Lie To Me: Deceit Detection via Online Behavioral Learning

Nisha Bhaskaran Ifeoma Nwogu Mark G. Frank Venu Govindaraju
Center for Unified Biometrics and Sensors (CUBS)
University at Buffalo, SUNY
[nb32, inwogu, mfrank83, govind]@buffalo.edu

***Abstract*— Inspired by the the behavioral scientific discoveries of Dr. Paul Ekman in relation to deceit detection, along with the television drama series *Lie to Me*, also based on Dr. Ekman's work, we use machine learning techniques to study the underlying phenomena expressed when a person tells a lie. We build an automated framework which detects deceit by measuring the deviation from normal behavior, at a critical point in the course of an investigative interrogation. Behavioral psychologists have shown that the eyes (via either gaze aversion or gaze extension) can be good "reflectors" of the inner emotions, when a person tells a high-stake lie. Hence we develop our deceit detection framework around eye movement changes. A dynamic bayesian model of eye movements is trained during a normal course of conversation for each subject, to represent normal behavior. The remaining conversation is broken into sequences and each sequence is tested against the parameters of the model of normal behavior. At the critical points in the interrogations, the deviations from normalcy are observed and used to deduce verity/deceit. An analysis on 40 subjects gave an accuracy of 82.5% which strongly suggests that the latent parameters of eye movements successfully capture behavioral changes and could be viable for use in automated deceit detection.**

I. Introduction

Lie to Me is a compelling American television drama series inspired by the scientific discoveries of Dr. Paul Ekman, a real-life psychologist who studies reading clues embedded in the human face, body and voice, to expose the truth and lies in different high-stake settings including criminal investigations.

Behavioral psychologists generally believe that deception causes physiological reactions such as high blood pressure, increased heart rate, and an increased respiration rate. The physiological reaction is the consequence of arousal that is associated with high-stake deception [5]. In [21], the authors suggest that the behavioral cues to deceit differ in low- and high-stakes situations, i.e. the nervous behaviors manifested or *leaked* in people telling lies when the stakes are high are different from when they are low, and high-stake cues are more readily detected. A high-stake lie is one told when the person lying stands to get a notable gain, or faces a notable loss by telling the truth.

For facial analysis, Ekman recommends looking at the eyes and the upper half of the face to detect deceit, since this where the true emotion might leak out [3]. "We are more adept at controlling the lower halves of our faces", says Ekman, "probably because this is where our mouth is and where our speech comes from". Guided by the behavioral studies of Ekman and Frank, we explore lie detection via cues "leaked" from the eyes.

In this paper, using machine learning techniques, we explore the latent parameters of a dynamic model of the movement of the eyes, created while a subject is engaged in a conversation that could potentially result in a high-stake lie being told. We observe the patterns generated at different phases of the conversation (an interrogation), we correlate them with verity/deceit detection, and we present our findings.

A. Related works in automated deceit detection

Some of the recent advances in automated verity/deceit decision-making include computer-based linguistic analysis for deception detection [2] which used decision trees to show that deceivers often displayed higher quantity of information and expressiveness and use less vocabulary and grammar. Zhang, Singh *et al*. [20] proposed an automated deceit detection system from involuntary facial expressions, measuring groups of facial action units in static images. The major challenges in this approach were (i) labor intensive manual tagging of the facial action units in static images, (ii)high computational costs and (iii) accuracy scores that could be compared against were not reported by the authors. Another pattern recognition based decision classifier was based on thermal imaging analysis [19] in which body heat changes were measured via thermal signals while subjects were being interviewed. Lastly, an automated deceit detection technique was developed by Nwogu et el. [12] where the authors extracted several features such as blink rate, gaze duration etc. and manually clustered the features. They subsequently tagged each cluster as truth or deceit and reported a deceit detection accuracy of only 64.28%.

Although there has been much research work in systemically detecting deceit in the behavioral psychology community, automation of the deceit detection processes is still a budding research area. This paper presents a joint research project where the inputs to the framework and the evaluated outputs were identified from the ongoing behavioral research work of Prof. CoAuthor.

II. Method

In this section, we provide an outline of the experiment conducted for data collection for the automated human behavior analysis system. The output of the experiment is a collection of interview videos along with a ground-truth

spreadsheet indicating whether the subject in each video was telling the truth or not. We also present the tracking and feature extraction techniques applied to the collected data in order to prepare it for deceit detection analysis. A dynamic model is constructed for each subject from a portion of the interview and the verity/deceit patterns are observed from the rest of the interview conversation.

A. Data Collection

The psychological experiment was designed by the Communication Laboratory of Prof. CoAuthor. He selected 132 subjects with a firm adherence to a party, faction or cause and were informed that the skill under examination for this experiment was deception. They were then told that there was a check made out to the group they adamantly opposed (e.g., democratic party subject with check made out to republican party, pro-life to pro-choice, etc.), and they were to decide to "steal" it or not. To increase the stakes of the experiment, there were penalties (money donated to the despised group) for either stealing the check and getting caught or not stealing it at all. Similarly, there were rewards (money donated to adherence group) for stealing the check and not getting caught. Lastly, if a subject did not steal the check but led the interrogator into thinking that he/she did, a small amount of money was donated to the despised group. The goal is to provide a natural setting where the subjects were encouraged to tell a relatively high-stake lie.

The subjects were informed that they would be interrogated afterwards by lie detection experts (retired FBI interrogators), to determine whether they lied or not about stealing the checks. These interviews/interrogations involved two types of questions: (1) questions involving basic conversations that had nothing to do with the check (**baseline questions**); and (2) questions involving the check removal and punishment for stealing (**critical questions**). In each interrogation, one critical question (for example, "did you steal the check" or "what did you do with the check") marked the end of the baseline period and the beginning of the critical period.

Lastly, the subjects filled out questionnaires to record ground truth (whether they stole the check or not and whether they lied or not about it). The questionnaires also requested they specify any strategies they implemented to maintain credibility during interrogation. Their strategies were recorded along with the ground truth. Hence, the inputs to our analysis were (i) the resulting interrogation videos (ii) the ground truth spreadsheet.

B. Analysis

The first step to processing the collected data is to manually identify and record the critical point, the point in time where the critical questioning began and the baseline conversation ended, in each video. Next, the two eye regions are identified and tracked via an incremental learning tracker; the eye pupils are located using low-level image processing techniques and the x- and y-locations of the pupil regions are recorded as the behavioral features. A dynamic Bayesian model is then trained on the features obtained at the start of the baseline period of the interrogation. The rest of the interrogation is subsequently divided into equal chunks of time and each chunk is tested against the trained model by computing its log-likelihood. The log-likelihood is plotted against time and the chunk containing the critical point is examined.

Justification for the analysis: Kim et al. [9] showed that in a high-stake deception, liars displayed significantly less gaze aversion than truth-tellers immediately after the critical point in an interrogation. Similarly, Fukuda showed that eye blink could aid in deceit detection [6]. Motivated by their findings, we selected the eye region, specifically the location of the pupils (a coarse measure of gaze change and blink rate) as our feature to track. Unlike their works, *we are not interested in explicitly measuring gaze direction or blinks, rather we are interested in measuring physiological manifestations or leaks through eye movements, that might occur at the point when a person is suddenly asked a critical question, after a substantial period of normal or baseline conversation*. The purpose of our analysis is to develop a model for normal behavior (trained during the baseline) and then intermittently test portions of the rest of the conversation against the model parameters, observing any significant deviations from normalcy, and paying special attention to the critical point.

1) Incremental Learning tracker: In our problem, tracking the eye regions is especially challenging because of the relatively small size, its appearance variability over time as well as illumination changes and occlusions due to blinking, gestures or head pose variations. We implemented an extension of an incremental learning tracker [16], a probabilistic tracker based on incremental principal component analysis.

We sampled the videos at 30fps to generate the set of images given as input to the tracker. The tracker is cast as an inference model with hidden state variables. The location of the target object at any time t is represented by the state variable X_t. Using the set of images I_t we have sampled from the videos, we estimate the value of the hidden state variable using Bayes theorem:

$$p(X_t|I_t) \propto p(I_t|X_t) \int p(X_t|X_{t-1})p(X_{t-1}|I_{t-1})dX_{t-1} \quad (1)$$

The tracking between two states is governed by two models, the **observation model** $p(I_t|X_t)$ and the **dynamic model** $p(X_t|X_{t-1})$.

We located the initial target position, i.e. the eye regions in the first frame. The regions were then tracked in consecutive frames by a method similar to the condensation algorithm. The condensation algorithm sampled the most probable (appearance based) points of the region for the next frame and selects the point with the maximum likelihood value. The two models which govern tracking are described below.

- The dynamic model

The patch defined by the eye region in a frame is represented by the state variable X_t. The state at any time t consists of six parameters $X_t = (x_t, y_t, \theta_t, s_t, \alpha_t, \phi_t)$ denoting x and y translation, rotation angle, scale, aspect ratio and the skewness at time t. Each parameter in X_t is modeled by a Gaussian distribution around its counterpart in X_{t-1}, so that the motion between frames is an affine transformation, and

$$p(X_t|X_{t-1}) = \mathcal{N}(X_t; X_{t-1}, \Psi) \tag{2}$$

where Ψ is the diagonal covariance matrix whose elements are the corresponding variances of the affine parameters.

- The observation model
  The image observations are modeled using a probabilistic interpretation of principal component analysis, such that given an image patch I_t predicated by X_t, we can assume that I_t was generated from a subspace of the target object spanned by U and centered at μ. The probability of a sample being generated from this subspace is inversely proportional to the distance d from the sample to the μ of the subspace. The likelihood of a sample being generated from the subspace is governed by

$$p(I_t|X_t) = \mathcal{N}(I_t; \mu, UU^\top + \epsilon E)\mathcal{N}(I_t; \mu, U\Sigma^{-2}U^\top) \tag{3}$$

  μ is the center of the subspace, Σ is the matrix of singular values corresponding to the columns of U and E is the identity matrix.
  Hence given a drawn particle X_t and a corresponding image patch I_t, the aim is to compute $p(I_t|X_t)$ iteratively using the eigenbasis U, initialized to be empty but incrementally updated based on the next most likely particle.

The most likely eye region (patch with maximum likelihood) in each frame is subjected to adaptive histogram equalization to enhance the image contrast for further processing. Circular hough transform is then applied to detect the iris in the eye region of the preprocessed image and the centroid of the iris is returned. When the subject blinks, the transform returns a null value and the x- and y- coordinates of the centroid are assigned a *special value*.

2) Model Training: The x- and y- coordinates of the centroid of each iris is the 4-dimensional feature vector, observed for every frame extracted from the interrogation videos. Hence our observed features are continuous and infinite. The basic assumption is that the information from the eye [features] can be approximated by a set of discrete states and the dependency between them can also be accounted for. A flat topology with 4 hidden states (determined empirically) is chosen and the output distribution densities are modeled as mixture of two Gaussian distributions (determined empirically after testing). Some properties of the dynamic bayesian model, shown in Figure 2, are given as:

- The number of distinct states N=4 (chosen empirically) where the individual states are given as (S_1, S_2, S_3, S_4); and the state at time t is given by q_t;
- The infinite and continuous observation symbols (generated by all the distinct states) are denoted by $\mathbf{V} = \{\mathbf{v}|\mathbf{v} \in \mathbb{R}^4\}$;
- The state transition probability distribution $\mathbf{A} = \{a_{ij}\}$:

$$a_{ij} = P(q_{t+1} = S_j|q_t = S_i), 1 \le i, j \le N \tag{4}$$

- The observation probability distribution in state S_j, $\mathbf{B} = \{b_j(\mathbf{v})\}$, is defined as

$$b_j(\mathbf{v}) = P(\mathbf{o_t} = \mathbf{v}|q_t = S_j), 1 \le j \le N \tag{5}$$

  $\mathbf{o}_t$: observation at time t
  $b_j(\mathbf{v})$ is a mixture of two Gaussian distributions.
- The set of observed sequence $\{\mathbf{o}_i \in \mathbf{O}\}$

The model parameters can be represented compactly as $\lambda = (\mathbf{A}, \mathbf{B})$.

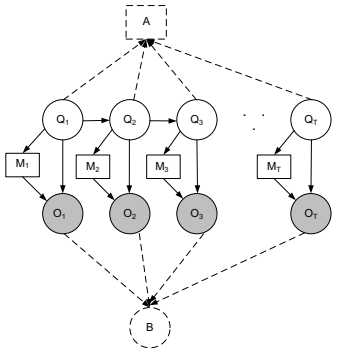

Fig. 1. Our dynamic model with mixture of gaussian output distributions

Hence, given an observed sequence $\mathbf{O}$, the parameters of the dynamic model can be trained by determining the parameters which maximize the probability $p(\mathbf{O}|\lambda)$. The dynamic model is therefore trained by expressing $b_j(\mathbf{O})$ in terms of every mixture component $b_j(\mathbf{o_t})$. Hence given a sequence of T observed values and an observation probability distribution of M mixtures of Gaussian, the problem can be redefined as:

$$P(\mathbf{O}, \mathbf{Q}|\lambda) = \left\{\prod_{t=1}^{T-1} a_{q_t q_{t+1}}\right\}\left\{\prod_{t=1}^{T} b_{q_t}(\mathbf{o}_t)\right\}$$
$$= \left\{\prod_{t=1}^{T-1} a_{q_t q_{t+1}}\right\}\left\{\prod_{t=1}^{T}\sum_{k_t=1}^{M}\left[c_{q_t k_t} b_{q_t k_t}(\mathbf{o}_t)\right]\right\} \tag{6}$$

For $\mathbf{K}$: one of the possible mixture component sequence and $\mathbf{Q}$ the state sequence,

$$P(\mathbf{O}, \mathbf{Q}, \mathbf{K}|\lambda) = \left\{\prod_{t=1}^{T-1} a_{q_t q_{t+1}}\right\}\left\{\prod_{t=1}^{T}\left[c_{q_t k_t} b_{q_t k_t}(\mathbf{o}_t)\right]\right\} \tag{7}$$

Finally,

$$P(\mathbf{O}|\lambda) = \sum_{\mathbf{Q}}\sum_{\mathbf{K}} P(\mathbf{O}, \mathbf{Q}, \mathbf{K}|\lambda) \tag{8}$$

Using the Expectation-Maximization (EM) algorithm, the model parameters $\overline{c_{jk}}, \overline{\mu_{jk}}, \overline{c\Sigma_{jk}}$ use an auxiliary function of equation 8. $\overline{c_{jk}}, \overline{\mu_{jk}}, \overline{c\Sigma_{jk}}$ represent the estimated values of the coefficient of mixture k in state S_j, and the means and covariance of Gaussian component k for state S_j.This EM algorithm returns the maximum likelihood parameters for our dynamic model.

The dynamic model is trained from a portion of the data obtained during the initial baseline period part of the interviewing process. The statistics of the model are then applied (in chunks) to the rest of the interview including the critical period. The deviations from normalcy are tested via a computation of the log likelihood. Deceit and non deceit videos showed distinctive patterns. The experimental setup and pattern evaluation are explained along with our results in the next section.

III. Experimental setup and results

We selected 40 videos for our experiment based on different skin colors, head pose variation, illumination and occlusions such as glasses. The videos were converted into frames sampled at 30 fps. The eye regions were initialized in the first frame and subsequently tracked using the incremental learning tracker. The contrast of the eye region images were enhanced and the x-y coordinates of the centroid of the iris was located and recorded in all the frames. A 4-state dynamic bayesian model with two Gaussian mixtures was trained on data obtained from the first 15-20 seconds of the baseline conversation in the video. This represented the model of normal behavior. On every ensuing 7-second sequence in the rest of the video, we tested against the model parameters and recorded the resulting the log-likelihood. The log likelihood for each subject was then plotted against the time sequences. The behavior of the plots around the critical point was evaluated for each video. We observed that the plots for deceit and non deceit showed distinctive patterns. Finally, a threshold value was selected such that when the log-likelihood fell below, it indicated a strong deviation from normal behavior (according to the trained model).

Subject Number	Gender	Ground-truth	Prediction
S020	M	L	P
S021	M	L	P
S026	M	L	P
S029	M	L	P
S034	F	T	P
S038	M	T	P
S086	F	T	NP
S096	M	T	P
S041	M	L	P
S045	F	L	P
S063	F	L	NP
S087	F	L	P
S111	M	T	P
S176	F	T	P
S222	F	T	P
S261	F	T	P
S088	F	L	P
S092	F	L	P
S110	F	L	NP
S165	F	L	P
S262	F	T	P
S271	F	T	P
S278	F	T	P
S279	M	T	P
S203	M	L	P
S234	F	L	P
S235	F	L	NP
S245	M	L	P
S284	F	T	P
S327	M	T	NP
S336	F	T	P
S340	F	T	NP
S265	M	L	NP
S288	M	L	P
S291	M	L	P
S295	F	L	P
S341	F	T	P
S344	M	T	P
S321	F	L	P
S339	F	L	P
Success rate for true videos	=	83.33%	
Success rate for lie videos	=	81.82 %	
Overall success rate	=	82.50 %	

TABLE I

Overview of experimental results

L=lie; T=True; P=Correct Prediction; NP=Wrong PRediction

Observation: The deceit graphs dipped well below the threshold while the non deceit videos reported no such drastic dips at the critical point. The graphs were clamped at the log likelihood value of -13000, in order to display the results on the same scale for all evaluated subjects. As such, any video which dipped below the threshold at the critical point was classified as a deceit video, otherwise, it was non-deceit.

An overview of the experimental result is presented in Table I. The distinctive output plots for the deceit and non-deceit video samples are shown in Figures 2 and 3 respectively. The last plot in each figure shows an error case.

IV. Discussion

A. Limitations

The primary limitation of our technique is that it is not an all-purpose general lie detection technique. It is relies strongly on the nature of the interrogation where the interrogator must initially achieve normalcy in conversation (baseline), before the critical question is posed. Also, in tracking a subject's eye region, if there is heavy reflection due to the presence of glasses, or if the subject performs gestures (such as constantly flicking the hair with the hand), which prevents successful eye region tracking, the model features cannot be accurately captured. Lastly, we assume that a deviation from normalcy is equivalent to deceit, but this might not always be the case. A subject who feels agitated by the critical question might display similar patterns as a liar, although this issue is somewhat mitigated by the fact that the interrogator gets "comfortable" with the subject, before posing the critical question.

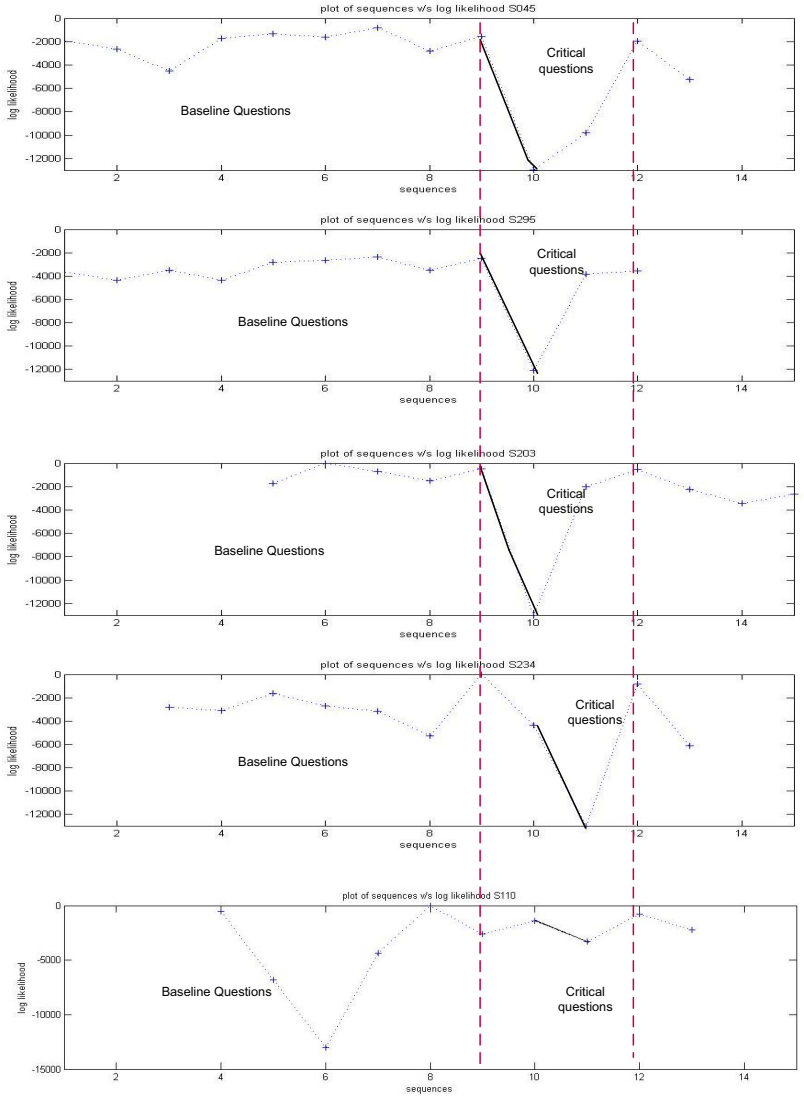

Fig. 2. The plots of subjects who lied in response to the critical question - their log-likelihoods dipped sharply right after the critical point. The last plot shows an example of an error

B. Conclusion

We have presented a working system and have shown results on real data (which is not easy to come by for this type of analysis). Although, our sample space is not yet statistically significant, the the preliminary results are very promising.

Our automated deceit detection system captures deviations from normal behavior and tags this deviation as deceit. Given a video, we provide a fully automated process to classify it into deceit and non deceit. The dynamic model captures a definitive distinguishing behavioral pattern in deceit and non deceit videos at their critical sections. The results obtained is compared with the ground truth and the accuracy of our system is 82.5% when tested on 40 subjects.

REFERENCES

[1] G. Bebis and K. Fujimura, "An Eigenspace Approach to Eye-Gaze Estimation," in *Intl. Conf Parallel and Distributed Comput. Systems, Las Vegas*, 2000.

[2] J. K. Burgoon, J. P. Blair, T. Qin, and J. F. Nunamaker, Jr., "Detecting deception through linguistic analysis," in *LNCS Proc. of Intell. and Security Informatics (ISI)*, 2008, p. 958.

[3] P. Ekman, *Telling Lies: Clues to Deceit in the Marketplace, Politics, and Marriage (Revised and Updated Edition)*. W. W. Norton & Company, 2001.

[4] I. R. Fasel, B. Fortenberry, and J. R. Movellan, "A generative framework for real time object detection and classification," *Computer Vision and Image Understanding*, vol. 98, no. 1, pp. 182–210, 2005.

[5] M. G. Frank and P. Ekman, "The ability to detect deceit generalizes across different types of high-stake lies." *J Pers Soc Psychol*, vol. 72, no. 6, pp. 1429–1439, Jun 1997.

[6] K. Fukuda, "Eye blinks: new indices for the detection of deception," *Int. Jo. of Psychophysiology*, vol. 40, no. 3, pp. 239 – 245, 2001.

[7] D. Hansen and A. Pece, "Eye tracking in the wild," *Computer Vision and Image Understanding*, vol. 98, no. 1, pp. 155–181, 2005.

[8] M. Isard and A. Blake, "Condensation C conditional density propa-

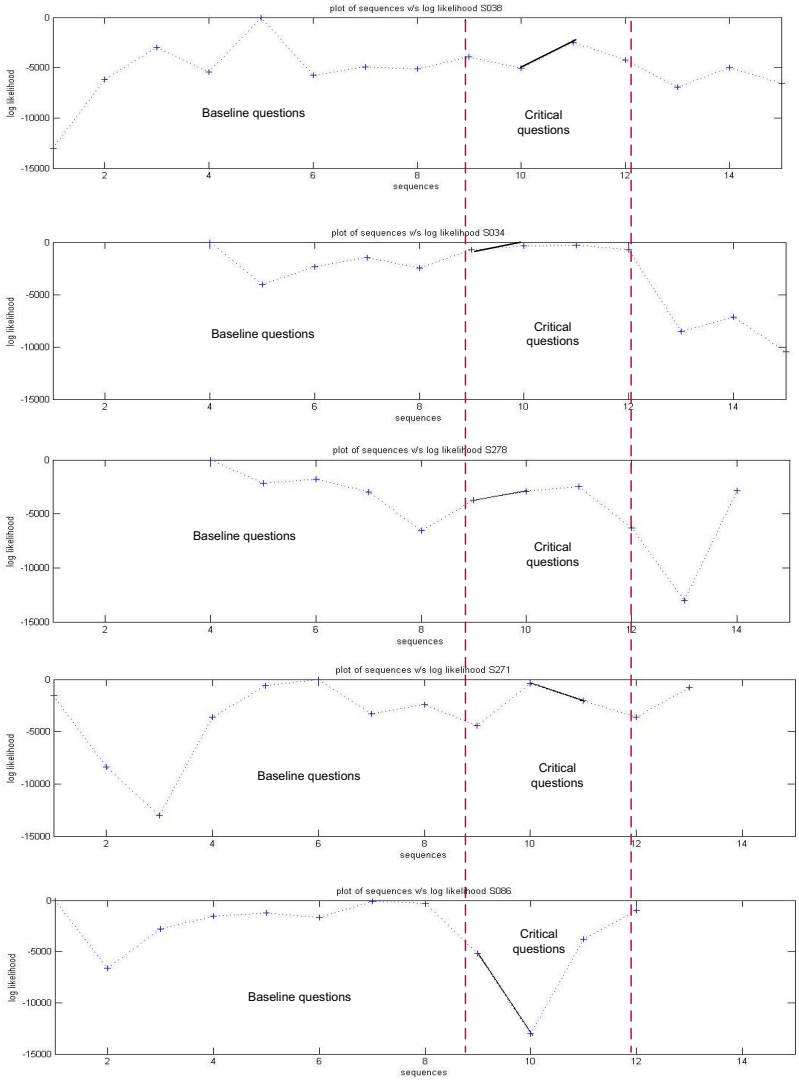

Fig. 3. The plots of subjects who did not lie in response to the critical question - their log-likelihoods did not show a dramatic deviation from normalcy at the critical period. The last plot shows an example of an error

gation for visual tracking," *Int. J. of Comp. Vis.*, vol. 28, no. 1, pp. 5–28, 1998.

[9] D. Kim, S. Kang, and M. G. Frank, "An Examination of Gaze Aversion in Different Conversational Phases During Deception," in *NCA 94th Annual Convention, San Diego, CA*, 2008.

[10] M. KP, "An introduction to graphical models," Intel Research, Tech. Rep., 2002.

[11] K. P. Murphy, *The hmm toolbox for MATLAB*.

[12] I. Nwogu, M. Frank, and V. Govindaraju, "An automated process for deceit detection," in *SPIE*, 2010.

[13] M. O'Sullivan, "The fundamental attribution error in detecting deception: The boy-who-cried-wolf effect," *Pers Soc Psychol Bull*, vol. 29, no. 10, pp. 1316–1327, October 2003.

[14] A. Prez, M. Crdoba, A. Garca, R. Mndez, M. Muoz, J. Pedraza, and F. Snchez, "A precise eye-gaze detection and tracking system," in *11th Int. Conf. in Central Europe of Comp. Graphics, Visualization and Comp. Vis., Czech Republic*, 2003.

[15] L. Rabiner, "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition," *IEEE*, vol. 17, no. 2, 1989.

[16] D. A. Ross, J. Lim, R.-S. Lin, and M.-H. Yang, "Incremental Learning for Robust Visual Tracking," *Int. J. Comput. Vis.*, vol. 77, no. 1-3, pp. 125–141, 2008.

[17] A. Salah, M. Bicego, L. Akarun, and a. M. T. E. Grosso, "Hidden Markov Model-based Face Recognition Using Selective Attention," in *SPIE Conference. on Human Vision and Electronic Imaging, San Jose, 2007*, 2007.

[18] P. Tsiamyrtzis, J. Dowdall, D. Shastri, I. Pavlidis, M. Frank, and P. Ekman, "Imaging Facial Physiology for the Detection of Deceit," *Int. J. of Comp. Vis.*, vol. 71, no. 2, pp. 197–214, 2005.

[19] P. Tsiamyrtzis, J. Dowdall, D. Shastri, I. Pavlidis, M. G. Frank, and P. Ekman, "Imaging facial physiology for the detection of deceit." *Int. J. of Comp. Vis.*, vol. 71, no. 2, pp. 197–214, 2007.

[20] S. Tulyakov, T. E. Slowe, Z. Zhang, and V. Govindaraju, "Facial Expression Biometrics Using Tracker Displacement Features," in *CVPR Workshop on Biometrics*, 2007.

[21] A. Vrij and S. Mann, "Telling and detecting lies in a high-stake situation: the case of a convicted murderer," *Applied Cognitive Psychology*, vol. 15, pp. 187–203, 2001.

[22] L. Zhang, R. Chu, S. Xiang, S. Liao, and S. Z. Li, "Face detection based on multi-block LBP representation." in *LNCS proceedings of the Int. Conference on Biometrics*, 2007, pp. 11–18.

[23] a. Q. J. Zhiwei Zhu, " Robust real-time eye detection and tracking under variable lighting conditions and various face orientations," *Computer Vision and Image Understanding*, vol. 98, no. 1, pp. 124–154, 2005.