High-Stakes Deception Detection Based on Facial Expressions

Lin Su
Centre for Intelligent Machines
McGill University
Montreal, Canada
linsu@cim.mcgill.ca

Martin D. Levine
Centre for Intelligent Machines
McGill University
Montreal, Canada
levine@cim.mcgill.ca

***Abstract*—During a forensic interview, high-stakes deception is very prevalent notwithstanding the heavy consequences that might result. This paper proposes an automated computer vision solution for detecting high-stakes deception based on facial clues. Four deceptive cues (eye-blink, eyebrow motion, wrinkle occurrence and mouth motion) were identified and integrated into a single facial behavior pattern vector for discerning deception and honesty. A Random Forest classifier was trained using an unconstrained video database and applied to classify facial patterns into either deceptive or truthful categories. The labeled database we created was based on open sources such as YouTube. The interview videos used for training and testing the classifier were selected on the basis of high-stakes criminal situations, such as murder or kidnapping, which were later verified by criminal trials. Despite the many uncontrolled factors (illumination, head pose and facial occlusion) in the videos, we have achieved an accuracy of 76.92% when discriminating liars from truth-tellers. This compares well with 80.9% [1], the best extant accurary obtained by experienced interrogators.**

Keywords—high-stakes; deception detection; facial expression; unconstrained database; Random Forest; classification; automated

I. INTRODUCTION

Deception is ubiquitous in the daily life of a person. But there are certain circumstances where lies are more likely to give rise to heavy consequences for both individuals and society [2]. These circumstances are termed *high-stakes situations*; for example, being interrogated by a police officer, defending oneself in a courtroom, or appealing for parole in front of a judge. To detect deceit in such high-stakes scenarios is even more demanding than in daily life, but people rarely have the capacity of being able to distinguish truth from lies because their judgment is easily biased [3]. Statistics show that the lie detection accuracy of untrained people is only slightly above chance [4]. Various technical methods, such as polygraphs, functional Magnetic Resonance Imaging (fMRI) and linguistic analysis, have been adopted to automate the process of detecting such deception. However, to the best of our knowledge, no computer vision research has attempted to discriminate high-stakes deception from truth using facial expressions.

This paper makes three contributions: (1) It sets a precedent for future research on high-stakes deception detection using facial clues. (2) The collected database is the first that consists of high-stakes deception videos of real-world situations. (3) The proposed method is at the forefront of the automated analysis of videos displaying facial expressions in completely unconstrained environments. The rest of this paper is organized as follows. Related work is firstly reviewed, followed by an introduction to the collected database. Then the theoretical basis, dynamic feature analysis and feature integration methods are presented. Finally the results are discussed.

A. Related Work

1) Deception Detection

To date, the focus of deception detection has been on three types of indicators: body language [5][6][7], gaze aversion [6] [7][8], and facial expression [6]. All of these used spontaneous video data.

The authors of [5] attempted to prove the theory that behavioral states are indicators of deception. They analyzed the position and velocity of face and hand blobs, classifying suspects into over-controlled, agitated and relaxed states. However, their work was not persuasive, since their experiments were based on a very small set of video data (18 subjects) collected from a low-stakes experimental environment. Additionally, study showed that most body cues occurred too rarely for statistical analysis [9].

In [8], deception was detected when eye behaviors (gaze direction and blink rate) deviated from a baseline determined for each individual. Nevertheless, according to DePaulo et al. [10], gaze aversion might have nothing to do with deception. Mann et al. [11] have also found that eye contact maintenance has no significant relationship to deception.

Other researchers have combined body language and facial micro-expressions to generate more convincing cues to deception [6]. (A micro-expression is considered to be a rapid and involuntary facial expression, which could reveal concealed emotions [12].) Head and hand movements were measured to quantify body language. For facial analysis, an Active Shape Model (ASM) was exploited to track mouth and eyebrow movements. This work achieved a high accuracy of 81.6%. However, again, their experimental data were collected in a low-stakes situation, where each of the 147 participants presented a deceptive and truthful opinion during an interview.

1051-4651/14 $31.00 © 2014 IEEE
DOI 10.1109/ICPR.2014.435

The Silent Talker presented in [7] used Artificial Neutral Networks (ANN) to discriminate deception and honesty based on four cues to deception: eye gaze, eye closure, head movement and blushing. Each cue was categorized into one of the defined discrete states. Then ANN was employed to integrate the cues to predict the emotional state, either deceptive or truthful. The authors achieved a classification accuracy of 79% based on their database, which was also collected from a low-stakes scenario in the laboratory.

In summary, the most prominent issue with past computer vision studies is that researchers rarely employed data obtained in *real* forensic circumstances. In this paper, an automatic deception detection system capable of providing valid predictions of lying in uncontrollable situations is proposed, and the decision classifier is based on video data captured in actual high-stakes situations.

2) Emotion Recognition

Current emotion recognition methods largely fall into two categories: appearance- and model-based. The former often involves image processing and feature extraction followed by machine learning classification techniques. Common face descriptors used in the literature include local binary patterns [13], Haar wavelets [14] and Gabor textures [15]. For the latter, a face model is first constructed based on facial landmarks, and the emotions are recognized by tracking spatial changes of the model. Typical face models include ASM [16] and CLM [17]. However, most emotion recognition methods lack the capacity to deal with unconstrained head pose, variable illumination and natural emotions.

B. Database

To date, there are only a few datasets that have been used in deception detection research, but *none* of these have been made available to the public and few of them were obtained in high-stakes situations. Therefore, we were required to collect our own database.

Courtesy of Professor Stephen Porter from the University of British Columbia-Okanagan in Kelowna, Canada, we have obtained a list of emotional pleaders who were publicly asking for help to find their missing relatives or the murderers that killed or dismembered them. (Professor Porter was unable to legally provide these videos but consented to send us a list of so-called pleaders.) Approximately half of the pleaders were later convicted of murdering the missing or dead person, based on conclusive evidence [9]. In addition, we gathered some videos absent from Porter's list, but containing circumstances that were similarly high-stakes. These situations were considered to be high-stakes if the suspects attempted to lie, and obviously the guilty ones were all liars. In total, we obtained 324 video clips: 51.23% of the clips contain guilty suspects, while 48.77% are innocent. The average length of the video clips is 20 seconds.

Different from most existing databases, ours is *completely uncontrived*: (1) The illumination conditions as well as the backgrounds are highly variable. (2) There is no constraint on the head pose and appearance of the suspects. (3) The facial expressions are completely natural, rather than being acted.

II. Background and Theoretical Basis

The theoretical basis of the proposed method can be dated back to 1872, when Darwin argued that some facial actions that are the most difficult to create voluntarily are also the hardest to voluntarily inhibit [18]. Recently, Stephen Porter and his team have proposed the possibility that purposely concealed emotions could be distinguished as inconsistencies among normal facial expressions [9][19]. The presence of concealed emotions is referred to as *emotional leakage*. Recently, Hurley and Frank [20] have also found that emotional leakage happens everywhere on the face and that facial countermeasures are very rare. According to Porter et al. [19], the specific emotions that can be used for distinguishing liars and truth-tellers in a forensic interview are sadness and happiness. Based on their findings and the Facial Action Coding System (FACS) [21], we have summarized the differences between truth-tellers and liars with regard to facial Action Units (AUs), as listed in Table 1.

In addition to emotional leakage, blinking can also be considered to be a clue to deception. Mann and Bull [22] have stated that suspects will blink less frequently when telling lies in high-stakes situations. Leal and Vrij [23] have found that the blinking patterns of liars and truth-tellers differ, in that liars show a decreased number of eye blinks when lying, followed by an increase. In [9], ten Brinke and Porter have also reported a higher blink rate observed in deceptive suspects. Therefore, blinking activity (AU45) could also be added to Table 1 as a cue for discerning deception and honesty.

To sum up, the following AUs are potential indicators for distinguishing liars from truth-tellers in high-stakes situations: AU1, AU2, AU4, AU12, AU15 and AU45 [19]. Each AU is related to the movement of a single facial muscle and can result in motion of a part of the face or appearance changes in a facial region. Also, multiple AUs can occur simultaneously. Table 2 summarizes these potential deception indicators (AUs), their associated facial movements and corresponding facial regions. Therefore, based on the psychological theories presented above, the proposed method aims to detect the AUs listed in Table 2 and use them to discern deceptive and honest suspects in high-stakes situations.

TABLE I. Differences Between Truth-Tellers and Liars

Emotion	Truth-Tellers	Liars
Sadness	AU1+4, AU15 (Genuine)	AU1 (Fake), AU2 (Fake), AU1+2 (Fake)
Happiness	NA	AU6+12 (Genuine), AU12 (Fake)

TABLE II. Potential Indicators of Deception

Action Unit	FACS Name	Facial Movement	Facial Region
AU1	Inner Brow Raiser	Horizontal wrinkles occur in the center of the forehead; inner eyebrows move up	Center of forehead; eyebrows
AU2	Outer Brow Raiser	Short horizontal wrinkles occur above the lateral portions of the eyebrows; outer eyebrows move up	Left and right forehead; eyebrows
AU4	Brow Lowerer	Vertical wrinkles occur between the eyebrows; partial or entire eyebrows are lowered	*Glabella*[a] area; eyebrows

AU12	Lip Corner Puller	Lip corners move up obliquely; may create or deepen *nasolabial* furrows	Mouth
AU15	Lip Corner Depressor	Lip corners move down obliquely; may create or deepen *nasolabial* furrows	Mouth
AU45	Blink	Eyelids close and open rapidly	Eyes

[a] The *glabellar* area is the region between the two eyebrows.

III. Dynamic Feature Analysis

A. Single Video Frame Preprocessing

After illumination compensation, PittPatt [24] was applied to locate three primary facial landmarks: left eye, right eye and nose base. The face was spatially normalized and nine facial regions were located according to an anthropometric face model [25], as shown in Fig. 1. No attempt was made to further improve these regions to take into account individual variations. In the following sections, feature analysis is conducted in each region of interest (ROI) in each frame of a single video clip.

B. Eye Blink Detection

An eye blink is a dynamic process with the eyelid closing and opening rapidly. A *blink event* involves a continuous process of eye closure, closed eye, and eye opening. Instead of treating blink detection as a conventional open-closed eye classification problem, we identify a blink event using *anomaly detection*, with non-blinking considered as the normal behavior. We obtained a likelihood heat map for each frame by applying the anomaly detection method in [26] to the eye ROI in a video. In order to binarize the likelihood maps and find the 'blink' frames, a global threshold for each video clip was determined by applying valley-emphasis thresholding [27] to the histogram of all likelihood values in the video. Based on this approach, a binary sequence was obtained from the likelihood sequence. This sequence was then assigned as the spatio-temporal feature descriptor for this eye ROI in the video.

C. Detecting Eyebrow Motion

Due to the potential existence of various natural head poses, we restricted our focus to upward and downward motion of the whole eyebrow, instead of treating the inner and outer corners separately. The eyebrow blob was firstly segmented by applying Otsu's method [28] to the eyebrow ROI using the L component of the $L^*a^*b^*$ color space. The displacement of the midpoint of the upper eyebrow contour from the horizontal middle line was measured as the eyebrow displacement value, as shown in Fig. 2. The displacement was tracked with time, producing a displacement curve, as shown in Fig. 3. A moving average filter was applied to the displacement curve to obtain a temporal baseline for the eyebrow motion. Then the difference

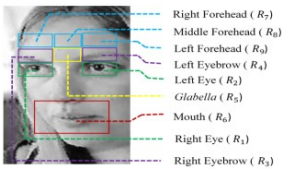

Fig. 1. Nine facial regions located on a spatially normalized face.

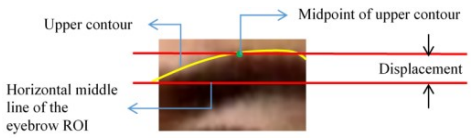

Fig. 2. Eyebrow displacement.

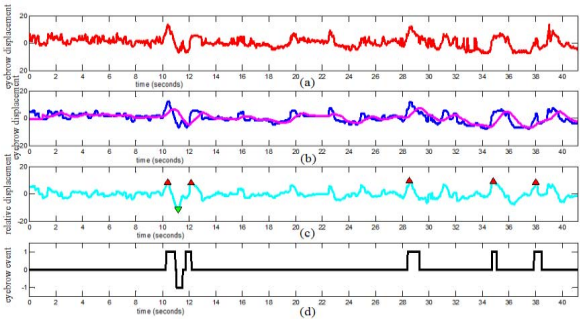

Fig. 3. Eyebrow motion detection. (a) Eyebrow displacement varies with time in a sample video clip. (b) The blue curve is the median filtered curve of (a) and the magenta is the moving average curve of the blue one. (c) Relative displacement curve. Red arrows: peaks higher than the threshold, green arrows: valleys. (d) Eyebrow event curve. The 1s indicate eyebrow raising, while −1s indicates lowering.

between the median-filtered displacement curve and its baseline was taken as the *relative displacement* of the eyebrow. Using the relative displacement curve, an *eyebrow raise event* was assumed if the height of the peak was higher than a threshold $\theta_{eyebrow}$. Similarly, an *eyebrow lower event* was located if the valley was lower than $-\theta_{eyebrow}$. Therefore, two binary feature descriptors were detected in the eyebrow ROI: *eyebrow raise event* and *eyebrow lower event*. The global threshold $\theta_{eyebrow}$ was chosen experimentally during the training process.

D. Detecting Wrinkles

Facial wrinkles are directional lines across the facial skin. Based on the anthropometric model, we focused on the horizontal forehead lines in three ROIs (R_7,R_8,R_9) and vertical *glabellar* frown lines in R_5. Since wrinkles in a predominant direction (either horizontal or vertical) need to be detected, oriented Gabor filters were employed to characterize the directional texture. The Gabor filter is defined by

$$gabor(x,y) = e^{-\frac{x'^2+\gamma^2 y'^2}{2\sigma^2}} e^{2\pi\frac{x'}{\lambda}+\varphi} \quad (1)$$

where $x' = x\cos\theta + y\sin\theta$, $y' = -x\sin\theta + y\cos\theta$ [29].

For horizontal forehead lines, three orientations were selected: $\varphi_i \in \{{}^3/_8\pi, {}^4/_8\pi, {}^5/_8\pi\}$; for vertical *glabellar* lines, we used $\varphi_i \in \{0, {}^1/_8\pi, {}^7/_8\pi\}$. Three frequencies $\lambda_j \in \{8,12,16\}$ were computed in each direction, thereby forming horizontal and vertical Gabor filter banks. Each directional bank contained nine filters. Suppose that the Gabor response in a certain direction and scale is denoted as $gabor(\varphi_i,\lambda_j)$. The total response of I directions and J scales is defined as:

$$GR = \frac{\sqrt{\sum_{j=1}^{J}\sum_{i=1}^{I} gabor(\varphi_i,\lambda_j)^2}}{I * J} \quad (2)$$

After computing the Gabor response, each ROI was represented by a Gabor-filtered response image. This image was then transformed into an entropy value indicating the overall strength of the edges:

$$Entropy = -\frac{\sum GR * \log(GR)}{L} \tag{3}$$

where L is the number of pixels in the image.

This provides a sequence of ROIs in a video clip with each frame having been characterized by a single entropy value. The *wrinkle event* can be detected by thresholding the entropy curve, as shown in Fig. 4. Each wrinkle ROI $R_i \in \{R_5, R_7, R_8, R_9\}$ in a video clip will have a binary sequence indicating where the wrinkle events have occurred. The threshold $\theta_{wrinkle}$ was also chosen during the training process.

E. Detecting Mouth Motion

The mouth ROI was used for detecting happiness (AU12: mouth corners pulling obliquely upwards) and sadness (AU15: mouth corners pulling obliquely downwards). Due to the variations in head pose and constant verbal utterances in the videos, traditional methods involving training a classifier on static images would be insufficient. Therefore, the mouth was initially segmented from the mouth ROI by thresholding the "pseudo-hue" [30] of the image:

$$PseudoHue = \frac{R}{R+G} \tag{4}$$

The feature points characterizing the mouth shape were located, as shown in Fig. 5. This actual event detector was based on a simple model of the expression behavior. A smile (AU12) was associated with an increase in the mouth angle ϕ, whereas for sadness (AU15) it was the opposite. In addition, there is an increase in the width W of the mouth for both AU12 and AU15. Therefore, the changes in $\phi(t)$ and $W(t)$ were measured simultaneously to determine AU12 and AU15. Thus AU12 was detected as a happiness event:

$$hp_{event} = hp_{\phi} \wedge hp_{W} \tag{5}$$

where $hp_{\phi} = \phi(t) > \phi_{hp}$ and $hp_W = W(t) > W_{hp}$, as shown in Fig. 6. Similary, AU15 was detected as a sadness event:

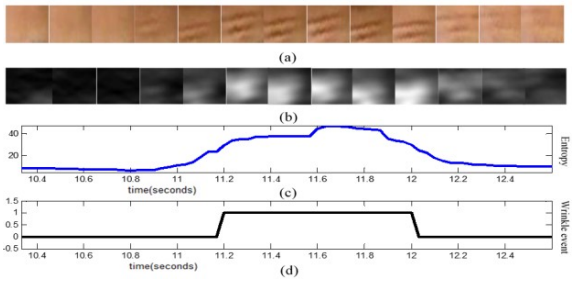

Fig. 4. Wrinkle detection. (a) A sequence of wrinkle ROIs taken from a video. (b) The corresponding Gabor responses. (c) Entropy curve for the frames in (b). (d) Binary sequence indicating where the wrinkle events occur by thresholding the entropy curve in (c).

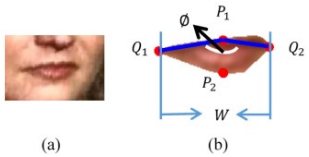

Fig. 5. Mouth segmentation. (a) The original mouth ROI. (b) Segmented mouth with mouth feature points located and measurements computed. ϕ is the angle between P_1Q_1 and P_1Q_2.

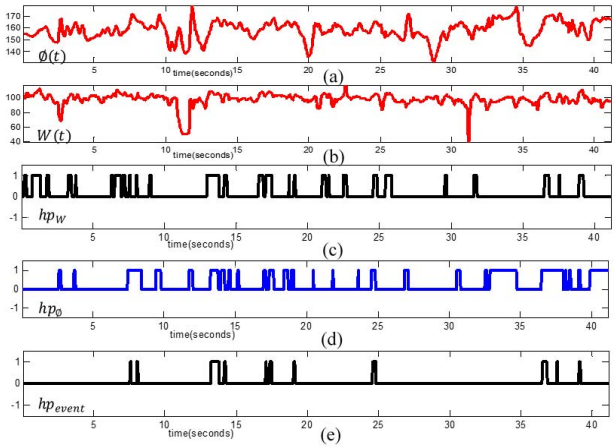

Fig. 6. Happiness event detection in a video. (a) Mouth angle curve $\phi(t)$. (b) Mouth width curve $W(t)$. (c) Happiness event candidates hp_W are detected where the width is larger than W_{hp}. (d) Happiness event candidates hp_ϕ are detected where the angle is larger than ϕ_{hp}. (e) Happiness event hp_{event} is determined by computing the logical conjunction of hp_W and hp_ϕ.

$$sd_{event} = sd_{\phi} \wedge sd_{W} \tag{6}$$

where $sd_\phi = \phi(t) < \phi_{sd}$ and $sd_W = W(t) > W_{hp}$.

Considering the individual differences in mouth shape and the influence of utterances and head pose on mouth scale and appearance, the thresholds ϕ_{hp}, W_{hp}, ϕ_{sd} were chosen as:

$$\begin{aligned} \phi_{hp} &= \text{mean}(\phi(t)) + \theta_{mouth} * \text{var}(\phi(t)) \\ W_{hp} &= \text{mean}(W(t)) + \theta_{mouth} * \text{var}(W(t)) \\ \phi_{sd} &= \text{mean}(\phi(t)) - \theta_{mouth} * \text{var}(\phi(t)) \end{aligned} \tag{7}$$

The threshold θ_{mouth} represents how much the angle and width deviate from an individual's baseline. However, we were unable to determine a universal formula or algorithm for it. Therefore, θ_{mouth} was selected based on experimentation with a training set.

IV. Feature Integration

A. Primary and Secondary Features

Denote the video database as $V = \{v_1, v_2, \ldots, v_N\}$. Each video clip v_α was decomposed into nine temporal sequences of specific regions on the face, $\mathcal{R}_\alpha = \{R_{\alpha,1}, R_{\alpha,2}, \ldots, R_{\alpha,9}\}$. For each region $R_{\alpha,\beta}$, one or two binary feature vectors (events) were computed using the methods discussed above. In total, 12 feature vectors were computed from the nine facial regions. Each feature corresponded to a facial Action Unit. Since the 12

feature vectors were computed directly from the facial regions and each involved only one region, we refer to them as *primary features*. The primary feature set of video clip v_α is denoted as $P_\alpha = \{p_{\alpha,1}, p_{\alpha,2}, \ldots, p_{\alpha,12}\}$ and their corresponding facial regions, Action Units and feature vectors are listed in Table 3.

Considering that some Action Units involve more than one facial region and multiple Action Units are likely to occur simultaneously, *secondary features* were also computed. For example, raising the left eyebrow involves both upward motion in the left eyebrow region ($p_{\alpha,4}$) and appearance change in the left forehead region ($p_{\alpha,12}$). Thus a secondary feature indicating the congruence of these events was generated by computing the logical conjunction of $p_{\alpha,4}$ and $p_{\alpha,12}$. Consequently, nine secondary features were also computed, resulting in a secondary feature set $S_\alpha = \{s_{\alpha,1}, s_{\alpha,2}, \ldots, s_{\alpha,9}\}$, as shown in Table 4.

After concatenating the 12 primary features and 9 secondary features, video clip v_α is given by a feature matrix denoted by:

$$\Omega_\alpha = [p_{\alpha,1}, p_{\alpha,2}, \ldots, p_{\alpha,12}, s_{\alpha,1}, s_{\alpha,2}, \ldots, s_{\alpha,9}] \quad (8)$$

TABLE III. PRIMARY FEATURES FOR VIDEO CLIP v_α

Action Unit	**AU45**	**AU45**	**AU1/2**	**AU1/2**	**AU4**	**AU4**
Event	right eye blinking	left eye blinking	right eyebrow raising	left eyebrow raising	right eyebrow lowering	left eyebrow lowering
Facial Region	$R_{\alpha,1}$	$R_{\alpha,2}$	$R_{\alpha,3}$	$R_{\alpha,4}$	$R_{\alpha,3}$	$R_{\alpha,4}$
Feature Vector	$p_{\alpha,1}$	$p_{\alpha,2}$	$p_{\alpha,3}$	$p_{\alpha,4}$	$p_{\alpha,5}$	$p_{\alpha,6}$

Action Unit	**AU12**	**AU15**	**AU4**	**AU2**	**AU1**	**AU2**
Event	mouth corners moving up	mouth corners moving down	wrinkle in *glabella* area	wrinkle in right forehead	wrinkle in mid-forehead	wrinkle in left forehead
Facial Region	$R_{\alpha,6}$	$R_{\alpha,6}$	$R_{\alpha,5}$	$R_{\alpha,7}$	$R_{\alpha,8}$	$R_{\alpha,9}$
Feature Vector	$p_{\alpha,7}$	$p_{\alpha,8}$	$p_{\alpha,9}$	$p_{\alpha,10}$	$p_{\alpha,11}$	$p_{\alpha,12}$

TABLE IV. SECONDARY FEATURES FOR VIDEO CLIP v_α

Action Unit	**AU45**	**AU1/2**	**AU4**	**AU1+2**	**AU1+2**
Event	eyes blink	eyebrows raising	eyebrows lowering	right eyebrow raising + wrinkle in right forehead	left eyebrow raising + wrinkle in left forehead
Facial Region	$R_{\alpha,1}$+ $R_{\alpha,2}$	$R_{\alpha,3}$+ $R_{\alpha,4}$	$R_{\alpha,3}$+ $R_{\alpha,4}$	$R_{\alpha,3}$+ $R_{\alpha,7}$	$R_{\alpha,4}$+ $R_{\alpha,9}$
Feature Vector	$s_{\alpha,1}$ $=p_{\alpha,1} \wedge p_{\alpha,2}$	$s_{\alpha,2}$ $=p_{\alpha,3} \wedge p_{\alpha,4}$	$s_{\alpha,3}$ $=p_{\alpha,5} \wedge p_{\alpha,6}$	$s_{\alpha,4}$ $=p_{\alpha,3} \wedge p_{\alpha,10}$	$s_{\alpha,5}$ $=p_{\alpha,4} \wedge p_{\alpha,12}$

Action Unit	**AU1+4**	**AU1+4**	**AU4**	**AU4**
Event	left eyebrow raising + wrinkle in mid-forehead	right eyebrow raising + wrinkle in mid-forehead	right eyebrow lowering + wrinkle in *glabella* area	left eyebrow lowering + wrinkle in *glabella* area
Facial Region	$R_{\alpha,4}$+ $R_{\alpha,8}$	$R_{\alpha,3}$+ $R_{\alpha,8}$	$R_{\alpha,3}$+ $R_{\alpha,5}$	$R_{\alpha,4}$+ $R_{\alpha,5}$
Feature Vector	$s_{\alpha,6}=p_{\alpha,4} \wedge p_{\alpha,11}$	$s_{\alpha,7}=p_{\alpha,3} \wedge p_{\alpha,11}$	$s_{\alpha,8}=p_{\alpha,5} \wedge p_{\alpha,9}$	$s_{\alpha,9}=p_{\alpha,6} \wedge p_{\alpha,9}$

of dimension $n_\alpha \times 21$, where n_α is the number of frames in video v_α. In this matrix, each column defines a feature vector for a specific event as time progresses and each row is a 1×21 feature vector for a single video frame.

B. Feature Temporal Volumes

We create a compact representation for each video frame, which incorporates its temporal context. As noted at the end of the previous section, a 1×21 binary feature vector denotes each frame. Considering a context of T consecutive frames centered at frame t, a Feature Temporal Volume (denoted as $FTV_{\alpha,t}$) can be obtained by *summing* the T feature vectors, as shown in Fig. 7.

In summary, video clip v_α can be represented by a "bag of FTVs" (BOF). Each FTV in the bag is a 1×21 vector, and there are m_α FTVs in the bag, where $m_\alpha = n_\alpha - T + 1$. Accordingly, the Feature Temporal Volumes (FTVs) are seen to be the elemental descriptors of a video clip. Furthermore, the *temporal context* surrounding each frame has taken into consideration to achieve a compact representation of the facial movements. The temporal size T of an FTV implies the number of frames considered as the duration of a so-called "compact" expression. Training was performed using 5-fold cross validation to determine the best thresholds, as well as the temporal size T. A Random Forest was trained as the decision classifier.

V. RESULTS AND DISCUSSION

Test results are shown in Fig. 8. The accuracy reflects the percentage of video clips that were correctly classified into deceptive and honest categories. Since the liars were considered as the positive samples, the true positive rate (TPR) reflects the precision of spotting them whereas the true negative rate (TNR) reflects the precision of spotting truth-tellers. Fig. 8 shows the variations in accuracy as well as the TPR and the TNR as functions of the FTV size T.

Several conclusions can be drawn from this figure: (1) The accuracy has a mean of 74.52% with a variance of 0.04%, implying that the choice of FTV size does not significantly influence the test accuracy. (2) The best accuracy, 76.92%, was

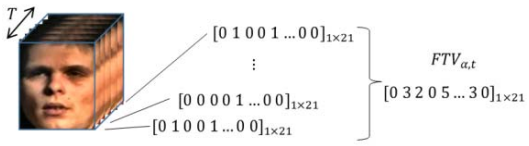

Fig. 7. Construction of a feature temporal volume $FTV_{\alpha,t}$

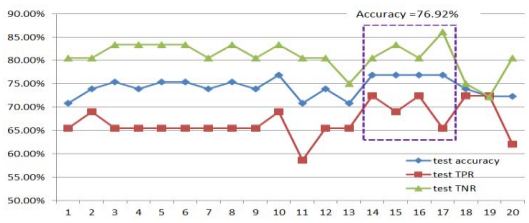

Fig. 8. The change in accuracy, TPR and TNR, as the FTV size varies. The results inside the dashed rectangle are the best obtained: accuracy=76.92%.

achieved when T=14~17. Note that the video frame rate was 30 frames per second. Therefore the ideal T is approximately half a second, which implies a reasonable duration for a compact facial expression. Thus, we can infer that most of the discriminating facial cues to deception are within half a second. (3) In general, the TNR is higher than the TPR, which means our system does a slightly better job of spotting truth-tellers than catching liars.

VI. CONCLUSION

As the first automated attempt in the literature at proving the viability of facial clues as deception detectors in unconstrained environments, the result is very promising. An accuracy of 76.92% for spotting liars and truth-tellers in high-stakes situations has been achieved. We also note that the best accuracy obtained by an experienced interogator, intensively trained in a lie-spotting workshop, is 80.9% [1].

Considering that the database we have used for our training and testing experiments was rather challenging, it is possible that better results could be obtained from data collected in an actual criminal interrogation scenario, where the environmental setting would be more controlled. In addition, this paper has shown that facial clues are potentially reliable indicators of deception in high-stakes situations.

REFERENCES

[1] M. O'Sullivan and P. Ekman, "The wizards of deception detection," in Deception Detection in Forensic Contexts, P. A. Granhag and L. A. Stromwall, Eds., ed Cambridge, England: Cambridge University Press, 2004, pp. 269-286.

[2] L. ten Brinke and S. Porter, "Discovering deceit: applying laboratory and field research in the search for truthful and deceptive behaviour," in Applied Issues in Investigative Interviewing, Eyewitness Memory, and Credibility Assessment, B. S. e. a. Cooper, Ed., ed New York: Springer Science and Business Media, 2013.

[3] A. Baker, L. ten Brinke, and S. Porter, "Will get fooled again: emotionally intelligent people are easily duped by high-stakes deceivers," Legal and Criminological Psychology, 2012.

[4] M. Hartwig, P. A. Granhag, L. A. Stromwall, A. G. Wolf, A. Vrij, and E. R. a. Hjelmsater, "Detecting deception in suspects: verbal cues as a function of interview strategy," Psychology, Crime and Law, vol. 17, pp. 643-656, 2011.

[5] T. O. Meservy, M. L. Jensen, J. Kruse, J. K. Burgoon, and J. F. Nunamaker, "Automatic extraction of deceptive bahavioral cues from video," Integrated Series in Information Systems, vol. 18, pp. 495-516, 2008.

[6] N. Micheal, M. Dilsizian, D. N. Metaxas, and J. K. Burgoon, "Motion profiles for deception detection using visual cues," in European Conference on Computer Vision, pp. 462-475, 2010.

[7] J. Rothwell, Z. Bandar, J. O'Shea, and D. McLean, "Silent talker: a new computer-based system for the analysis of facial cues to deception," Applied cognitive psychology, vol. 20, pp. 757-777, 2006.

[8] N. Bhaskaran, I. Nwogu, M. G. Frank, and V. Govindaraju, "Lie to me: deceit detection via online behavioral learning," presented at the International Conference on Automatic Face and Gesture Recognition and Workshops, 2011.

[9] L. ten Brinke and S. Porter, "Cry me a river: identifying the behavioural consequences of extremely high-stakes interpersonal deception," Law and Human Behavior, vol. 36, pp. 469-477, 2011.

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

[11] S. Mann, A. Vrij, E. Nasholm, L. Warmelink, S. Leal, and D. Forrester, "The direction of deception: neuro-linguistic programming as a lie detection tool," Journal of Police and Criminal Psychology, vol. 27, pp. 160-166, 2012.

[12] P. Ekman, Telling Lies: Clues to Deceit in the Marketplace, Politics, and Marriage. New York: W.W. Norton, 1992.

[13] S. Moore and R. Bowden, "Local binary patterns for multi-view facial expression recognition," Computer Vision and Image Understanding, vol. 115, pp. 541-558, 2011.

[14] M. Satiyan and R. Nagarajan, "Recognition of facial expression using Haar-like feature extraction method," International Conference on Intelligent and Advanced Systems (ICIAS), 2010.

[15] S.W. Chew, P.J. Lucey, S. Sridharan, and C.B. Fookes, "Exploring visual features through Gabor representations for facial expression detection," in Proceedings of International Conference on Auditory-Visual Speech Processing (AVSP), 2010.

[16] F. Tsalakanidou and S. Malassiotis, "Real-time 2D+3D facial action and expression recognition," Pattern Recognition, vol. 43, pp. 1763-1775, 2010.

[17] S.W. Chew, P. Lucey, S. Lucey, J.M. Saragih, J.F. Cohn, and S. Sridharan, "Person-independent facial expression detection using Constrained Local Models," IEEE International Conference Automatic Face and Gesture Recognition and Workshops, 2011.

[18] C. Darwin, The Expression of the Emotions in Man and Animals. London: J. Murray, 1872.

[19] L. ten Brinke, S. Porter, and A. Baker, "Darwin the detective: observable facial muscle contractions reveal emotional high-stakes lies," Evolution and Human Behavior, 2012.

[20] C. M. Hurley and M. G. Frank, "Executing facial control during deception situations," Journal of Noverbal Behavior, vol. 35, pp. 119-131, 2011.

[21] P. Ekman and W. Friesen, Facial Action Coding System: A Technique for the Measurement of Facial Movement. Palo Alto: Consulting Psychologists Press, 1978.

[22] S. Mann and R. Bull, "Suspects, lies, and videotape: an analysis of authentic high-stake liars," Law and Human Behavior, vol. 26, pp. 365-376, 2004.

[23] S. Leal and A. Vrij, "The occurence of eye blinks during a guilty knowledge test," Psychology, Crime and Law, vol. 16, pp. 349-357, 2010.

[24] "PittPatt SDK," 5.2.2 ed: Pittsburgh Pattern Recognition, 2011.

[25] A. S. M. Sohail and P. Bhattacharya, "Detection of facial feature points using anthropometric face model," in Signal Processing for Image Enhancement and Multimedia Processing, ed: Springer, 2008, pp. 189-200.

[26] M. J. Roshtkhari and M. D. Levine, "An on-line, real-time learning method for detecting anomalies in videos using spatio-temporal compositions," Computer Vision and Image Understanding, vol. 117, pp. 1436-1452, 2013.

[27] J. L. Fan and B. Lei, "A modified valley-emphasis method for automatic thresholding," Pattern Recognition Letters, vol. 33, pp. 703-708, 2012.

[28] N. Otsu, "A threshold selection method from gray-level histograms," Automatica, vol. 11, pp. 23-27, 1975.

[29] I. Fogel and D. Sagi, "Gabor filters as texture discriminator," Biological Cybernetics, vol. 61, pp. 103-113, 1989.

[30] A. Hulbert and T. Poggio, "Synthesizing a colour algorithm from examples," Sciences, vol. 239, pp. 482-485, 1998.