Trimodal Analysis of Deceptive Behavior

Mohamed Abouelenien
Computer Science and Engineering
University of Michigan
Ann Arbor, MI 48109, USA
zmohamed@umich.edu

Rada Mihalcea
Computer Science and Engineering
University of Michigan
Ann Arbor, MI 48109, USA
mihalcea@umich.edu

Mihai Burzo
Mechanical Engineering
University of Michigan, Flint
Flint, MI 48502, USA
mburzo@umich.edu

ABSTRACT

The need arises for developing a more reliable deception detection system to address the shortcomings of the traditional polygraph tests and the dependability on physiological indicators of deceit. This paper describes a new deception detection dataset, provides a novel comparison between three modalities to identify deception including the visual, thermal, and physiological domains, and analyzes whether certain facial areas are more capable of indicating deceit. Our experimental results show a promising performance especially with the thermal modality, and provide guidelines for our data collection process and future work.

Categories and Subject Descriptors

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

Keywords

trimodal; deception; thermal, visual, physiological

1. INTRODUCTION

Deceptive behavior is found on a daily basis in different human interactions. Applications such as security, business, online interactions, and criminal investigation triggered research interest in different fields such as computer vision, psychology, physiology, and language processing. While earlier research focused on polygraph tests and physiological measurements as a contact-based mean for lie detection, recent methodologies analyzed gestures, facial expressions, and language usage [18, 19, 15, 12].

Visual clues of deception include facial emotions, expression intensity, hands and body movements, and microexpressions – defined as spontaneous expressions that exist for a short amount of time. These modalities were shown to be capable of discriminating between deceptive and truthful behavior [3, 9]. More recently, thermal imaging analysis was used to identify deception relying on an increase in the blood flow in facial veins [6, 19]. However, most of the visual clues rely heavily on the manual analysis performed by human experts and psychologists. Moreover, research results so far have been conflicted regarding the visual clues and facial thermal areas that provide the best discriminative features. Furthermore, employing polygraph tests was shown to be unreliable in many cases as it requires decisions from human experts, which is subject to bias and error.

Motivated by the aforementioned challenges and shortcomings, this paper targets three goals. First, we develop a novel deception dataset using different modalities. Second, we provide a novel comparison between the non-contact, automatically-generated visual and thermal features extracted from the dataset. Additionally, we compare their performance to that achieved using physiological measurements. Third, we track and analyze both the facial and periorbital thermal regions to determine whether specific areas in the face are more capable of indicating deceit.

2. RELATED WORK

Recent research focused on non-contact approaches as a consequence of the shortcomings and errors of the contact-based methods and polygraph tests. Bartlett et al. [2] introduced a system to detect spontaneous facial reactions occurring with a deceptive action and showed that deception detection rates improved with spontaneous expression data rather than posed expressions. Ekman [3] analyzed microexpressions and identified their relation to deceptive behavior. Pietikainen [13] applied temporal interpolation using kernel learning to create a lie detection system. Meservy et al. [8] used a hierarchical Hidden Markov Model which utilized blob analysis to detect deceit based on facial skin color.

More recently, thermal features were extracted from the facial area in order to detect deceit. Garbey et al. [4] developed a bioheat transfer model to detect the anatomy of the facial blood vessels in order to identify deceptive behavior. Warmelink et al. [17] extracted statistical thermal features to create a deception system that can be deployed in airports, and reached approximately 60% accuracy.

In order to identify whether certain regions of the face are capable of indicating deception, Pavlidis and Levine [11] used thermodynamic modeling to detect blood flow rates in the periorbital area. Zwiggelaar [15] tracked the surrounding regions of the eyes corners to detect deception in within-person responses. Jain et al. [6] computed the mean of the 10% highest temperatures in the area surrounding the tear ducts to differentiate between deception and truthfulness. By averaging the maximum temperatures in the eyes

WMDD'15, November 09-13, 2015, Seattle, WA, USA

http://dx.doi.org/10.1145/2823465.2823470.

regions, Park et al. [10] were able to identify deceptive behavior.

3. DATASET

A thermal camera and two visual cameras were used to record deceptive and truthful responses. Additionally, we connected four contact-based physiological bio-sensors to the participants. A blood volume pulse, skin conductance, and skin temperature sensors were attached to the fingers of the non-dominant hand of the subject. The abdominal respiration sensor was placed to surround the thoracic region. The physiological features were extracted using the Biograph Infinity Physiology suite[1] at a rate of 2048 Hz.

Two Logitech visual cameras were used to record the participants' responses. One was used to record the facial area while the other recorded the upper body. The cameras had a frame rate of 30 fps and a resolution of 800x600. Thermal videos were acquired using a FLIR Thermovision A40 thermal camera with a frame rate of 30 fps and a resolution of 340x240.

3.1 Participants

Participants were graduate and undergraduate students with different ethnic backgrounds including Caucasian, Hispanic, Asian, and African-American. The age range was between 22 and 38 years. The dataset consisted of a total number of 30 subjects including 25 males and 5 females.

3.2 Topics

Participants were asked to sit comfortably in a seat at the experiment station, and were instructed to respond truthfully and deceptively to the introduced topics. The experimental procedure was explained to them and they were asked to avoid excessive movements with their bodies in order to avoid distortion of the physiological signals and to keep them in the field of view of the cameras.

Three topics were prepared for the subjects, namely, "Abortion," "Best Friend," and "Mock Crime." Participants were told the topic matter and were asked to respond freely once truthfully and once deceptively for each of the first two topics. The interviewer did not have any role other than introducing the topic to the subjects. However, the interviewer had more involvement in the "Mock Crime" scenario by interrogating the subjects. For this scenario, the subjects provided either a deceptive or a truthful response.

3.2.1 Abortion

The subjects were asked to provide a verbal response on their opinion on abortion assuming they were in a debate on this particular topic. They were asked to first provide a truthful response of their real opinion on abortion whether supportive or opposing. Then they were asked to respond deceptively on the same topic acting as though it was their true opinion. Each response was recorded independently.

3.2.2 Best Friend

For this topic, the participants were instructed to truthfully describe their best friend. After recording the first session, they were asked to lie about a person they dislike and describe him/her as though he/she was their best friend. Hence, both descriptions were positive about a certain individual, however, the first response was truthful while the second was deceptive.

3.2.3 Mock Crime

For this topic, participants were assigned randomly to either provide a deceptive or a truthful response. An envelope containing a $20 bill was placed on a table in an office for the deceptive scenario while an empty envelope was placed for the truthful scenario. All participants were asked to deny taking the bill, which was truthful when the envelope was empty and deceptive otherwise. This was followed by a one-on-one interview using the following questions:

1. Are the lights on in this room?
2. Regarding that missing bill, do you intend to answer truthfully each question about that?
3. Prior to 2012, did you ever lie to someone who trusted you?
4. Did you take that bill?
5. Did you ever lie to keep out of trouble.
6. Did you take the bill from the private area of the lab?
7. Prior to this year, did you ever lie for personal gain?
8. What was inside the white envelope?
9. Please describe step by step, in as much detail as you can, what you did while you were behind the white board. Please aim at a clear description of about 2-3 minutes.
10. Do you know where that missing bill is now?

Hence, each of the 30 participants provided two responses for each of the "Abortion" and "Best Friend" topics and a single response for the "Mock Crime" scenario. One "Mock Crime" response was not included due to an error in the data collection. Therefore, the final dataset contained 149 responses including 30 truthful and 30 deceptive responses for each of the "Abortion" and "Best Friend" scenarios, and 16 deceptive and 13 truthful "Mock Crime" responses. Additional dataset analysis is provided in [1].

4. METHODOLOGY

4.1 Physiological Measurements

A set of physiological measurements including the heart rate, skin conductance, skin temperature, and respiration rate were collected using the output produced by the four sensors. Statistical measurements were extracted from the raw data such as the mean, maximum, minimum, power means, and standard deviations to form a total of 60 physiological features. The features of each response were averaged to result in a single feature vector.

4.2 Thermal Features

To determine whether a specific region in the face had higher capability of indicating deceit, the participants' frames were segmented into two regions, the whole face and the periorbital region. These two areas were then tracked during the entire response of each participant. Finally, thermal maps were created from these regions and presented as a feature vector for each response.

First, the whole face and the periorbital regions were manually located for each participant from the first recorded frame by specifying the pixel locations of their bounding

[1] http://www.thoughttechnology.com/physsuite.htm

boxes. One minute with no activity, which preceded every recording, was used to extract baseline thermal features under regular conditions. We will refer to this period as the "normalization minute." Once the two regions were located, points of interest were detected using the Shi-Tomasi corner detection algorithm. These points were found at locations with varying temperatures in the facial and periorbital regions.

This was followed by tracking these points using a fast Kanade-Lucas-Tomasi (KLT) tracking method [16]. The method assumed the occurrence of a small displacement between a given frame and a successive displaced frame, which was suitable for our experimental design.

The displacement between successive frames was estimated such that the error was minimized using the intensity second moment matrix, and was used to track the interesting points during the video responses. A threshold of 95% was enforced as a rate of correct points matching between successive frames. In order to specify the new location of the bounding boxes of our two regions in the incoming frame, geometric transformation [5] was applied.

Given that the tracked area could be polygon-shaped due to head movements, the rectangular region masking the polygon was geometrically located and cropped. Using image binarization, the backgrounds were eliminated to improve the precision of the tracked regions. In particular, the cropped rectangular areas were multiplied by the binarized images to eliminate areas outside our two regions of interest.

We uniformly sampled 500 frames for feature extraction from the response of each participant to increase the efficiency of the process. An additional set of 500 images were sampled from the frames of the normalization minute.

In order to specify the relation between the thermal variations in the facial and periorbital regions and deceptive behavior, a thermal map was created using the Hue Saturation Value (HSV) pixel representations. The thermal 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. This representation resulted in a total of 780 HSV features. It should be noted that pixels with a value of zero were eliminated. The thermal features were averaged for each of the two regions for each response. The histograms were normalized to form a probability distribution over the bins.

As different persons have different thermal temperatures under regular conditions, 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. The features from the responses were divided by the corresponding features from the normalization minute in order to achieve thermal correction. This resulted in a feature vector indicating the variation in the thermal distribution in the facial and periorbital area, whether it was a thermal increase or decrease.

4.3 Visual Features

In order to automatically identify multiple facial expressions and actions, 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 were specified by the Facial Action Coding System, which was developed by psychologists, and provided taxonomy of facial features using muscle movements. Examples of these action units included inner brow raiser, nose wrinkle, lip raiser, cheek raiser, and others.[2] Moreover, CERT provided twelve facial expressions such as yaw, pitch, roll, smile detector, anger, contempt, disgust, fear, joy, sad, surprise, and neutral. The software tool detected faces in each frame followed by specifying the eyes corners, nose, and mouth corners and center. The algorithm determined the log-likelihood ratio of the presence of these regions in specific locations, which specified the intensity of the facial actions. The global facial expressions were specified using a combination of different action units.

The automated visual feature extraction process resulted in a set of 40 CERT features including 28 action units and 12 global facial expressions for each of the 149 responses.

4.4 System Training and Classification

The 149 thermal and visual feature vectors are used to train a decision tree classifier as recommended in [14] for deception detection. We opted to use a leave-one-out cross validation scheme to report the overall accuracy, as well as the recall of the deception and truthfulness classes, given the size of our dataset. The baseline performance is 51% and 49% for the deception and truth classes, respectively. We hypothesized that there would be subtle variations in the participants' thermal and visual responses as they acted deceptively. Additionally, we report the performance of individual topics to analyze the effect of each topic, as well as the role of the interviewer's involvement in the "Mock Crime" scenario. Furthermore, we evaluate different topics using an across-topic training scheme, where the classifier is trained using instances from two topics while instances from the third topic are used for testing.

5. EXPERIMENTAL RESULTS

Table 1: Percentage accuracy and recall of the deceptive and truthful classes for individual modalities and for the physiological, visual, and thermal periorbital modalities combined "All". The best performance is highlighted in bold.

	Phys	AU	Exp	CERT	Tface	Tperi	All
Accuracy	53.0	34.2	42.3	35.6	55.0	**58.4**	56.4
Deception	**60.5**	35.5	48.7	42.1	47.4	56.6	55.3
Truthful	45.2	32.9	35.6	28.8	**63.0**	60.3	57.5

Table 1 lists the percentage accuracy as well as the recall of the deceptive and truthful classes for individual modalities such as the physiological, CERT actions units, CERT expressions, all CERT features, thermal face, and thermal periorbital denoted by "Phys", "AU", "Exp", "CERT", "Tface", and "Tperi", respectively. The table also lists the perfor-

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

mance of the physiological, visual, and thermal periorbital modalities combined denoted by "All".

The table shows that the best overall accuracy is achieved by the thermal periorbital modality followed by the modalities combined. The thermal face and the physiological modalities achieve a performance that is slightly higher than random guessing. Moreover they obtain the highest recall for the truthful and deceptive classes, respectively. The visual modality suffers a deteriorated performance for all types of features. The performance indicates that the thermal features are the most promising lead of identifying deception.

On the other hand, the participants are clearly able to control their facial muscle movements and expressions as to hide any deceptive behavior. As the thermal and physiological features are harder to control, they are more capable of discriminating between deception and truthfulness.

In order to further analyze whether specific facial muscle movements or expressions can be considered as potential clues of deception, backward feature selection is used for all CERT features. The algorithm specifies a list of eight action units and six expressions, which provides the highest accuracy of 63.09%. The list includes brow lowering, chin raising, cheek raising, lip puckering, eye closure, distress brow, left turning AU 10, left AU 14, yaw, roll, contempt, disgust, sadness, and neutral. The potential of these specific features in indicating deceit will be considered for our future work and when more data is collected.

Table 2: Percentage accuracy and recall of the deceptive and truthful classes for the thermal periorbital modality and all three modalities combined for individual topics and across-topic learning.

	AB	BF	MC	Test AB	Test BF	Test MC
Thermal Periorbital						
Accuracy	41.7	45.0	41.4	51.7	51.7	69.0
Deception	43.3	40.0	50.0	56.7	50.0	68.8
Truthful	40.0	50.0	30.8	46.7	53.3	69.2
Physiological + Visual + Thermal Periorbital						
Accuracy	40.0	40.0	31.0	51.7	51.7	79.3
Deception	43.3	40.0	37.5	46.7	50.0	81.3
Truthful	36.7	40.0	23.1	56.7	53.3	76.9

Table 2 lists the percentage accuracy as well as the recall of the deceptive and truthful classes using the thermal periorbital modality and the modalities combined for the three individual topics, abortion, best friend, and mock crime denoted by "AB", "BF", and "MC", respectively. Additionally, the table shows the performance of the across-topic learning scheme. For example, "Test AB" indicates that the best friend and mock crime instances were used for training, while the abortion instances were used for testing.

The table clearly indicates that the performance of different modalities experience a deteriorated performance for the per-topic analysis. This can be attributed to the small sample size used for each individual topic, which does not provide the classifier with enough data to learn from. However, as the sample used for the across-topic learning increases, the performance improves and exceeds that of random guessing.

The across-topic analysis also shows that the performance using the thermal periorbital and combined modalities is topic independent. The results are more dependent on the data size compared to the topics used. For instance, when the classifier is trained with the abortion and best friend instances and tested with the mock crime instances, the accuracy and recall increase significantly for the thermal periorbital and the combined modalities, reaching close to 80% overall accuracy.

6. CONCLUSION

Motivated by the need of developing deception detection approaches that are more diverse and reliable, this paper targeted three goals. First, the paper provided a deception detection dataset which utilized multiple modalities for recording. Second, we conducted a novel comparison between the non-contact based visual and thermal modalities as well as the contact-based physiological measurements. Third, we analyzed whether certain regions in the face were capable of providing discriminant features to differentiate between deceptive individuals and truth tellers.

Our experimental results indicate that the thermal features are difficult to control and hence, provided a promising approach to address the deception detection problem. In particular, the periorbital region was found to have variations in its thermal distribution, which are enough to identify deceptive behavior with detection rates higher than random guessing. As it is easier for human to control their facial expressions and muscle movements, the performance of the automatically-generated visual features was poor and the automated feature extraction method was not able to capture any subtle changes between deceptive and truthful responses, if any. However, the feature selection process provided us with specific action units and expressions, which seemed to be promising in detecting deception. These features will be further analyzed as we collect more data.

The size of the data used to train the classifier played a significant role in improving the performance. On the other hand, the performance was topic-independent, which is promising for the development of a system trained on certain scenarios to be used in multiple applications. The significance of the role of the interviewer's involvement in the mock crime scenario cannot be concluded for the time being. With these promising guidelines and leads, we are in the process of collecting more data using more sophisticated equipment to reliably identify deceptive behavior.

7. ACKNOWLEDGMENTS

This material is based in part upon work supported by National Science Foundation awards #1344257 and #1355633, by grant #48503 from the John Templeton Foundation, and by DARPA-BAA-12-47 DEFT grant #12475008. 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, the John Templeton Foundation, or the Defense Advanced Research Projects Agency.

8. REFERENCES

[1] 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.

[2] 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.

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

[4] M. Garbey, A. Merla, and I. Pavlidis. Estimation of blood flow speed and vessel location from thermal video. In *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004*, volume 1, pages I–356–I–363 Vol.1, June 2004.

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

[6] U. Jain, B. Tan, and Q. Li. Concealed knowledge identification using facial thermal imaging. In *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1677–1680, March 2012.

[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] T. O. Meservy, M. L. Jensen, W. Kruse, J. K. Burgoon, and J. Nunamaker, Jay F. Automatic extraction of deceptive behavioral cues from video. In *Terrorism Informatics*, volume 18 of *Integrated Series In Information Systems*, pages 495–516. Springer US, 2008.

[9] 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.

[10] K. K. Park, H. W. Suk, H. Hwang, and J.-H. 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.

[11] I. Pavlidis and J. Levine. Thermal image analysis for polygraph testing. *IEEE Engineering in Medicine and Biology Magazine*, 21(6):56–64, Nov 2002.

[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] T. Pfister and M. Pietikäinen. Electronic imaging & signal processing automatic identification of facial clues to lies. *SPIE Newsroom*, January 2012.

[14] T. Qin, J. K. Burgoon, J. P. Blair, and J. F. Nunamaker. Modality effects in deception detection and applications in automatic deception-detection. In *Proceedings of the 38th Hawaii International Conference on System Sciences*, 2005.

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

[16] S. N. Sinha, J.-m. Frahm, M. Pollefeys, and Y. Genc. Gpu-based video feature tracking and matching. Technical report, The University of North Carolina at Chapel Hill, 2006.

[17] L. Warmelink, A. Vrij, S. Mann, S. Leal, D. Forrester, and R. P. Fisher. Thermal imaging as a lie detection tool at airports. *Law and Human Behavior*, 35(1):40–48, 2011.

[18] M. Yancheva and F. Rudzicz. Automatic detection of deception in child-produced speech using syntactic complexity features. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 944–953, Sofia, Bulgaria, August 2013. Association for Computational Linguistics.

[19] 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.