Is Interactional Dissynchrony a Clue to Deception? Insights From Automated Analysis of Nonverbal Visual Cues

Xiang Yu, Shaoting Zhang *Member, IEEE*, Zhennan Yan, Fei Yang, Junzhou Huang, Norah E. Dunbar, Matthew L. Jensen, Judee K. Burgoon, and Dimitris N. Metaxas, *Senior Member, IEEE*

***Abstract*—Detecting deception in interpersonal dialog is challenging since deceivers take advantage of the give-and-take of interaction to adapt to any sign of skepticism in an interlocutor's verbal and nonverbal feedback. Human detection accuracy is poor, often with no better than chance performance. In this investigation, we consider whether automated methods can produce better results and if emphasizing the possible disruption in interactional synchrony can signal whether an interactant is truthful or deceptive. We propose a data-driven and unobtrusive framework using visual cues that consists of face tracking, head movement detection, facial expression recognition, and interactional synchrony estimation. Analysis were conducted on 242 video samples from an experiment in which deceivers and truth-tellers interacted with professional interviewers either face-to-face or through computer mediation. Results revealed that the framework is able to automatically track head movements and expressions of both interlocutors to extract normalized meaningful synchrony features and to learn classification models for deception recognition. Further experiments show that these features reliably capture interactional synchrony and efficiently discriminate deception from truth.**

***Index Terms*—Deception detection, expression recognition, face tracking, gesture detection, interactional synchrony.**

Manuscript received March 6, 2013; revised November 24, 2013 and February 18, 2014; accepted June 3, 2014. Date of publication June 27, 2014; date of current version February 12, 2015. This work was supported in part by the National Science Foundation under Grant IIS-1064965, Grant IIS-1065013, Grant CNS-1059281, Grant CNS-1059218, Grant IIS-0964597, and Grant IIS-0964385, in part by the National Space Biomedical Research Institute through NASA under Grant NCC 9-58, in part by the Center for Identification Technology Research, a National Science Foundation (NSF) Industry/University Cooperative Research Center, and in part by the University of North Carolina at Charlotte. This paper was recommended by Associate Editor M. Pantic. *(Corresponding author: Shaoting Zhang).*

X. Yu, Z. Yan, F. Yang, and D. N. Metaxas are with the Department of Computer Science, Rutgers University, Piscataway, NJ 08854 USA (e-mail: xiangyu@cs.rutgers.edu; zhennany@cs.rutgers.edu; feiyang@cs.rutgers.edu; dnm@cs.rutgers.edu).

S. Zhang is with the Department of Computer Science, University of North Carolina at Charlotte, Charlotte, NC 28223 USA. (e-mail: shaoting@cs.rutgers.edu).

J. Huang is with the Department of Computer Science and Engineering, University of Texas at Arlington, Arlington, TX 76019 USA (e-mail: jzhuang@uta.edu).

N. E. Dunbar and M. L. Jensen are with the Center for Applied Social Research, University of Oklahoma, Norman, OK 73019 USA (e-mail: ndunbar@ou.edu; mjensen@ou.edu).

J. K. Burgoon is with the Center for Management of Information, University of Arizona, Tucson, AZ 85721 USA (e-mail: jburgoon@cmi.arizona.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCYB.2014.2329673

I. Introduction

IMPLICIT in all interpersonal interactions is the need to gauge whether an interlocutor is truthful and authentic in his or her presentation of self. The expectation of truthfulness, in fact, is one of the foundations of human discourse [36]. Yet, notwithstanding the importance of this largely outside-of-consciousness assessment process, voluminous research has shown that humans, unaided by technology, are very poor at detecting deception [1], [8], [78]. Average detection accuracy is estimated at 54%, or only slightly above chance, and detection of deception specifically, as opposed to detection of truthfulness, is approximately 47% [8]. Those accuracy estimates have included both lay and professional judges, although some recent evidence points to experts achieving higher accuracy rates under interviewing conditions more characteristic of their usual professional setting and task [11].

One reason cited for humans' poor detection in interpersonal dialog is that deceivers take advantage of the give-and-take of interaction to adapt to any signs of skepticism in the interviewer's verbal and nonverbal feedback. Deceivers adjust their messages to make their responses more plausible and their demeanor more credible [9], [82]. That same give-and-take, however, has the potential to offer subtle clues to deception through the disruption of interactional synchrony. Interactional synchrony refers to interaction that is nonrandom, patterned, and aligned in both timing and configuration of kinesic behavior (i.e., head, face, body, and limb movement) with the rhythms and forms of expression in the vocal-verbal stream. It is considered a key marker of interaction involvement, rapport, and mutuality. Synchrony may take the form of simultaneous synchrony, in which two or more people's behaviors mimic or match one another (e.g., similar postures and facial expressions) in the same time frame and behavioral changes occur at the same junctures. This is speaker-listener synchrony. Synchrony may also be concatenous, in which one person's behavior is followed by similar behavior from the next speaker (e.g., each using rapid nodding while speaking). This serial form of synchrony captures speaker-speaker and listener–listener coordination.

A. Hypothesis

The current investigation explores both simultaneous and concatenous synchrony. It is premised on the possibility that

engaging in deception disrupts interactional synchrony and may therefore be a clue to its presence. Practitioners have suggested using rapport-building techniques or interactional synchrony as an effective method for detecting deception: in FBI interviews with terrorists and in police investigations [43], [56], [71]. However, few systematic studies of rapport, coordination, synchrony or reciprocity have examined the effects of synchrony on deception or vice versa [12], [29]. The emphasis typically has been on interviewers using interactional synchrony to promote more verbal disclosures and confessions by interviewees.

Our approach is a novel perspective on the role of synchrony in revealing deception in that we are focusing on the interaction between the interviewee and interviewer rather than only the interviewee side of the equation. Deception has been shown to be a cognitively and emotionally taxing activity, especially when the stakes are high and the consequences of being discovered are serious [40], [77]. Interactional synchrony entails a very close linkage among behavioral, physiological, and emotional manifestations such that synchrony is positively correlated with rapport and empathy between interlocutors; conversely, incongruent feeling states and behavioral states can disrupt coordination, synchrony, and perceived rapport [48]. Mimicry, named the chameleon effect [3], [14], [27], is a nonconscious tendency to imitate others' verbal and nonverbal behaviors whereas we reserve the term "mirroring" for visually based static behaviors and not dynamic ones. In contrast, interactional synchrony is the smooth meshing in time of the rhythmic, patterned activity of two interlocutors and thus is a better fit for the behaviors we are interested in here because of the temporal component. If the behaviors involved are visual ones and are identical in form (e.g., one person's posture is just like the partner's), the pattern is mirroring. If he behaviors instead reflect some temporal, rhythmic, and smoothly meshed coordination between interactants, the pattern is called interactional synchrony.[1] Because deceivers may experience various negative emotional states (or at least emotional states that diverge from those of an interlocutor) and because deceivers may be too preoccupied with constructing plausible verbal responses to attend to or coordinate their nonverbal behaviors with another, we expect interactional synchrony to be attenuated and disrupted when interviewees are deceptive as compared to when they are truthful. Even skilled deceivers may be unable to counter this decrement in interactional enmeshment because conscious efforts to produce synchronous behavior patterns through mirroring another's posture or matching another's degree of animated gesturing and facial expressions may appear "inauthentic" and "off" [35]. Deception thus may be one cause of poor interactional synchrony and dissynchrony may be one sign that deception is taking place.

Our hypothesis tested this possibility. Specifically, we expected that interviews with deceivers are less synchronous than interviews with truth-tellers. Initial research using manual coding to evaluate synchrony suggested that deception alters the level of synchrony between deceivers and receivers. The modality used for the questioning and the type of questioning also affected the outcome [87]. Our goal is to determine whether computer vision techniques can expand on previous research by automatically detecting synchrony behavior, which can then be used to distinguish truthful from deceptive individuals. Testing this hypothesis required developing the computer vision methods to assess simultaneous and concatenous synchrony. Those methods are a central focus of the current work.

[1]This definition of synchrony is differentiated from mimicry, mirroring, and other forms of behavioral adaptation described in [10]

B. Moderators

Little is known about whether moderator variables alter the patterns of synchrony. Two possible factors were investigated here: the modality of interaction [face-to-face (FtF) or video-conferencing] and sanctioning of the deception. Video-conferencing is becoming a widespread medium for communication and sanctioning may alter how deception is behaviorally expressed [30]. Few experiments have examined video-conferencing and instead compare FtF interactions to those in text-only modalities [34].

In addition, few experiments directly compare the situation where the experimenter has sanctioned the deception to unsanctioned deception [32] and instead tend to focus on one or the other. In many experiments, participants are told by an experimenter to deceive their partner which may result in less nervousness, guilt, and dissynchrony. In other experiments, participants are allowed to choose whether or not to lie, which results in a lack of random assignment, such that only confident or skilled deceivers may choose to deceive.

Our experiment examines both of these moderators, modality, and sanctioning. We asked the following research questions: 1) is the synchrony between interviewer and interviewee affected by the modality they are using (FtF or video-conferencing) and 2) does the sanctioning of the deception (sanctioned or unsanctioned by experimenter) affect the synchrony between interviewer and interviewee?

C. Method

In overview, testing data were derived from a cheating experiment in which some subjects cheated during a trivia game with a confederate and some did not, but all were encouraged to appear as credible as possible when interviewed about the game by expert interviewers from the Department of Defense [30]. Thus, cheaters were expected to be deceptive and noncheaters, to be truthful. This kind of deception has been considered high stakes by other researchers [49] and the subjects in our paper were reminded that they were in violation of the university's honor code during their interviews. They knew they could face disciplinary action for this act and thus, were under pressure to evade detection. All interviews were videotaped. Separate cameras captured the interviewer and the interviewee and the time-aligned videos were rendered in split screen form. Modality and sanctioning were experimentally manipulated such that some participants were interviewed FtF and others were interviewed via Skype. Some were told that the experimenters were aware of their cheating

but that they were to deny it to the interviewer (sanctioned version) whereas others received no such explicit approval of their cheating (unsanctioned cheating).

The videotaped interviews were analyzed using computer vision methods for automated analysis. The autonomous tracking of interactional synchrony cues are proposed to deal with the cases where manual codings are not available. Also, as video is increasingly available (e.g., video data through the internet), manual coding would not be a suitable method for synchrony tracking because coding is time- and labor-intensive. Autonomous nonverbal cues extraction significantly improves such situations. Moreover, it is difficult for a person to track synchrony across multiple cues (e.g., shakes, nods, smiles, and gaze) in real time. Thus, the main considerations of proposing autonomous strategies to code the synchrony feature are to: 1) save the coding process time and labor, especially when the applications are large-scale, e.g., web-based video dataset; 2) improve the deception detection accuracy by investigating different nonverbal cues, its fusion and feature selection methods; and 3) investigate the factors which may influence the synchrony feature. For example, the modality difference, computer-mediated communication (CMC) versus FtF, deceiver sanctioning, the type of nonverbal cues, the confessor groups, etc. Considering the computer vision algorithms, constrained local models are used for detailed face tracking and head gestures [21], [65]. Our focus in this paper is on the automated tracking of head gestures and expressions of both the subject and the interviewer, extracting normalized meaningful synchrony features, and learning classification models for deception recognition.

Our major contributions are as follows.

1) Combining our developed face tracking, head gesture detection, and expression recognition modules, a fully automatic visual cue extraction system is introduced and synchrony information is modeled to detect deception.
2) Insights are put into investigating how synchrony influences the interaction of the two participants. Single-channel visual cue features are examined to show how they contribute to the classification of truth or deception. In-depth analysis are conducted for designing deception detection systems.
3) The deception detection problem is formulated as a classification problem. Feature selection toward synchrony features and two-layer classifier design significantly improve the detection accuracy from 54%, which is what a typical unaided individual can expect, to 74%.

The rest of the paper is organized as follows: Section II reviews relevant work on deception detection, face tracking, head gesture, and facial expression recognition. Section III presents the framework of our method. Section IV reports experimental results on two databases of different modalities. Section V concludes this paper.

II. Relevant Work

A. Deception Detection

Deception is defined as a message knowingly transmitted to create a false impression or knowledge on the part of the receiver [9], [76]. Early research focused on cardiorespiratory measures to detect evidence of lies in the form of the polygraph [47], [50]. In recent years, scholars have also added other physiological data such as neurological activity to identify liars [46].

The other main approach of detecting deception is to investigate nonverbal and verbal behavioral indicators of deception. For instance, there are many speech cues, such as pauses, voice pitch, interruptions, delay of response, and response length that are associated with deceit [75], [79], [80]. Verbal content indicators have included such features as the amount of detail, logical inconsistencies, spontaneous corrections, or other cues [26], [80]. Nonverbal behaviors that have been examined for indications of honesty or mendacity, have included eye contact, blinking, head movement, posture changes, gestures, and leg movements [5], [13], [38], [61], [64], [70], [73], [88]. Two ways in which these indicators reveal veracity is through signaling arousal and emotional states: "the emotional arousal hypothesis suggests that deception produces various emotional states which may influence nonverbal signals" [76]. Indicators may also reflect other processes such as cognitive load. Understudied is the extent to which observed behaviors reflect the social interaction between interlocutors rather than internal states of senders. This latter aspect motivated the current paper.

Measurement of nonverbal behavior can be manual, non-computational methods, or automatic computational methods [25], [52]. Early research relied on trained observers' rating, counting, or timing behaviors, which is the behavioral coding method [45]. This type of method is tedious and requires significant investment of labor and time. Moreover, there is no fixed rule for those methods to segment, annotate, and label observations, which may lead to confusion. Hence, researchers have turned to automatic techniques to measure individual behaviors and to assess the degree of similarity between the dynamic nonverbal behaviors of dyadic partners. These methods rely on motion-tracking devices, image processing, and video-based computer algorithms [13], [64], [73].

B. Face Tracking

Face tracking is a fundamental problem in computer vision. The face is a nonrigid and pose-variant object, which increases the difficulty in tracking. Moreover, illumination, facial expression, and occlusion are other factors that make the problem even harder.

There are two types of methods dealing with face-tracking problems. One is to extract local features and use standard trackers to trace the variation of the features. The other is to directly approximate the shift between the consecutive frames of face images where the local features are the same. In extracting features, the active shape model (ASM) [20] is one of the most successful methods so far. It sets up a series of landmarks which capture the face profile. Those landmarks extract gradient or pixel value and do a local search to find the proper locations along the face profile. A linear shape space is trained to constrain those landmarks in a face shape. The active appearance model [17] is another well-known

Fig. 1. Sample snapshots from tracked facial data showing a subject (left) and an interviewer (right). Red dots represent tracked facial landmarks (eyes, eyebrows, etc.), while ellipse in top left corner depicts the estimated 3-D head pose of the subject; top right corners show the detected expressions, and head gestures for subject and interviewer.

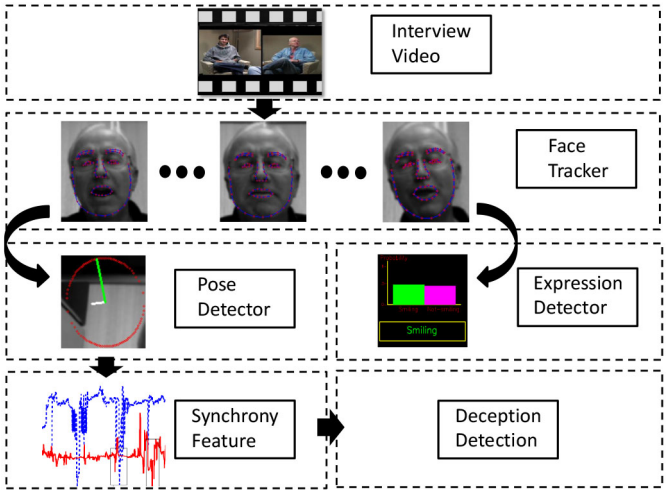


Fig. 2. Workflow of deception detection framework consisting face tracker, head pose detector, expression detector, synchrony feature extractor, and deception detection classifier.

algorithm to capture face profile. Besides the shape landmarks, it imports face appearance and attempts to minimize the error between the reference appearance and the searched appearance. Nevertheless, the global shape or appearance may encounter problems of local variations. The constrained local model developed by Cristinacce and Cootes [21] represents a face as a combination of shape and local feature templates. It is fitted by optimizing the shape parameters to match the image's local appearances to the templates. Improved methods are proposed based on those three above models [18], [19], [54], [85].

C. Head Movement Detection and Facial Expression Recognition

Detection of head gestures has a long tradition including work by Kapoor and Picard [42] who proposed a method to recognize head nods and head shakes based on two Hidden Markov Models trained and tested on a 2-D dataset from an eye gaze tracker. Kawato and Ohya [44] developed a method using "between eyes" templates. Recently, sequential analysis tools have become more and more important in gesture recognition. For instance, a conditional random field is imported in Quattoni *et al.*'s work [60]. They model the head motion as a temporal sequence and establish a graphical structure to analyze the behavior. Moreover, due to the complexity of real problems, Hidden-state CRFs, and Dynamic CRFs are gradually employed in analyzing time series data [55], [68].

Facial expression recognition is another important topic in the communities of computer vision. The previous work can be categorized into two classes, image-based methods and video-based methods [31], [59]. Image-based methods neglect the dynamic characteristics during an expression sequence. However, video-based methods deal with the dynamics to classify expression. Many experiments have demonstrated the importance of the facial dynamics [6], [15]. In video-based expression recognition, temporal segmentation of expression action events and the representation of the dynamics are two major problems. Torre *et al.* [69] used condensation to trace the local appearance dynamics and Cohn [16] applied key point tracker to represent the dynamics.

D. Interactional Synchrony

Synchrony is the dynamic and reciprocal adaptation of behaviors between interactive partners [25]. It is reflected by the relevant features of the interactive motion, i.e., head motion, facial expression, etc. Although using synchrony to evaluate the deceptiveness of statements is a rather new technique, an initial investigation using human coding has suggested that disruptions in synchrony can distinguish between deceivers and truth-tellers. It is perhaps due to their cognitive load or their violation of conversational norms [29].

To evaluate synchrony, correlation is one of the mainstream methods [2], [4], [7]. After extracting the motion or expression features, a time-lagged cross correlation is applied over the two sequential single-channel features with certain time slot window and thus the response indicates the degree of synchrony. Another strategy is to use recurrence analysis [62], [67]. For a given two time series, every vector with delay t is compared with every vector in the second time series. A recurrence matrix is created. By thresholding the distance between the two vectors, the degree of synchrony is provided. The third type of methods are spectral methods. Some methods focus on dealing with the two time series' relative phase [58], [63]. Others focus on measuring the overlap between the movement frequencies of the two interactants [24], [61].

III. Methodology

We have developed a framework that is capable of analyzing synchrony and detecting deception. The framework includes tracking facial movements module, facial expression recognition module, and head movement detection module. The sample interface is shown in Fig. 1. Based on the single-channel features extracted by those modules, i.e., head nodding, head shaking, smile, etc., we designed the temporal causality like strategy to generate synchrony features. Using the higher level features, a data-dependent learning-based classifier is designed to differentiate deceptive groups from truthful groups. The whole flowchart is shown in Fig. 2. Each of the modules is illustrated in detail in the following subsections.

A. Multipose Face Tracking

Face tracking is a challenging problem. The shapes of faces change dramatically with various identities, poses, and expressions. Furthermore, poor lighting conditions may cause a low contrast image or cast shadows on faces, which adversely affect the performance of the tracking system. We have developed a robust face tracker [41] based on ASMs [20] together with a nonlinear shape manifold.

ASMs are landmark-based models that attempt to learn a statistical distribution over variations in shapes for a given class of objects. The ASMs consist of a global shape model and a set of local landmark detectors

$$s \approx \bar{s} + \Psi \Delta\lambda. \tag{1}$$

The global shape model captures shape variations. As (1) shows, s stands for a predefined n facial landmark positions, $s = \{x_1, y_1, \ldots, x_n, y_n\}$. $\bar{s}$ is the mean shape from face training dataset. Ψ is the corresponding Eigen matrix depicting the face variation directions. Any shape s is approximated by the mean shape $\bar{s}$ and the variation from certain directions. $\Delta\lambda$ is the varying coefficients along the Eigen basis directions.

The local profile models capture the local appearances around each landmark point and are used for selecting the best candidate landmark positions. We adopt a logistic regression-based learning method [86] to obtain weights w, b for template detectors

$$p(v_i = 1|s_i, I) = \frac{1}{1 + exp(wf + b)}, f = \Phi(I, s_i). \tag{2}$$

In (2), we formulate the possibility of locating a candidate landmark position as $p(v_i = 1|s_i, I)$ knowing the facial image I, the i^{th} landmark candidate position s_i. Here, v_i is a random variable indicating whether the candidate position is in the positive location. f is feature vector extracted by Φ, which is the feature extraction strategy. In general, histogram of gradient (HoG) [22] and local binary pattern (LBP) [81] are widely used feature extractors in appearance detection. To locate the facial features in varying poses, we learn seven clusters of shape models, each covering a range of face poses. At each frame the system traverses the nonlinear facial shape manifold looking for the landmark configuration whose shape and texture at each landmark yield the minimum distance between what is observed in the image and the reconstructed shape. The learned model allows the complex, nonlinear shape manifold to be approximated in a piecewise linear subregions. Each subregion defines a hyper-ellipsoid on this manifold. Facial shapes of similar pose are constrained to lie in the same linear subspaces.

Fitting ASMs in every frame is computationally expensive and causes jittering. To solve this problem, we track the features using the KLT tracker [66] across consecutive frames. KLT tracker can register two local features and compute the displacement of the feature by minimizing the intensity matching cost. We track landmarks in successive frames using a sum of squared differences point tracker and running the relatively "expensive" step of face search only periodically to prevent any error accumulation. This scheme allows us to have a measure of tracking success (confidence) for each landmark, so we can have early detection and correction when drifting from the target.

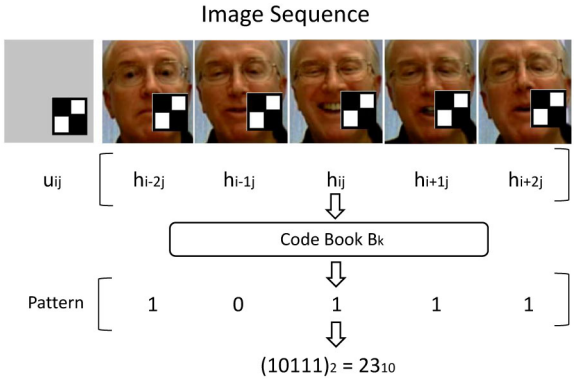


Fig. 3. Example of the coded feature on one haar-like dynamic feature unit u_{ij}.

B. Head Movement and Facial Expression

From the landmark positions in each frame, we are able to estimate the 3-D poses (pitch, yaw, and tilt) and detect the relevant head gestures (head shaking, nodding, head toward front, and head turning away). To estimate the face pose, a linear partial least squares (PLS) regression model [37] is built for all linear regions in the shape manifold. This regression model takes the x and y coordinates of the 79 landmarks as input, and predicts the pitch, yaw and tilt angles

$$A = XB + F^*. \tag{3}$$

Matrix A represents the head pose parameters pitch, yaw. and tilt. Matrix X is concatenated by all the landmarks' coordinates. B is the mapping matrix which we pursue and F^* is the residual matrix indicating the variation. The head nod is a gesture in which the head is tilted in alternating up and down, and head shake means that the head is turned left and right, repeatedly in quick succession. Therefore, by differentiating the pitch value and yaw angles in each frame, we are able to detect the head nod and shake, respectively.

A facial expression classifier is also built to detect facial expressions such as smiles [84]. We use a ranking-based framework of facial expression analysis, which can recognize expressions and estimate its intensity. Our method is based on an observation that the pair-wise ordinal relationship along the temporal domain is obvious, despite the lack of quantitative measurement of intensity. Therefore, it is relatively easy to model the intensity estimation as a ranking problem.

Our framework consists of three components as follows.

1) Facial appearance feature representation. We use the haar-like features to represent facial appearance due to its efficiency in facial appearance representation [74]. Then the dynamic haar-like feature is encoded into binary codes as illustrated in Fig. 3. The original haar-like dynamic feature u_{ij}, is extracted by canonical haar-like detector located at the position of (i, j) in the facial image. Such unit feature u_{ij} are then translated into binary code as Fig. 3 shows. The output binary code

Algorithm 1 Expression Feature Organization

1: **Input:** a subject with expression E_i and E_j.
2: **Output:** all pairwise instances $\{(x_k, x_{k+1})\}$
3: **repeat**
4: label intensity of E_i and E_j separately from apex to the start state.
5: concatenate two labeled sequence as: $S_{E_i} = R(I_{E_i,Apex}) \prec R(I_{E_i,start}) \prec R(I_{E_j,start}) \prec R(I_{E_j,Apex})$.
6: based on $\{S_{E_i}\}$, build pairwise instances $\{(x_k, x_{k+1})\}$ satisfying $R(x_{k+1}) \prec R(x_k)$.
7: **until** all pairwise instances x_k, x_{k+1} are visited.

is denoted as feature vector x, which is the input of the weak classifiers $h_t(x)$.

2) Ordinal pair-wise data organization. It describes how to organize the data properly for the ranking model. Detail procedure is demonstrated in Algorithm 1. $R(I)$ indicates the ranking score of instance I. Given a subject with expression E_i and E_j, which is represented by score in $(0, 1)$, we first label the expression in amplifying order. Then given two generated feature vectors x_k and x_{k+1}, after the classification score returning, we arrange the two feature vectors in the same order of E_i and E_j according to the classification scores, which are also ranged in $(0, 1)$.
3) Building ranking model, which is the core component. Due to a large number of haar-like features, we use the RankBoost [33] to select a subset of haar-like features to build a final strong ranker in

$$H(x) = \sum_{t}^{T} \alpha_t h_t(x) \tag{4}$$

where $h_t(x)$ is a weak ranker based on the following loss function:

$$L(H) = min \sum_{x_0,x_1} exp\left(\sum_{t}^{T} \alpha_t \left(h_t(x_0) - h_t(x_1)\right)\right) \tag{5}$$

t is the index and T is the total number of the weak classifiers. α_t is the weight for each weak classifier contributing to the strong classifier. We choose exponential function as our loss function. In (5), x_0 and x_1 are feature vectors extracted as binary coded features in Fig. 3, which belong to the same category, e.g., both of the vectors are indicating smiling. The objective function aims at minimizing the classification error of samples in the same group categorized into different groups.

In most learning models, the Vapnik–Chervonenkis (VC) dimension grows linearly in the number of parameters. If the training size is not large enough, the learning model may result in overfitting. Regularization is needed in this situation to prevent such a phenomenon [57]. $l1$ regularization has shown its advantages in vision and learning areas [39], [57], [84]. To further improve the performance, we introduce the $l1$ based regularization into the RankBoost. The final ranking score given by a ranking function can be used for expression intensity estimation and recognition. The loss function imports the constraints such that those positive dominating weak classifiers are sparse, as shown in (6)

$$L(H) = min \sum_{x_0,x_1} exp\left(\sum_{t}^{T} \alpha_t \left(h_t(x_0) - h_t(x_1)\right)\right)$$
$$s.t. \sum_{t}^{T} \alpha_t \le r, \alpha_t \ge 0. \tag{6}$$

To summarize, we first extract the haar-like feature and encode it into binary codes. Then by arranging the binary codes into ordinal pair-wise data, we use Rankboost to train a strong ranker to indicate the intensity of these emotions.

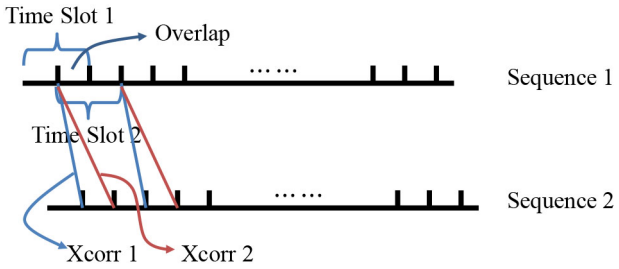


Fig. 4. Sequences cross correlation scheme.

C. Synchrony Features

The interaction of people in a dialog is directly indicated by head gestures and facial expression. However, the inner property of such interaction should be depicted by more profound features. The subtle and significant way people influence each other can be seen through their nonverbal synchrony. Synchrony refers to similarity in rhythmic qualities and enmeshing or coordination of the behavioral patterns of both parties in an interaction [12]. Such synchrony can either be simultaneous or concatenous. In [29], synchrony can be indicated by nodding or shaking, facial mirroring, etc. Providing pairs of interview videos, we can capture head nodding or shaking and facial expressions (especially smiling) in videos by our proposed facial tracking and facial expression detection methods. Based on such single-channel features, we intend to check the simultaneous response from both two people in an interview.

Individual feature vectors of two interview videos from one interviewer and one interviewee are viewed as two corresponding data sequences. We get large response while doing correlation over two sequences if the two sequences have similar magnitude at the same position, which may measure the simultaneous response. If two sequences have similar magnitude at different position, we take a time sliding window to compensate the time delay and calculate their correlation. Cross correlation is a standard method of estimating the degree to which two sequences are correlated. The cross correlation of two signals with a latency d is defined as

$$C(d) = \frac{\sum_i \left(z_i - \bar{z}\right)\left(y_{i-d} - \bar{y}\right)}{\sqrt{\sum_i \left(z_i - \bar{z}\right)^2} \sqrt{\sum_i \left(y_{i-d} - \bar{y}\right)^2}} \tag{7}$$

where z_i and y_i are the i th element of sequence z and sequence y, $\bar{z}$ and $\bar{y}$ are the mean value of sequence z and sequence y.

In order to accommodate concatenous synchrony, we divide two sequences into overlapped time slots, as shown in Fig. 4. The two sequences are required to have the same length. Then, we equally divide each sequence into m time slots. Starting from either of the sequences at current time slot, we take $[-t, t]$ time slots to calculate their correlation with the current time slot. The largest cross correlation response is chosen as the current time slot's feature value. We repeat such procedure for every time slot in a sequence. As a result we obtain a cross correlation based synchrony feature vector with length m.

D. Feature Selection and Deception Detection

Once we obtained effective single-channel feature, e.g., head-nodding, smiling, looking forward, asynchrony feature is formed by combining those single-channel features. The single-channel features are normalized, and then a weighted vector concatenation is applied to generalize a uniform feature vector. The weights for different single-channel features can either be tuned by k-fold cross validation or empirical setting. Such combination may not lead to optimal feature representation. Moreover, in the single-channel feature extraction step, noise may be introduced. Inside the combined feature, different feature elements may be correlated. Therefore, we chose the most effective feature elements out of the original synchrony feature vector to remove noise and redundancy.

This feature selection was achieved by genetic algorithm (GA). In this process, we randomly set the initial feature selectors consisting of 0/1 elements, in which 0 indicates not selecting and 1 means selecting. In each generation, such random feature selectors would crossover and mutate to generate new descendants. Then among the whole population, the algorithm will choose the top candidates which achieves the best performance in the classification task. Such iteration repeats until the accuracy cannot be increased or the maximum iteration limit is reached.

Obtaining the effective and precise selected synchrony features, we formulate the deception detection problem as a classification problem. We intend to differentiate the truthful group from the deceptive group, which is a two-class classification problem. Since the training volume is not large and the aim is to minimize the misclassification rate, we maximize the margin of those two groups and thus choose support vector machine (SVM) [72] as our first layer classifier. In addition, inside the deceptive group, it can be further classified as the sanctioned cheating group in which people are allowed to lie before they take the examination dialog, and the unsanctioned cheating group in which people are not provided with any information about deceiving. Similarly, we adopt SVM as our second layer classifier. In such way, we designed a multilayer SVM classifier for the deception detection task.

As a summary, in this section, we have introduced a face-tracking module to trace the head movement. Based on the head movement, we further set up the head pose detection module and facial expression detection module to obtain gesture and emotion features. Such single-channel features are not efficient enough for deception detection. We built a causal relationship-based synchrony feature as higher level feature. After feature selection, we designed a two-layer SVM classifier to achieve the deception detection task.

IV. Experiments

In this section, we first introduce the experimental protocols, and then show the tracking, gesture, and expression detection results as the input for the deception detection. Further, different single-channel features are examined and the feature selection algorithm is investigated to improve the recognition accuracy. Finally two-class and three-class classification output is illustrated, which reveals the advantages of our framework for analyzing synchrony and hence discriminating truth.

A. Experiment Settings

The analysis began with creating a database of 242 videotaped interviews of 121 interviewer–interviewee pairs. Interviewees were students who participated in a trivia game and in some cases were induced by a confederate to cheat. All participants were then interviewed by expert interviewers about the game interaction and whether they cheated during the game. Approximately half of the interviews were conducted over Skype and the other half were conducted FtF to produce two modality conditions, CMC and FtF. Since a few of the pairs are incompletely recorded, we selected 100 out of 121 pairs of videos as our training and testing data. These video pairs vary from 4500 frames to 15 000 frames. Although video pairs' lengths are different, we ensure inside each video pair, the interviewer sequence, and interviewee sequence keep the same length, which allows using a fixed number of time slots to analyze the video sequences.

To generate the synchrony feature, we set up certain width of time window. Each window has the same size, which are 180 frames comprised of 6 s prior to the current frame and 6 s following the current frame. Such window is tunable in practice. However, since too large window size may mediate the synchrony pattern and too small window size cannot capture significant pattern in synchrony, such window size is tuned according to the experimental performance. In our case, we modified the window size, generated the synchrony feature, and tested it on randomly chosen video segments. If the performance improved, we modified the the window size again. This process was repeated until the window size reached it optimum value. As our video pairs have length above 4500 frames, we set the window size to overlap a half in consecutive manner so that the procedure lasts for the entire sequence and generate a holistic synchrony feature.

B. Tracking, Gesture, and Expression Detection

In the synchrony detection step, we extracted head nodding, head shaking, smiling and head direction (looking forward or looking away). The visual result is shown in Fig. 6. More tracking results are shown in Fig. 10. Head motion can be detected by analyzing pitch, yaw, and tilt as demonstrated in Section II-B. Pitch depicts nodding action and yaw

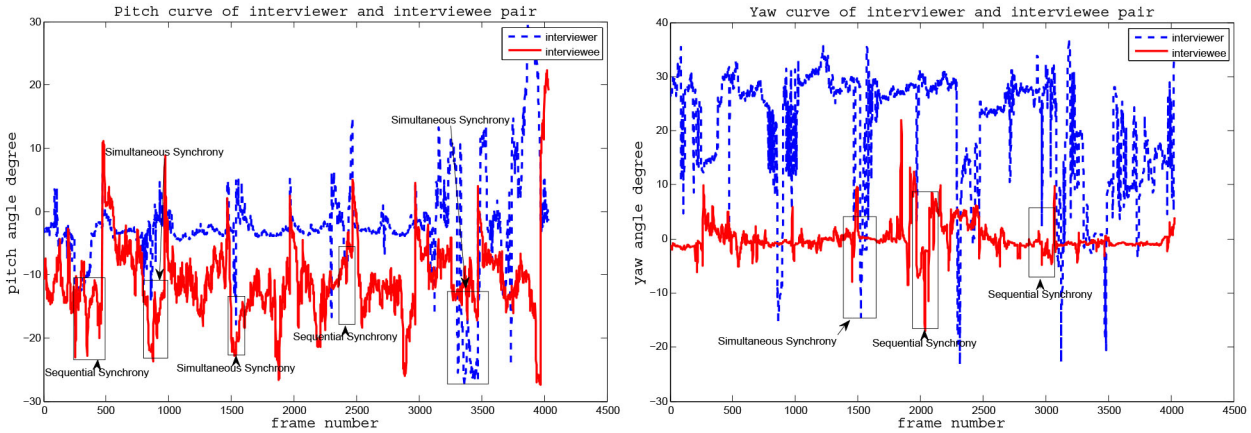


Fig. 5. Synchrony pattern illustration in pitch and yaw angle curves. Left: Pitch curve of a pair of interviewer and interviewee. Right: Yaw curve of a pair of interviewer and interviewee. *x*-axis stands for frame number and *y*-axis represents angle degree (pitch or yaw).

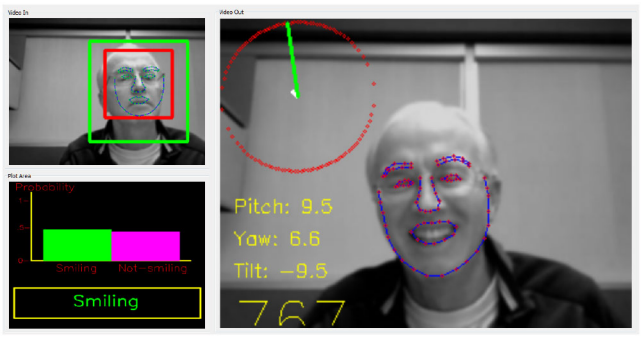


Fig. 6. Face tracking and expression, head pose estimation visual result. Left Top: Initial face detection and landmark initialization. Left Bottom: Score plot of expression (smile or not smile) recognition. Right: Facial landmark tracking result and head pose estimation(depicted by pitch, yaw, and tilt).

reveals shaking action. The nodding and shaking synchrony patterns are shown in Fig. 5. Based on such single-channel features, we further combine the interviewer's feature vector with interviewee's feature vector to form higher level features. A correlation-based method is adopted to identify synchrony. Then a two-layer classification scheme is designed to separate three classes (truthful group, sanctioned cheating group, and unsanctioned cheating group). We first classify the truthful group from the cheating group using nonlinear SVM classifier, which is a two class classification task. Then based on the result of the first step, we continue to classify the cheating group into sanctioned cheating group and unsanctioned cheating group by another nonlinear SVM classifier. During the feature selection part, at each step we separately train a feature selector using GAs. The feature selector is an efficient way to promote the performance in recognition task because the raw features may have noise or redundancy.

C. Evaluation of Synchrony Features

Before using all features, different types of features should be investigated to find the effective ones for classification. Our strategy is to leave each single feature out of the whole feature vector and then test the recognition accuracy. We also identify the single feature's recognition accuracy and visualize the feature vector in plots to see the separability of each of the four types of features (head nodding, head shaking, smiling or not smiling, and look forward or look away). During this step, we examine the average precision of classifying three classes, i.e., truthful, sanctioned cheating, and unsanctioned cheating classes, to evaluate each feature. Table I shows the average precision of different feature combinations over the three-class classification.

TABLE I
DETECTION ACCURACY EVALUATION OF FOUR FEATURES, "NOD," "SHAKE," "SMILE," AND "LOOK FORWARD." "ALL BUT ONE" MEANS THAT ALL FEATURES ARE USED EXCEPT THE ONE OF THAT COLUMN. "SINGLE" MEANS USING ONLY THE FEATURE OF THAT COLUMN

		Nod	Shake	Smile	Look forward
CMC	All but one	**0.422**	**0.437**	0.356	0.311
	Single	0.35	0.33	**0.38**	**0.43**
FtF	All but one	0.442	0.473	**0.554**	-
	Single	**0.513**	**0.464**	0.435	-

Table I shows that in CMC, when the feature "Nod" or "Shake" is excluded from the whole feature vector, the performance is higher than the rest. When the feature "Smile" or "Look forward" is excluded, the performance drops. For FtF, the trend is opposite: "Nod" and "Shake" are more significant in classification. When testing each single feature's accuracy, it appears that "Look forward" and "Smile" are more accurate than "Nod" and "Shake" for CMC. And again, for FtF modality, "Nod" and "Shake" achieves higher accuracy than "Smile," where "Look forward" is not applicable in the FtF dataset. The reason is for FtF data, interviewer and interviewee sit FtF. The look-forward feature should be defined by their local head coordinates. But only one camera was capturing the scene, only allowing the global camera coordinate. Thus, the frontal face cannot be obtained by the camera coordinate.

In Fig. 7, the vertical dotted lines separate the plot into four regions representing the four separate features. The first column indicates the feature "Nod," the second one is the feature "Shake," the third is the feature "Smile," and the last one is the feature "Look forward." We plot the average feature vector of each group in the subplots. The feature vector is 800 numbers long, of which each region is with length 200 numbers. With

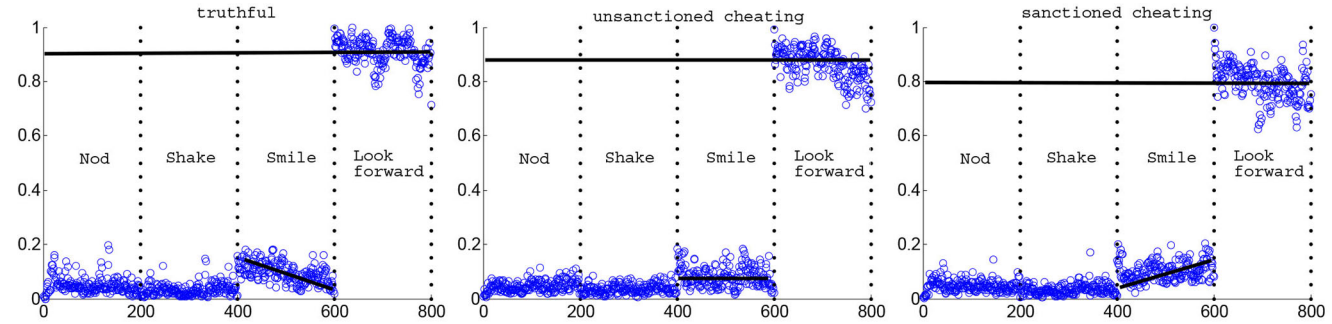


Fig. 7. Mean feature vector patterns of three groups. Left: Truthful group's mean feature vector pattern. Middle: Unsanctioned cheating group's mean feature vector. Right: Sanctioned cheating group's mean feature vector.

TABLE II
ACCURACY OF PROPOSED SYNCHRONY FEATURE, FEATURE IN [53] AND EACH SINGLE CHANNEL FEATURE. "TP" AND "FP" STAND FOR TRUE AND FALSE POSITIVE

		TP	FP	Precision	Recall
CMC	Proposed	**0.667**	**0.167**	**0.668**	**0.667**
	[53]	0.420	0.289	0.429	0.420
	Nod	0.370	0.314	0.374	0.370
	Shake	0.370	0.316	0.360	0.370
	Smile	0.437	0.282	0.432	0.437
	Look forward	0.462	0.261	0.478	0.462
FtF	Proposed	**0.651**	**0.187**	**0.668**	**0.651**
	[53]	0.442	0.280	0.442	0.442
	Nod	0.372	0.338	0.394	0.372
	Shake	0.395	0.319	0.395	0.395
	Smile	0.349	0.349	0.339	0.349

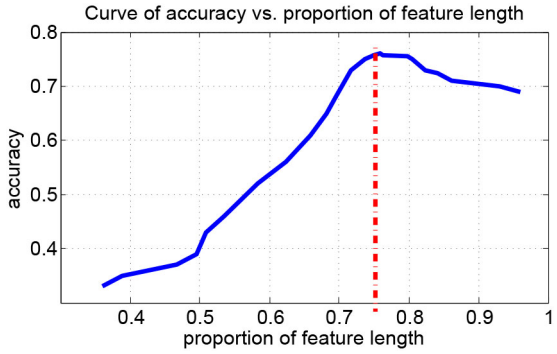


Fig. 8. Relationship between proportion of feature length and classification accuracy.

the black line showing the trends in the figure, we see that in region three, the pattern of the feature vector is obviously different. In the subplot for the truthful condition, it is going down; in the subplot for unsanctioned cheating, it is flat; in subplot of sanctioned cheating, it is going up. In region four, the average value of those numbers is going down from above 0.9 to less than 0.9 until around 0.8.

In practice, we have applied nod, shake, smile, looking forward, and hit-miss rate proposed in [53] for the single-channel features. Besides the "all-but-one" evaluation, we would like to investigate each single-channel feature comparing to features proposed in [53] and our fused synchrony feature. The comparison may reveal whether the proposed synchrony feature is a valid one for deception detection and whether it is more advantageous other than features proposed in [53] or single-channel features.

The feature proposed in [53] is mainly the hit-miss rate. The problem in that paper is to detect deception from truth, which is a two-class classification problem. The proposed method deals with three-class categorization problem, which is to classify truthful, sanctioned cheating, and unsanctioned cheating groups.

In Table II, we listed the comparison of true positive rate, false positive rate, and precision. For the CMC database, the fusion feature achieves largest true positive and precision rate, with smallest false positive rate. In contrast, the feature proposed in [53] shows limited advantages over other single-channel features. The reason may be that our proposed fusion feature includes the feature proposed in [53] and we applied efficient feature selection method to improve the accuracy. The same observation appears to the FtF database, which consistently reveals that the synchrony feature is a more efficient higher level feature in detecting deception.

D. Evaluation of Feature Selection

We applied GA for feature selection in our framework. In GAs, there are several parameters to influence the recognition rate. Basically, they are the length of selected elements, crossover segment number, mutation ratio, the amount of population for each generation, etc. The length of selected feature elements are decided by other factors such as crossover segment number, mutation ratio, the amount of population, etc. Since all other factors are finally reflected to the length of elements, for simplicity, we investigate how the recognition rate varies with the selected feature length.

To experiment with the parameters in the GA, we set the crossover segment number to vary from 2 to 100, the crossover time varied from 1 to 3, and the mutation ratio varied from 0.005 to 0.05. For each set of parameters, we independently ran the GA five times. Each time, we got the recognition rate together with the selected feature length. When all the sets of parameters were listed, we analyzed the relationship between recognition rate and the feature length. Then the curve is plotted in Fig. 8.

In the plot, the horizontal axis means the proportion of the original feature length and the vertical axis stands for the recognition rate. From Fig. 8, we observe that as the proportion decreases from 1 to 0.8, the recognition rate is increasing until it reaches 0.75, when the accuracy reaches

TABLE III
CONFUSION MATRIX OF CLASSIFYING TRUTHFUL AND DECEPTIVE CASES OF CMC AND FTF MODALITIES

		Truthful	Deceptive
CMC	Truthful	**10**	5
	Deceptive	6	**24**
FtF	Truthful	**12**	4
	Deceptive	7	**20**

maximum. And as the proportion goes down from 0.7 to 0.4, the accuracy decreases almost monotonically. We think that the feature vector with full length may contain redundancy and noise. After feature selection, the useful feature elements are selected, by which the redundancy and noise are removed. As a result, the accuracy is expected to increase. When the feature length is continuously shortened, some of the useful information in the feature vector may be eliminated. In this case, the accuracy may decrease as the figure shows. As a consequence, in this experiment, 0.75 is considered optimal proportion of feature length to achieve the best accuracy.

With optimal chosen parameters, comparing the accuracy of features with and without feature selection, for two-class classification, the original accuracy is 54.2% and the accuracy with feature selection is 74.2%; for three-class classification, the original accuracy is 46.9% while the accuracy after feature selection is 66.8%. The accuracy with feature selection is more than 10% higher than that without feature selection, which indicates that the feature selection is a key step of accuracy improvement.

E. Evaluation of the Classification Accuracy

1) Two-Class Classification: Even with feature selection's promotion, it is still possible to improve the accuracy since proper classifier design could enhance performance. The initial three-class classification using nonlinear SVM scheme may not be perfect because it contains at least three intersections of misclassification, which are intersections of truthful and unsanctioned cheating groups, truthful and sanctioned cheating groups, unsanctioned cheating and sanctioned cheating groups. Although the problem is to divide the data into truthful, unsanctioned cheating and sanctioned cheating groups, it is at least a two-class' classification problem, which is truthful and cheating groups' classification. We could continue to solve a two-class' classification problem on the unsanctioned cheating and sanctioned cheating groups in the same way. Hence, we get only two intersections of misclassification, misclassification of truthful and cheating groups and misclassification of unsanctioned cheating and sanctioned cheating groups, which is expected to decrease the error recognition rate. We set both 15 test samples for truthful group and cheating group. Thus, 70 samples are the training samples, 16 in the truthful group and 54 in the cheating group. The performance is shown in Table III.

The confusion matrix in Table III shows that for the CMC dataset in the truthful group, ten samples are correctly classified while five are not; in the cheating group, which is the combination of unsanctioned cheating and sanctioned cheating groups, 24 samples are correctly classified and only six are not. Table IV shows the classification accuracy details. In CMC, truth precision is 62.5% and deception precision is 82.8%, for an overall average of 72.7%. In FtF, the precision values are 63.2% for truth and 83.3% for deception for an overall precision of 75.8%, which is roughly at the same level as CMC.

TABLE IV
ACCURACY OF CLASSIFYING THE TRUTHFUL AND DECEPTIVE CASES. "TP" AND "FP" STAND FOR TRUE POSITIVE AND FALSE POSITIVE

		TP	FP	Precision	Recall
CMC	Truthful	0.667	0.200	**0.625**	0.667
	Deceptive	0.800	0.333	**0.828**	0.800
	Average	0.734	0.267	**0.727**	0.734
FtF	Truthful	0.750	0.259	**0.632**	0.750
	Deceptive	0.741	0.250	**0.833**	0.741
	Average	0.744	0.253	**0.758**	0.744

TABLE V
CONFUSION MATRIX OF CLASSIFYING TRUTHFUL, UNSANCTIONED, AND SANCTIONED CHEATING CASES OF CMC AND FTF

		Truthful	Unsanctioned	Sanctioned
CMC	Truthful	**10**	5	3
	Unsanctioned	4	**9**	2
	Sanctioned	2	2	**11**
FtF	Truthful	**11**	1	4
	Unsanctioned	5	**7**	1
	Sanctioned	2	2	**10**

TABLE VI
ACCURACY OF CLASSIFYING THE TRUTHFUL, UNSANCTIONED, AND SANCTIONED CHEATING CASES. "TP" AND "FP" STAND FOR TRUE POSITIVE AND FALSE POSITIVE

		TP	FP	Precision	Recall
CMC	Truthful	0.667	0.200	**0.625**	0.667
	Unsanctioned	0.600	0.133	**0.692**	0.600
	Sanctioned	0.733	0.167	**0.688**	0.733
	Average	0.667	0.167	**0.668**	0.667
FtF	Truthful	0.750	0.259	**0.632**	0.750
	Unsanctioned	0.538	0.067	**0.778**	0.538
	Sanctioned	0.714	0.172	**0.667**	0.714
	Average	0.651	0.187	**0.668**	0.651

2) Three-Class Classification: After the classification of truthful and cheating groups, based on the cheating categorization result, we continue to classify the cheating group into unsanctioned cheating and sanctioned cheating groups. The classification scheme is the same as first step. However, the training and classification is data-dependent, especially in feature selection and nonlinear SVM classifier training.

Table V shows our final confusion matrices over all the three categories. In each category, the number of correctly recognized samples dominates misclassified numbers. Further Table VI illustrates that the precisions of all classes are above 60%, two of which are approaching 70%. The average accuracy is 66.8%, which is clearly a significant improvement over 47% [8].

F. Evaluation of Confessors in Deception Detection

In the experiment, some of the interviewees confessed to deception during the interview. Before they confess to the

TABLE VII
CONFUSION MATRIX OF CLASSIFYING TRUTHFUL, UNSANCTIONED, AND SANCTIONED CHEATING CASES OF CMC AND FTF IN CONFESSION GROUP EXCLUDED CONDITION

		Truthful	Unsanctioned	Sanctioned
CMC	Truthful	**27**	3	1
	Unsanctioned	10	**18**	2
	Sanctioned	4	6	**7**
FtF	Truthful	**14**	2	0
	Unsanctioned	2	2	1
	Sanctioned	2	1	**3**

TABLE VIII
ACCURACY OF CLASSIFYING THE TRUTHFUL, UNSANCTIONED, AND SANCTIONED CHEATING CASES IN CONFESSION GROUP EXCLUDED CONDITION. "TP" AND "FP" STAND FOR TRUE AND FALSE POSITIVE

		TP	FP	Precision	Recall
CMC	Truthful	0.871	0.298	**0.659**	0.871
	Unsanctioned	0.600	0.188	**0.667**	0.600
	Sanctioned	0.412	0.049	**0.700**	0.412
	Average	0.667	0.201	**0.671**	0.667
FtF	Truthful	0.875	0.364	**0.778**	0.875
	Unsanctioned	0.400	0.136	**0.400**	0.400
	Sanctioned	0.500	0.048	**0.750**	0.500
	Average	0.704	0.251	**0.702**	0.704

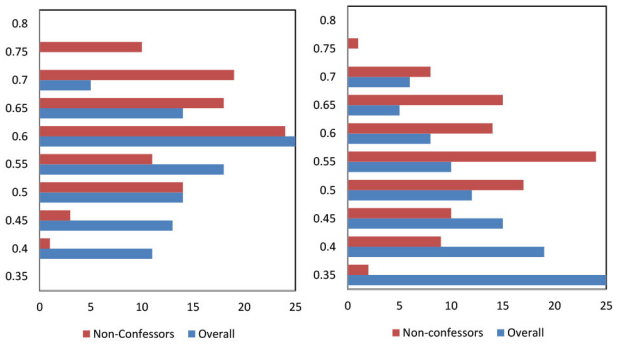


Fig. 9. Accuracy histogram of overall and nonconfessor groups. Left: Accuracy histogram of CMC database. Right: Accuracy histogram of FtF database.

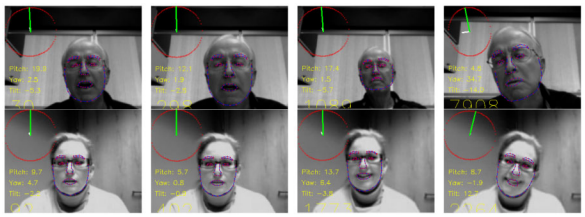


Fig. 10. More visual results of the multipose tracking system. The first row are results from one interviewer. The second row are results from the corresponding interviewee.

interviewer, the pattern may appear the same as cheating mode. After the confession, they may have felt relieved and then performed in a similar fashion to truth-tellers. The confessor group inside the cheating group may have influenced detection. We aim to determine if the confessors were more synchronous than the nonconfessors by evaluating the dataset excluding the confessor group. Comparing to the results in Section III-C, we expect to find the degree of synchrony by including and excluding such confession group.

Table VII reports the confusion matrices of the three-class classification result on both CMC and FtF databases. The diagonal elements of the two matrices dominate all the other elements which reflect that our classification scheme groups most of the samples correctly. Further, comparing Table VIII with Table VI, the excluding confessor classification achieves at least as good as the including confessor scheme. Moreover, it shows that in "Truthful" group of both CMC and FtF, the excluding scheme achieves 80.6% accuracy for CMC and 87.5% for FtF, while the including scheme achieves 66.7% for CMC and 68.8% for FtF. Nevertheless, the average precision of excluding confessor cases is slightly higher than including cases in both CMC and FtF datasets.

To test the differences between the confessor group and nonconfessor group, we set up hypothesis-test experiments both for CMC and FtF databases. For the CMC database, we have 109 valid video pairs and consequently 109 valid synchrony feature vectors. Among the whole dataset, there are 78 nonconfessors and 31 confessors. We propose the following hypotheses.

1) *H0:* The confessor group has no difference with the nonconfessor group.
2) *H1:* The confessor group has difference with certain significance level, which is to reject H0.

Our experiment is set up under the Monte Carlo framework. The first experiment is to randomly divide the 109 feature vectors into two sets, one for training and the other for testing. Generally, the training and testing volume is equal. Then we record the test accuracy of a one-time experiment and repeat this experiment 100 times. Each time the training and testing datasets are partitioned again. Thus, we independently conduct the experiment 100 times and obtain 100 classification accuracy records for the 109 feature vectors. The second experiment was to remove the 31 confessors from the whole 109 feature vector group. With only 78 nonconfessors, we conduct the same experiment as before. We independently repeat the training and testing process 100 times. Each time the training and testing datasets are repartitioned. The final classification accuracy is recorded for comparison.

After the two separate experiments, we obtain two accuracy vectors, each of which is with 100 elements. Then, we apply t-test over the two vectors under the null hypothesis with significance level 0.05, which is acceptable for most hypothesis-test experiments. For CMC experiment, the t-value is 5.87 and the threshold from t-table is 1.66. Since 5.87 is larger than 1.66, we reject the null hypothesis with probability 0.95. For FtF experiment, the t-value is 8.55 and the threshold from t-table is 1.66. Since 8.55 is larger than 1.66, we again reject the null hypothesis with probability 0.95.

The accuracy histogram is visualized for comparison in Fig. 9. The blue columns represent the overall group accuracy and the red columns stand for nonconfessor group accuracy. The vertical axis is the accuracy interval. Each interval calculates the accuracy below that threshold. The horizontal axis means how many times the accuracy appears inside the interval of vertical axis. Clearly, the distribution for the two groups

is different. This supports the idea that the confessor groups undermine classification accuracy.

V. Conclusion

In this investigation, we hypothesized that the introduction of deception into an interview would disrupt the synchrony of a dyad. We also examined whether the modality of the interview (CMC versus FtF) and the sanctioning of deception by the interviewee would affect the diagnosticity of synchrony. The analysis of the CMC and FtF conditions were proposed in parallel fashion, but the four features (head nod, head shake, smile, and look forward) had different significance in the two modality conditions, possibly due to the physical location of the camera or individuals in each. Both in two-class and three-class classification, the performance of CMC and FtF datasets achieved the same degree of accuracy, which suggests that the degree of synchrony was not influenced much by the modality of communication. This finding is consistent with other synchrony research based on manual coding [28]. Nevertheless, from the three-class classification result, the sanctioned cheating group is well separated from the unsanctioned cheating group, which indicates sanctioning is a key factor to influence synchrony and as a result discriminates unsanctioned cheating from sanctioned cheating. This finding does not contradict findings from synchrony research using manual coding which found the unsanctioned deceivers most distinguishable. We state that sanctioning is a key factor for deception detection but no judgment is made whether sanctioning group or unsanctioning group is easier to detect. Moreover, the manual assessment of synchrony was at a gestalt level, not at the level of detail presented here. Those deceivers who confessed during the interview also influenced the classification process. Once the confessor group is removed, the truth-tellers are much better separated from the deception groups than before.

Automatic methods can often detect events of synchrony which are missed by the human coders for whatever reason. In particular, we found that while the human coders in the Dunbar *et al*. study [29] would label a given video as having no synchrony in it, our software did detect a number of synchrony events, producing disagreement between the results of the manual analysis and the results of the automatic analysis. Despite a small percentage of false negatives in detecting the events of interest (i.e., nodding, shaking, smiling), the results of the automatic analysis are supportive of the initial hypothesis of synchrony being detectable and discriminating among conditions. This means that monitoring synchrony events, while establishing implicit models of deception, may be useful for automatic deception detection.

False negatives (for shaking and nodding) are attributed to the poor resolution of the input video and to the fact that the camera was not frontal to the faces. In particular, the face was quite small, and although it was correctly tracked, the displacement of the facial landmarks was sometimes not large enough to register as a nodding or shaking event. We believe that using videos of better quality, with facial close-ups, will improve our results and confirm our findings.

In this paper, we investigated how the degree of interactional synchrony can signal whether deceit is present or absent. An automated framework has been introduced to analyze videos effectively, and a new group of features has been proposed that not only register synchrony but also can detect deception at a reasonable accuracy. Future analysis will consider if the trend discovered thus far by our computerized methods generalizes to the greater sample population and also to other scenarios in which deception may be present. Furthermore, we will improve our face tracking system by incorporating 3-D deformable models [23], [51], [83].

References

[1] M. G. Aamodt and H. Custer, "Who can best catch a liar?: A meta-analysis of individual differences in detecting deception," *Forensic Examiner*, vol. 15, no. 1, pp. 6–11, 2006.

[2] K. Ashenfelter, S. Boker, J. Waddell, and N. Vitanov, "Spatiotemporal symmetry and multifractal structure of head movements during dyadic conversation," *J. Exp. Psychol. Human Percept. Perform.*, vol. 35, no. 4, pp. 1072–1091, 2009.

[3] J. Bailenson and N. Yee, "Digital chameleons automatic assimilation of nonverbal gestures in immersive virtual environments," *Psychol. Sci.*, vol. 16, no. 10, pp. 814–819, 2005.

[4] A. Barbosa, E. Bateson, M. Oberg, and R. Dechaine, "An instantaneous correlation algorithm for assessing intra and inter subject coordination during communicative behavior," in *Proc. Workshop Modeling Human Commun. Dyn. Neural Inf. Process. Syst. (NIPS)*, 2010.

[5] N. Bhaskaran, I. Nwogu, M. Frank, and V. Govindaraju, "Lie to me: Deceit detection via online behavioral learning," in *Proc. Automat. Face Gesture Recognit.*, Santa Barbara, CA, USA, 2011, pp. 24–29.

[6] M. J. Black and Y. Yacoob, "Recognizing facial expressions in image sequences using local parameterized models of image motion," *Int. J. Comput. Vis.*, vol. 25, no. 1, pp. 23–48, 1997.

[7] S. Boker, M. Xu, J. Rotondo, and K. King, "Windowed cross-correlation and peak picking for the analysis of variability in the association between behavioral time series," *Psychol. Methods*, vol. 7, no. 3, pp. 338–355, 2002.

[8] C. F. Bond, Jr., and B. M. DePaulo, "Accuracy of deception judgements," *Pers. Soc. Psychol. Rev.*, vol. 10, no. 3, pp. 214–234, 2006.

[9] D. B. Buller and J. K. Burgoon, "Interpersonal deception theory," *Commun. Theory*, vol. 6, no. 3, pp. 203–242, 1996.

[10] J. K. Burgoon, N. E. Dunbar, and C. H. White, "Interpersonal adaptation," in *Handbooks of Communication Science Vol. 6. Interpersonal Communication*, P. J. Schultz and P. Cobley (Ser. Eds.) and C. R. Berger (Vol. Ed.). Berlin, Germany: De Gruyter Mouton, 2014, pp. 225–248.

[11] J. K. Burgoon, J. F. Nunamaker, Jr., and D. N. Metaxas, "Noninvasive measurement of multimodal indicators of deception and credibility," Final Report to the Defense Academy for Credibility Assessment (Grant IIP-0701519), Jul. 10, 2010.

[12] J. K. Burgoon, L. A. Stern, and L. Dillman, *Interpersonal Adaptation: Dyadic Interaction Patterns*. Cambridge, U.K.: Cambridge Univ. Press, 2007.

[13] N. Campbell, "Multimodal processing of discourse information: The effect of synchrony," in *Proc. 2nd Int. Symp. Universal Commun. (ISUC)*, Osaka, Japan, 2008, pp. 12–15.

[14] T. Chartrand and J. Bargh, "The chameleon effect: The perception behavior link and social interaction," *J. Pers. Soc. Psychol.*, vol. 76, no. 6, pp. 893–910, 1999.

[15] I. Cohen, N. Sebe, L. Chen, A. Garg, and T. Huang, "Facial expression recognition from video sequences: Temporal and static modeling," *Comput. Vis. Image Underst.*, vol. 91, nos. 1–2, pp. 160–187, 2003.

[16] J. Cohn, "Automated analysis of the configuration and timing of facial expression," in *What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System*, 2nd ed. Oxford, U.K.: Oxford Univ. Press, 2005, pp. 388–392.

[17] T. Cootes, G. Edwards, and C. Taylor, "Active appearance models," in *Proc. 5th Eur. Conf. Comput. Vis. (ECCV)*, Freiburg, Germany, 1998, pp. 484–498.

[18] T. Cootes and C. Taylor, "Constrained active appearance models," in *Proc. Int. Conf. Comput. Vis. (ICCV)*, Vancouver, BC, Canada, 2001, pp. 748–754.

[19] T. Cootes, K. Walker, and C. Taylor, "View-based active appearance models," in *Proc. Autom. Face Gesture Recognit.*, Grenoble, France, 2000, pp. 227–232.

[20] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active shape models—Their training and application," *Comput. Vis. Image Underst.*, vol. 61, no. 1, pp. 38–59, 1995.

[21] D. Cristinacce and T. Cootes, "Feature detection and tracking with constrained local models," in *Proc. Brit. Mach. Vis. Conf. (BMVC)*, 2006, pp. 929–938.

[22] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Comput. Vis. Pattern Recognit. (CVPR)*, 2005.

[23] D. Decarlo and D. Metaxas, "Optical flow constraints on deformable models with applications to face tracking," *Int. J. Comput. Vis.*, vol. 38, no. 2, pp. 99–127, 2000.

[24] E. Delaherche and M. Chetouani, "Multimodal coordination: Exploring relevant features and measures," in *Proc. 2nd Int. Workshop Social Signal Process.*, New York, NY, USA, 2010, pp. 47–52.

[25] E. Delaherche *et al.*, "Interpersonal synchrony: A survey of evaluation methods across disciplines," *IEEE Trans. Affect. Comput.*, vol. 3, no. 3, pp. 349–365, Jul./Sep. 2012.

[26] B. DePaulo *et al.*, "Cues to deception," *Psychol. Bull.*, vol. 129, no. 1, pp. 74–118, 2003.

[27] A. Dijksterhuis, "Automatic social influence: The perception-behavior link as an explanatory mechanism for behavior matching," *Social Influence: Direct and Indirect Processes*. Philadelphia, PA, USA: Psychology Press, 2001.

[28] N. E. Dunbar, M. L. Jensen, D. C. Tower, and J. K. Burgoon, "Synchronization of nonverbal behaviors in detecting mediated and non-mediated deception," *J. Nonverb. Behav.*, 2014.

[29] N. E. Dunbar *et al*. (2011, Jan.). A dyadic approach to the detection of deception. in M. Jensen, T. Meservy, J. Burgoon, and J. Nunamaker. Report of the HICSS-44 Credibility Assessment and Information Quality in Government and Business [Online]. Available: http://www.hicss.hawaii.edu/Reports.htm

[30] N. E. Dunbar *et al.*, "Effects of veracity, modality and sanctioning on credibility assessment during mediated and unmediated interviews," *Commun. Res.*, 2013.

[31] B. Fasel and J. Luettin, "Automatic facial expression analysis: A survey," *Pattern Recognit.*, vol. 36, no. 1, pp. 259–275, 2003.

[32] T. H. Feeley and M. A. de Turck, "The behavioral correlates of sanctioned and unsanctioned deceptive communication," *J. Nonverbal Behav.*, vol. 22, no. 3, pp. 189–204, Sep. 1998.

[33] Y. Freund, R. Iyer, R. Schapire, and Y. Singer, "An efficient boosting algorithm for combining preferences," *J. Mach. Learn. Res.*, vol. 4, pp. 933–969, Nov. 2003.

[34] G. A. Giordano, J. S. Stoner, R. L. Brouer, and J. F. George, "The influences of deception and computer-mediation on dyadic negotiations," *J. Comput. Mediat. Commun.*, vol. 12, no. 2, pp. 362–383, Jan. 2007.

[35] D. Goleman, *Social Intelligence: The New Science of Human Relationships*. New York, NY, USA: Random House, 2006.

[36] H. P. Grice, *Studies in the Ways of Words*. Cambridge, MA, USA: Harvard Univ. Press, 1989.

[37] M. A. Haj, J. Gonzalez, and L. S. Davis, "On partial least squares in head pose estimation: How to simultaneously deal with misalignment," in *Proc. IEEE Comput. Vis. Pattern Recognit. (CVPR)*, Providence, RI, USA, 2012.

[38] C. Hart, L. Hudson, D. Fillmore, and J. Griffith, "Managerial beliefs about the behavioral cues of deception," *Indiv. Differ. Res.*, vol. 4, no. 3, pp. 176–184, 2006.

[39] J. Huang, T. Zhang, and D. Metaxas, "Learning with structured sparsity," in *Proc. Int. Conf. Mach. Learn.*, 2009.

[40] M. L. Jensen, P. Lowry, and J. Jenkins, "Effects of automated and participative decision support in computer-aided credibility assessment," *J. Manag. Inf. Syst.*, vol. 28, no. 1, pp. 201–233, 2011.

[41] A. Kanaujia, Y. Huang, and D. N. Metaxas, "Tracking facial features using mixture of point distribution models," in *Proc. Indian Conf. Comput. Vis. Graph. Image Process. (ICVGIP)*, 2006, pp. 492–503.

[42] A. Kapoor and R. Picard, "A real-time head nod and shake detector," in *Proc. Percept. User Interfaces (PUI)*, New York, NY, USA, 2001.

[43] S. M. Kassin *et al.*, "Police interviewing and interrogation: A self-report survey of police practices and beliefs," *Law Human Behav.*, vol. 31, no. 4, pp. 381–400, 2007.

[44] S. Kawato and J. Ohya, "Real-time detection of nodding and head-shaking by directly detecting and tracking the between-eyes," in *Proc. Autom. Face Gesture Recognit.*, 2000, pp. 40–45.

[45] M. Kipp, "Spatiotemporal coding in anvil," in *Proc. 6th Int. Lang. Resour. Eval. (LREC)*, Marrakech, Morocco, 2008.

[46] D. Langleben, "Detection of deception with fMRI: Are we there yet?" *Legal Criminol. Psychol.*, vol. 13, no. 1, pp. 1–9, 2008.

[47] J. Larson, "The polygraph and deception," *Welfare Mag.*, vol. 18, no. 1, pp. 646–669, 1927.

[48] R. W. Levenson and A. Ruef, "Physiological aspects of emotional knowledge and rapport," in *Empathic Accuracy*, W. Ickes, Ed. New York, NY, USA: Guilford, 1997, pp. 44–72.

[49] T. Levine, A. Shaw, and H. Shulman, "Increasing deception detection accuracy with strategic questioning," *Human Commun. Res.*, vol. 36, no. 1, pp. 216–231, 2010.

[50] W. Marston, "Synstolic blood pressure symptoms of deception," *J. Exp. Psychol.*, vol. 2, no. 2, pp. 117–163, 1917.

[51] D. Metaxas, *Physics-Based Deformable Models: Applications to Computer Vision, Graphics, and Medical Imaging*, vol. 389. Boston, MA, USA: Springer, 1997.

[52] D. N. Metaxas and S. Zhang, "A review of motion analysis methods for human nonverbal communication computing," *Image Vis. Comput.*, vol. 31, nos. 6–7, pp. 421–433, 2013.

[53] N. Michael, M. Dilsizian, D. Metaxas, and J. Burgoon, "Motion profiles for deception detection using visual cues," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Heraklion, Greece, 2010.

[54] S. Milborrow and F. Nicolls, "Locating facial features with an extended active shape model," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Marseille, France, 2008, pp. 504–513.

[55] L. Morency, A. Quattoni, and T. Darrell, "Latent-dynamic discriminative models for continuous gesture recognition," in *Proc. Comput. Vis. Pattern Recognit. (CVPR)*, 2007.

[56] J. Navarro, "A four-domain model for detecting deception," *FBI Law Enforcement Bull.*, vol. 72, no. 6, p. 19, 2003.

[57] A. Ng, "Feature selection, l1 vs. l2 regularization, and rotational invariance," in *Proc. Int. Conf. Mach. Learn.*, New York, NY, USA, 2003.

[58] O. Oullier, G. de Guzman, K. Jantzen, J. S. Kelso, and J. Lagarde, "Social coordination dynamics: Measuring human bonding," *Soc. Neurosci.*, vol. 3, no. 2, pp. 178–192, 2008.

[59] M. Pantic and L. Rothkrantz, "Automatic analysis of facial expressions: The state of the art," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 12, pp. 1424–1445, Dec. 2000.

[60] A. Quattoni, M. Collins, and T. Darrell, "Conditional random fields for object recognition," in *Proc. Neural Inf. Process. Syst. (NIPS)*, 2004.

[61] D. Richardson and R. Dale, "Looking to understand: The coupling between speakers' and listeners' eye movements and its relationship to discourse comprehension," *Cognit. Sci.*, vol. 29, no. 6, pp. 1045–1060, 2005.

[62] D. Richardson, R. Dale, and K. Shockley, "Synchrony and swing in conversation: Coordination, temporal dynamics and communication," in *Embodied Communication in Humans and Machines*. Oxford, U.K.: Oxford University Press, 2008.

[63] M. Richardson, K. Marsh, R. Isenhower, J. Goodman, and R. Schmidt, "Rocking together: Dynamics of intentional and unintentional interpersonal coordination," *Human Mov. Sci.*, vol. 26, no. 6, pp. 867–891, 2007.

[64] R. Rienks, R. Poppe, and D. Heylen, "Differences in head orientation behavior for speakers and listeners: An experiment in a virtual environment," *ACM Trans. Appl. Percept.*, vol. 7, no. 2, pp. 1–13, 2010.

[65] J. M. Saragih, S. Lucey, and J. F. Cohn, "Deformable model fitting by regularized landmark mean-shift," *Int. J. Comput. Vis.*, vol. 91, no. 2, pp. 200–215, 2011.

[66] J. Shi and C. Tomasi, "Good features to track," in *Proc. Comput. Vis. Pattern Recognit. (CVPR)*, 1994, pp. 593–600.

[67] K. Shockley, M. Santana, and C. Fowler, "Mutual interpersonal postural constraints are involved in cooperative conversation," *J. Exp. Psychol. Human Percept. Perform.*, vol. 29, no. 2, pp. 326–332, 2003.

[68] C. Sutton, K. Rohanimanesh, and A. McCallum, "Dynamic conditional random fields: Factorized probabilistic models for labeling and segmenting sequence data," in *Proc. Int. Conf. Mach. Learn. (ICML)*, Banff, AB, Canada, 2004.

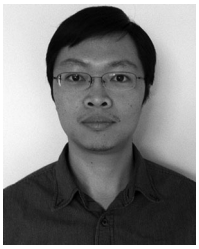
[69] F. Torre, Y. Yacoob, and L. Davis, "A probabilistic framework for rigid and non-rigid appearance based tracking and recognition," in *Proc. Int. Conf. Autom. Face Gesture Recognit.*, Grenoble, France, 2001.

[70] P. Tsiamyrtzis *et al.*, "Imaging facial physiology for the detection of deceit," *Int. J. Comput. Vis.*, vol. 71, no. 2, pp. 197–214, 2005.

[71] B. E. Turvey, "Introduction to terrorism: Understanding and interviewing terrorists," in *Criminal Profiling: An Introduction to Behavioral Evidence Analysis*, 3rd ed. Oxford, U.K.: Oxford Univ. Press, 2011, pp. 569-583.

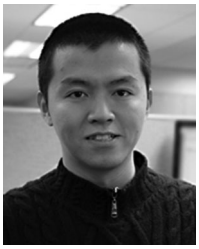
[72] V. N. Vapnik, *The Nature of Statistical Learning Theory*. New York, NY, USA: Springer, 1995.

[73] G. Varni, G. Volpe, and A. Camurri, "A system for real-time multimodal analysis of nonverbal affective social interaction in user-centric media," *IEEE Trans. Multimedia*, vol. 12, no. 6, pp. 576–590, Oct. 2010.
[74] P. Viola and M. Jones, "Robust real-time face detection," *Int. J. Comput. Vis.*, vol. 57, no. 2, pp. 137–154, 2004.
[75] A. Vrij, "Behavioral correlates of deception in a simulated police interview," *J. Psychol. Interdiscip. Appl.*, vol. 129, no. 1, pp. 15–28, 1995.
[76] A. Vrij, *Detecting Lies and Deceit: The Psychology of Lying and the Implications for Professional Practice*. New York, NY, USA: Wiley, 2000.
[77] A. Vrij, "Why professionals fail to catch liars and how they can improve," *Legal Criminol. Psychol.*, vol. 9, no. 2, pp. 159–181, 2004.
[78] A. Vrij, *Detecting Lies and Deceit: Pitfalls and Opportunities*. Chichester, U.K.: Wiley, 2008.
[79] A. Vrij, K. Edward, and R. Bull, "People's insight into their own behavior and speech content while lying," *Brit. J. Psychol.*, vol. 92, no. 2, pp. 373–389, 2001.
[80] A. Vrij, K. Edward, K. Roberts, and R. Bull, "Detecting deceit via analysis of verbal and nonverbal behavior," *J. Nonverb. Behav.*, vol. 24, no. 4, pp. 239–263, 2000.
[81] X. Wang, T. X. Han, and S. Yan, "An HOG-LBP human detector with partial occlusion handling," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Kyoto, Japan, 2009.
[82] C. H. White and J. K. Burgoon, "Adaptation and communicative design: Patterns of interaction in truthful and deceptive conversations," *Human Commun. Res.*, vol. 27, no. 1, pp. 9–37, 2001.
[83] F. Yang, J. Wang, E. Shechtman, L. Bourdev, and D. Metaxas, "Expression flow for 3D-aware face component transfer," *ACM Trans. Graph.*, vol. 27, no. 3, p. 60, 2011.
[84] P. Yang, Q. Liu, and D. Metaxas, "Rankboost with l1 regularization for facial expression recognition and intensity estimation," in *Proc. Int. Conf. Comput. Vis. (ICCV)*, Kyoto, Japan, 2009.
[85] X. Yu, J. Huang, S. Zhang, W. Yan, and D. N. Metaxas, "Pose-free facial landmark fitting via optimized part mixtures and cascaded deformable shape model," in *Proc. Int. Conf. Comput. Vis. (ICCV)*, Sydney, NSW, Australia, 2013.
[86] X. Yu, F. Yang, J. Huang, and D. Metaxas, "Explicit occlusion detection based deformable fitting for facial landmark localization," in *Proc. IEEE Int. Conf. Face Gesture Recognit.*, Shanghai, China, 2013.
[87] X. Yu *et al.*, "Is interactional dissynchrony a clue to deception: Insights from automated analysis of nonverbal visual cues," in *Proc. Hawaii Int. Conf. Syst. Sci. (HICSS)*, 2013.
[88] X. Yu *et al.*, "Automated analysis of interactional synchrony using robust facial tracking and expression recognition," in *Proc. Autom. Face Gesture Recognit.*, Shanghai, China, 2013.



Xiang Yu received the B.E. and the M.S. degrees from the Electronic Engineering Department, Tsinghua University, Beijing, China, in 2007 and 2010, respectively. He is currently pursuing the Ph.D. degree from the Department of Computer Science, Rutgers University, Piscataway, NJ, USA;, under the supervision of Prof. D. N. Metaxas.

His current research interests include computer vision and machine learning, especially on face alignment, tracking, and extended facial analysis.

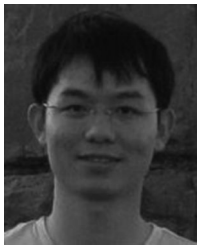


Shaoting Zhang received the B.E. degree from Zhejiang University, Zhejiang, China, in 2005, the M.S. degree from Shanghai Jiao Tong University, Shanghai, China, in 2007, and the Ph.D. degree in computer science from Rutgers University, Piscataway, NJ, USA, in 2012.

He is an Assistant Professor with the Department of Computer Science, The University of North Carolina (UNC) at Charlotte, Charlotte, NC, USA. Before joining UNC Charlotte, he was a Research Assistant Professor at the Department of Computer Science, Rutgers University, New Brunswick, NJ, USA, in 2012–2013. His current research interests include interface of medical imaging informatics, large-scale visual understanding, and machine learning.

Zhennan Yan is currently pursuing the Ph.D. degree from the Department of Computer Science, Rutgers University, Piscataway, NJ, USA, under the supervision of Prof. D. Metaxas, and co-advised by Prof. S. Zhang.

His current research interests include medical image processing and deformable models.



Fei Yang received the B.E. degree in computer science from Tsinghua University, Beijing, China, in 2003, the M.E. degree in computer science from The Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China, in 2006, and the Ph.D. degree in computer science from Rutgers University, Piscataway, NJ, USA, in 2013.

He is a Research Scientist of AI Group, Facebook, Cambridge, MA, USA. His current research interests include computer vision and machine learning, especially on face recognition, face tracking, and image classification.

Junzhou Huang received the B.E. degree in control science and engineering from Huazhong University of Science and Technology, Wuhan, China, in 1996, the M.S. degree in pattern recognition and intelligence systems from the Institute of Automation, Chinese Academy of Sciences, Beijing, China, in 2003, and the Ph.D. degree in computer science at Rutgers University, Piscataway, NJ, USA.

Since 2011, he has been Assistant Professor at the Computer Science and Engineering Department, the University of Texas at Arlington, Arlington, TX, USA.

Norah E. Dunbar received the Ph.D. degree from the University of Arizona, Tucson, AZ, USA, in 2000.

She is a Professor of Communication at the University of Oklahoma, Norman, OK, USA, and relocated to the University of California Santa Barbara, Santa Barbara, CA, USA, in 2014. Her current research interests include nonverbal and interpersonal communication, with special emphasis on dominance and power relationships, interpersonal synchrony, and interpersonal deception. She has published over 35 journal articles and book chapters, in journals including *Journal of Computer-Mediated Communication, Journal of Nonverbal Behavior, Communication Research, Communication Monographs,* and *Journal of Social and Personal Relationships*. Her work has also been featured in anthologies such as *The Encyclopedia of Communication Theory, The Sage Handbook of Nonverbal Communication,* and *The Sourcebook for Nonverbal Research Measures*.

Matthew L. Jensen received the Ph.D. degree from the University of Arizona, Tucson, AZ, USA, in 2007.

He is an Assistant Professor of Management Information Systems at the Price College of Business, University of Oklahoma, Norman, OK, USA, and a Co-Director of the Center for Applied Social Research at the University of Oklahoma. His current research interests include understanding credibility and deception in virtual environments, computer-aided decision making, and human-computer interaction. His research has been published in *Information Systems Research, Journal of MIS, Group Decision and Negotiation*, and various IEEE outlets.

Judee K. Burgoon received the bachelor's degree with a double major in speech and english and a double minor in social studies and education, from Iowa State University, Ames, IA, USA, in 1970, the master's degree in speech communication from Illinois State University, Normal, IL, USA, in 1972, and the Ph.D. degree in the joint fields of communication and educational psychology from West Virginia University, Morgantown, WV, USA, in 1974.

She is a Professor of Communication, Family Studies, and Human Development at the University of Arizona, Tucson, AZ, USA, where she is also a Director of Research for the Center for the Management of Information and Site Director for the Center for Identification Technology Research, a National Science Foundation Industry/University Cooperative Research Center. She has authored or edited 13 books and monographs and over 300 published articles, chapters, and reviews related to nonverbal and verbal communication, deception, and computer-mediated communication. Her research has been funded by the National Science Foundation, Department of Defense, Department of Homeland Security, and the Office of the Director of National Intelligence, among others.

Dimitris N. Metaxas received the B.E. degree from the National Technical University of Athens, Athens, Greece, in 1986, the M.S. degree from the University of Maryland, College Park, MD, USA, in 1988, and the Ph.D. degree from the University of Toronto, Toronto, ON, Canada, in 1992.

He is a Professor at the Computer Science Department, Rutgers University, Piscataway, NJ, USA, and is directing the Computational Biomedicine Imaging and Modeling Center. His current research interests include the development of formal methods upon which computer vision, computer graphics, and medical imaging can advance synergistically.