 2003.
[25] R.E. Kalman, "A New Approach to Linear Filtering and Prediction Problems," Transactions of the ASME, Journal of Basic Engineering, Vol.82D, No.1, pp.35-45, 1960.
[26] J. K. Burgoon, D. B. Buller, and W. G. Woodall, Nonverbal Communication: The Unspoken Dialogue. New York, New York: HarperCollins, 1989.
[27] J. Yang, Y. Xu, and C. S. Chen, "Human Action Learning via Hidden Markov Model," IEEE Trans. On Systems, Man and Cybernetics, Part A: Systems and Humans, vol. 27(1), pp. 34 – 44, January 1997.
[28] 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 7-8, 2000.
[29] A. D. Wilson, and A. F. Bobick, "Realtime Online Adaptive Gesture Recognition," International Workshop on Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems, Corfu, Greece, September 26-27, 1999.

9. Acknowledgements

Portions of this research were supported by funding from the U. S. Air Force Office of Scientific Research under the U. S. Department of Defense University Research Initiative (Grant #F49620-01-1-0394) and Department of Homeland Security - Science and Technology Directorate under cooperative agreement NBC2030003. The views, opinions, and/or findings in this report are those of the authors and should not be construed as an official U.S. Government position, policy, or decision.