Analyzing Thermal and Visual Clues of Deception for a Non-Contact Deception Detection Approach

Mohamed Abouelenien
Computer Science and Engineering
University of Michigan
zmohamed@umich.edu

Rada Mihalcea
Computer Science and Engineering
University of Michigan
mihalcea@umich.edu

Mihai Burzo
Mechanical Engineering
University of Michigan-Flint
mburzo@umich.edu

ABSTRACT

With increased levels of security threats and the long-term consequences of falsely accusing the innocent and freeing the guilty, there is a growing need for reliable and efficient deception detection systems. Polygraph tests are invasive and require elongated time and human expertise, which is subject to bias and error. In this paper, we analyze thermal and visual clues of deception using a dataset collected from 30 subjects and multiple scenarios. We analyze expressions and other visual features and provide the first comparison between thermal facial regions to identify areas with higher capability of indicating deceit. Our experimental results show that our non-contact feature fusion model outperforms traditional physiological measurements, paving the road for non-invasive deception detection methodologies.

Categories and Subject Descriptors

I.2 [**Artificial Intelligence**]: Miscellaneous

Keywords

deception; thermal; visual; physiological; multimodal

1. INTRODUCTION

Deception detection has been receiving increased attention from different research communities. Existing methodologies rely mainly on polygraph tests that extract physiological measurements. However, in many cases these results were proven to be incorrect due to the discovery of new evidence and due to other contributing factors such as stress, exhaustion, and intrusiveness of the contact-based sensors [11, 12]. Additionally, decisions of whether a deceptive behavior was present were made by human experts, which is usually subject to bias and error.

Researchers in deception detection have also looked for clues in facial expressions, gestures, thermal variations, word patterns and consistency of statements, among other methods that can be related to deceptive behavior [8, 13, 14]. Visual clues such as micro-expressions (defined as involuntary expressions that exist for a short period of time), expression intensity, movements, and other local features were shown to have the capability of identifying deceit [4,8]. Recently, thermal imaging was also used to provide clues of deception [14]. However, there seems to be disagreement on which facial area provides features with higher capability of discriminating between deception and truthfulness.

The research reported in this paper departs from earlier work and adds a new dimension through an extensive analysis of the visual modality. In particular, the paper makes three main contributions. First, we analyze 149 instances of deceptive and truthful responses from 30 subjects. Second, we provide the first study of the relationship between thermal variations and deception by tracking three different facial areas with rotation, namely, the whole face, the forehead, and the periorbital areas. Third, this is the first work to fuse automatically-extracted thermal and visual features to detect deceit.

2. RELATED WORK

Recent research in deception detection has focused on visual non-contact based approaches. Ekman [4] analyzed micro-expressions and identified their relation to deception. Bartlett et al. [3] developed a real-time system to detect spontaneous facial expressions which occur with a deceptive action. Owayjan et al. [8] utilized dynamic geometric-based templates on deceptive videos to extract features identifying deceit.

In thermal imaging, Pavlidis et al. [10] extracted thermal features from the face using a high definition thermal camera to analyze whether differences occur when a subject responded truthfully or deceptively.

Attempts were also made to identify which areas in the face provided discriminant thermal features. Park et al. [9] averaged the maximum temperatures in the periorbital region of video recorded subjects to distinguish between truthful and deceptive behavior. Zwiggelaar [13] detected the corners of the eyes to track the surrounding region using thermal imaging, and reported improved deception detection rates for within-person responses.

3. DATASET

Deceptive and truthful responses were collected using a thermal camera and two visual cameras as well as four physiological bio-sensors. The dataset consists of recordings of 30 subjects including 25 males and 5 females using three scenarios, "Abortion," "Best Friend," and "Mock Crime". Details

PETRA '16, June 29-July 01, 2016, Corfu Island, Greece
 ISBN 978-1-4503-4337-4/16/06...$15.00
DOI: http://dx.doi.org/10.1145/2910674.2910682

of the dataset collection can be found in [1, 2].

4. METHODOLOGY

4.1 Thermal Clues

In order to identify which region of the face provides features which have higher capability of indicating deceit, the subjects' faces were segmented into three regions: whole face, forehead, and periorbital. This was followed by tracking these regions throughout the response. Finally, feature vectors were generated by creating thermal maps for each region of interest.

4.1.1 Tracking

Initially, the whole face, the forehead, and the periorbital regions were manually located for each subject from the first recorded frame by specifying the pixel location of their bounding boxes. The actual response of each subject was preceded by at least one minute of recording with no activity in order to provide a baseline for the thermal activity in regular conditions. We will refer to this period as the "normalization minute." Once the three regions were located, points of interest were detected using the Shi-Tomasi corner detection algorithm, which in this case were found at locations with varying temperatures in the three regions.

Once the interesting points were detected, they were tracked using a fast Kanade Lucas Tomasi (KLT) tracking method.

We computed the Forward-Backward Error [6] by tracking the points back and forth among the frames to eliminate outliers and uncertain points. To map the interesting points from one frame to the next, geometric transformation [5] was applied to specify the location of the new boundary box for each of the three regions. We set a threshold of 95% as a rate of correct points matching between successive frames.

4.1.2 Extracting Thermal Features

The rectangular area that masked the polygon was geometrically located and cropped. The backgrounds and the extended regions of the polygon due to masking were eliminated. This was performed by binarizing the image and multiplying it by its original cropped rectangular image.

For each of the three regions, we decided to uniformly sample 500 frames for feature extraction from the response of each subject. Another set of 500 images were sampled from the frames of the normalization minute.

In order to measure the thermal variations in our three regions of interest as a potential indicative of deceptive behavior, a thermal map was created using two pixel representations, grayscale and Hue Saturation Value (HSV). The map was formed by extraction of the mean of the pixels values in the region of interest, the maximum pixel value representing the highest temperature, the minimum pixel value representing the lowest temperature, the difference between the maximum and minimum values, the mean of the 10% highest pixel values representing the mean of 10% highest temperatures, and a histogram over the values of the pixels, which resulted in a total of 260 grayscale features and 780 HSV features. The thermal features were averaged for each region of interest for each response, and the histograms were normalized to form a probability distribution over the bins.

A thermal correction process was performed to account for the normal inter-personal temperature variations. The same set of features was extracted from the 500 frames of the normalization minute. To achieve the thermal correction, the features from the responses were divided by the corresponding features from the normalization minute.

4.2 Visual Features

In order to detect facial expressions and other features we decided to use the Computer Expression Recognition Toolbox (CERT) [7]. CERT is a software tool that detects universal facial expressions and facial action units. These units are specified by the Facial Action Coding System, which was developed by psychologists and behavioral scientists and provided taxonomy of features using muscle movements. Examples of these action units include inner brow raiser, nose wrinkle, lip raiser, cheek raiser, chin raiser, eye widen, and others.[1] Additionally, CERT provides twelve facial expressions such as yaw, pitch, roll, smile detector, anger, contempt, disgust, fear, joy, sad, surprise, and neutral.

In addition to the CERT features, we also extract features that are more personalized to the subjects. In particular, we calculated the normalized blinking rates and the mean head orientation angle along the entire length of the response. Templates of the subjects' open eyes were cropped from a single frame. Normalized cross correlation coefficient was calculated by using the template of each subject as a sliding window through each frame of his response.

The max correlation coefficient of each frame was computed for each video response. As the coefficients approached a value of one, a high similarity between the template and the eyes region was found and the subjects were assumed not to be blinking.

As head movements were shown to contribute to identifying deceit, we opted to detect the mean head rotation angle for the subjects during their responses. For this task, template images containing the subjects' faces in a forward position without tilt were used. Interesting corner points in the templates as well as the individual frames were detected using the minimum eigenvalue algorithm. Feature descriptors were extracted to describe the interesting points after eliminating noisy points and outliers. Geometric transformation between the template faces and the faces in each frame was performed and hence, the rotation angle was specified

Following the visual feature extraction process, we obtained 40 CERT features including 28 action units and 12 global facial expressions, one feature for the blinking rates, and one feature with the mean head rotation angle for each of the 149 responses.

4.3 Physiological Features

We also collected physiological features, using the output produced by the four bio-sensors. In addition to raw data, statistical measurements are also extracted and stored, such as means, maximum, minimum, power means, and standard deviations, for a total of 60 physiological features.

4.4 Deception Classification

Our study hypothesizes that there will be subtle variations in the subjects' thermal and visual features as they respond deceptively. The 149 thermal and visual feature vectors were used to train a decision tree classifier. Given the size of our dataset, we opted to use a leave-one-out cross validation scheme to report the overall accuracy, as well as the recall of each of the deception and truthfulness classes.

[1] http://www.cs.cmu.edu/ face/facs.htm

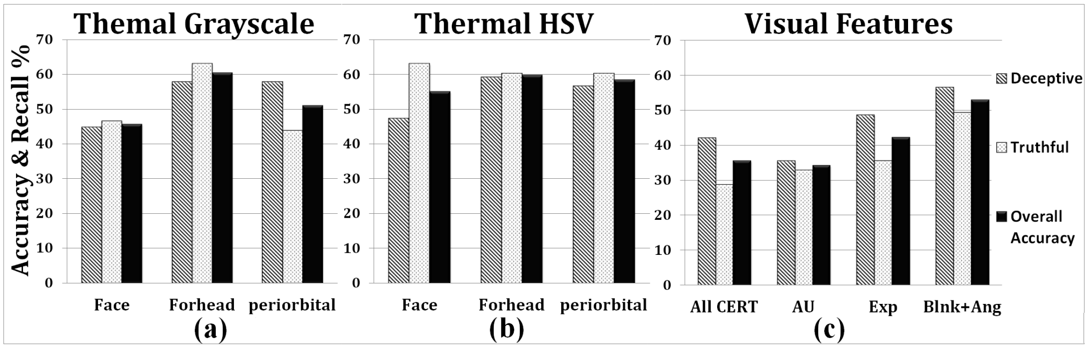

Figure 1: Accuracy as well as recall of the truthful and deceptive classes for (a) thermal gray (b) thermal HSV (c) visual features. AU, Exp, Blnk+Ang denote Action Units (AU), expressions, blink and head angle.

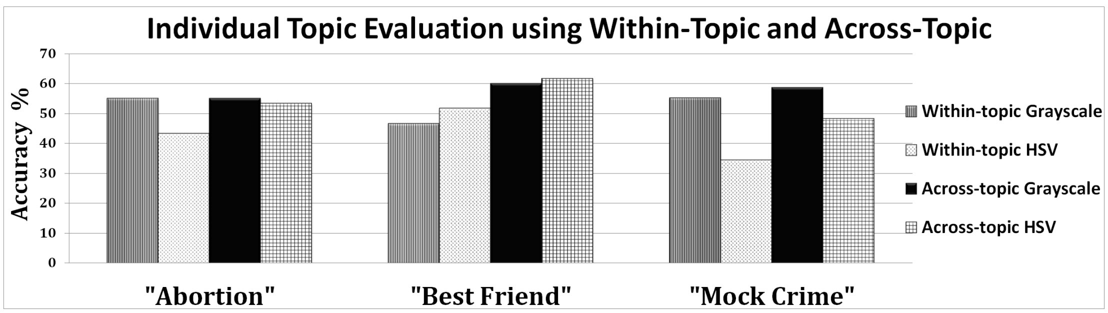

Figure 2: Overall accuracy for each topic using within-topic and across-topic training using fused forehead thermal and visual features.

Additionally, we report the performance of individual topics to analyze the effect of each scenario, as well as the role of the interviewer's involvement on performance. Moreover, we evaluate different scenarios using across-topic training scheme, where instances from one topic are tested while instances from other topics are used for training.

5. EXPERIMENTAL DISCUSSION

Figure 1 presents the accuracy and per-class recall using (a) thermal grayscale; (b) thermal HSV; and (c) CERT, blinking rates, and head orientation features. It can be noted that using thermal features from the forehead outperforms features from the whole face and the periorbital areas. In general, the HSV features provide richer information compared to the grayscale features, which is reflected in their performance. The performance using the grayscale whole face features is below the baseline.

In contrast, the visual features experience deteriorated performance in identifying deceit. In particular, the performance of the CERT features indicates that facial muscle movements and expressions fail in discriminating between deception and truthfulness. Moreover, we evaluate the performance of the action units and facial expressions separately. However, the performance is still below the baseline. On the other hand, the blinking rates and head orientation provide a performance which is above that of random guessing, which indicates that personalized measurements can be a promising indicator of deceit.

Based on these results, we decided to fuse the forehead features with the blinking rates and head orientation features. The performance of this classifier is shown in Figure 3. Cross-referencing Figure 3 with Figure 1 (a) and (b), we note the effect of the combined classifier: for instance, fusion of forehead grayscale thermal and visual features achieves an accuracy improvement of 2.2% over using thermal features and 16.5% over visual features. This combined classifier is used in all subsequent analyses described below.

To gain additional insights, we performed two analyses. First, we explored the role played by topic (or domain) in deception detection. Figure 2 shows the overall accuracy for each individual topic. Within-topic training refers to using a leave-one-out validation scheme on instances of the same topic. On the other hand, across-topic training for the "Abortion" topic, for example, refers to testing its instances using a classifier trained with "Best Friend" and "Mock Crime" instances.

The figure leads to two observations. One is that the performance is topic-independent and larger training data is more beneficial. Although across-topic training uses instances from a different topic, yet it provides more training instances which in general improves the performance for all scenarios compared to within-topic training. The other observation is that there is no major difference in performance among topics. However, it can be noted that "Best Friend" achieves a slightly improved performance while "Mock Crime" suffers from the lack of sufficient number of instances. The improvement in "Best Friend" using thermal and visual features can be attributed to the topic being more personal compared to other scenarios.

The second analysis consists of a comparison with a classifier based on physiological features alone. Figure 3 shows the accuracy and class recall using the physiological features, as compared to the fusion of forehead thermal and

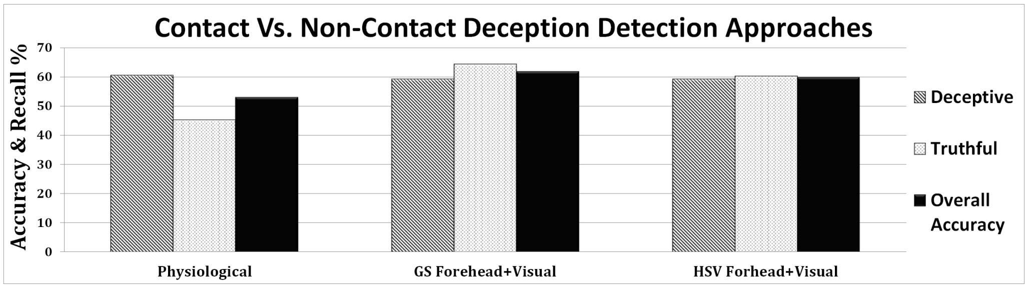

Figure 3: Accuracy as well as recall of each of the truthful and deceptive classes using contact physiological sensors vs. using non-contact fused thermal and visual features.

visual features.The overall accuracies of the combined thermal and visual classifiers (GS+visual 61.74%, HSV+visual 59.73%) clearly exceed the one obtained with the physiological features (53.02%), for error rates reduction of 18.56% and 14.28% respectively, which demonstrates the advantage of our proposed non-contact deception detection method as compared to earlier contact-based measurements.

6. CONCLUSION

This paper analyzed thermal and visual deception clues in order to create an automated improved non-contact system that can be deployed in airports, interrogations rooms, courts, etc. This paper provided a novel dataset for deception detection using three different scenarios, and presented the first comparative study of facial regions that are capable of indicating deceit. The paper also presented an analysis of automatically-generated visual features associated with deception, and a model which integrated non-contact features for improved performance.

Our experimental results indicate that the thermal forehead region provides richer information for discriminating between deception and truth. One of the factors that can be associated to this improvement is the presence of facial hair and other features found in the whole face and the periorbital areas. Moreover, subjects were capable of controlling their visual features compared to the thermal ones in order to hide their deceptive behavior. However, based on our analysis there were some muscle movements and facial expressions which can be associated with deception. As visual features were normalized for each subject, the performance improved. Moreover, the fusion of thermal and visual features improves the deception detection rates. This indicates that a non-contact deception detection approach is promising, and our method can form the basis for future real-life deception detection applications.

7. ACKNOWLEDGMENTS

This material is based in part upon work supported by National Science Foundation awards #1344257 and #1355633, and by grant #48503 from the John Templeton Foundation. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation or the John Templeton Foundation.

8. REFERENCES

[1] M. Abouelenien, R. Mihalcea, and M. Burzo. Trimodal analysis of deceptive behavior. In *Proceedings of the 2015 ACM on Workshop on Multimodal Deception Detection*, WMDD '15, pages 9–13, New York, NY, USA, 2015. ACM.

[2] M. Abouelenien, V. Perez-Rosas, R. Mihalcea, and M. Burzo. Deception detection using a multimodal approach. In *16th ACM International Conference on Multimodal Interaction, ICMI 2014*, November 2014.

[3] M. Bartlett, G. Littlewort, M. Frank, C. Lainscsek, I. Fasel, and J. Movellan. Automatic recognition of facial actions in spontaneous expressions. *Journal of Multimedia*, 1(6):22–35, 2006.

[4] P. Ekman. *Telling Lies: Clues to Deceit in the Marketplace, Politics and Marriage.* Norton, W.W. and Company, 2001.

[5] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision.* Cambridge books online. Cambridge University Press, 2003.

[6] Z. Kalal, K. Mikolajczyk, and J. Matas. Forward-backward error: Automatic detection of tracking failures. In *20th International Conference on Pattern Recognition (ICPR)*, pages 2756–2759, Aug 2010.

[7] G. Littlewort, J. Whitehill, T. Wu, I. Fasel, M. Frank, J. Movellan, and M. Bartlett. The computer expression recognition toolbox (cert). In *2011 IEEE International Conference on Automatic Face Gesture Recognition and Workshops (FG 2011)*, pages 298–305, March 2011.

[8] M. Owayjan, A. Kashour, N. Al Haddad, M. Fadel, and G. Al Souki. The design and development of a lie detection system using facial micro-expressions. In *2012 2nd International Conference on Advances in Computational Tools for Engineering Applications (ACTEA)*, pages 33–38, Dec 2012.

[9] K. Park, H. Suk, H. Hwang, and J. Lee. A functional analysis of deception detection of a mock crime using infrared thermal imaging and the concealed information test. *Frontiers in Human Neuroscience*, 7, 2013.

[10] I. Pavlidis, N. L. Eberhardt, and J. A. Levine. Human behavior: Seeing through the face of deception. *Nature*, 415(6867):35–35, 2002.

[11] V. Pérez-Rosas, M. Abouelenien, R. Mihalcea, Y. Xiao, C. Linton, and M. Burzo. Deception detection using real-life trial data. In *17th ACM International Conference on Multimodal Interaction, ICMI 2015*, November 2015.

[12] V. Pérez-Rosas, M. Abouelenien, R. Mihalcea, Y. Xiao, C. Linton, and M. Burzo. Verbal and nonverbal clues for real-life deception detection. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 2336–2346, Lisbon, Portugal, September 2015. Association for Computational Linguistics.

[13] B. Rajoub and R. Zwiggelaar. Thermal facial analysis for deception detection. *IEEE Transactions on Information Forensics and Security*, PP(99):1–10, 2014.

[14] Y. Zhou, P. Tsiamyrtzis, P. Lindner, I. Timofeyev, and I. Pavlidis. Spatiotemporal smoothing as a basis for facial tissue tracking in thermal imaging. *IEEE Transactions on Biomedical Engineering*, 60(5):1280–1289, May 2013.