Deception Detection Using a Multimodal Approach

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

Verónica Pérez-Rosas
Computer Science and Engineering
University of North Texas
Denton, TX 76207, USA
vrncapr@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

In this paper we address the automatic identification of deceit by using a multimodal approach. We collect deceptive and truthful responses using a multimodal setting where we acquire data using a microphone, a thermal camera, as well as physiological sensors. Among all available modalities, we focus on three modalities namely, language use, physiological response, and thermal sensing. To our knowledge, this is the first work to integrate these specific modalities to detect deceit. Several experiments are carried out in which we first select representative features for each modality, and then we analyze joint models that integrate several modalities. The experimental results show that the combination of features from different modalities significantly improves the detection of deceptive behaviors as compared to the use of one modality at a time. Moreover, the use of non-contact modalities proved to be comparable with and sometimes better than existing contact-based methods. The proposed method increases the efficiency of detecting deceit by avoiding human involvement in an attempt to move towards a completely automated non-invasive deception detection process.

Categories and Subject Descriptors

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

General Terms

Keywords

multimodal, deception, thermal

1. INTRODUCTION

DePaulo et al. [2], define the act of deception as a deliberate attempt to mislead others. As deception occurs on a daily basis through explicit lies, misrepresentations, or omissions, it has attracted the interest of researchers from multiple research fields. Also with a growing number of multimodal communications, the need arises for efficient and enhanced methodologies to detect deceptive behaviors.

Previous attempts on detecting deceit were usually conducted using physiological sensors and trained experts. However, a major drawback for these strategies is that human judgment on different cases is usually biased and achieves poor classification accuracy [1]. Moreover, these approaches require a large amount of time and effort for analysis.

Psychological and theoretical approaches [3] have also been proposed in order to increase human capability of detecting deceptive behavior. Additionally, efforts were exerted to collect behavioral clues that can be indicative of deception. However, these clues can also be results of other states such as the fear of being caught in a lie [2]. For instance, travelers in airports that are the subjects of custom verifications where deception often occurs, can also be stressed due to long flights, fear of flights, and many other reasons.

While it was reported that audible and verbal clues are better deception indicators compared to visual clues [1], recent studies argued that a sudden increase in the blood flow in the periorbital area of the face, which results from spontaneous lies can be detected using thermal imaging. This is especially true when a person tries to find an answer to unexpected questions and unanticipated events [24].

This paper addresses the problem of automatic deception detection using a multimodal approach. The paper makes two important contributions. First, we create a new dataset with the participation of 30 subjects. The subjects were asked to discuss two different topics – ("Abortion" and "Best Friend," described in detail below) – in both truthful and deceptive manners, while they were recorded using a microphone, a thermal camera, and several physiological sensors. Second, in order to automate and improve the detection of deceptive behaviors, avoid human efforts and the limitations associated with individual methods, and increase the efficiency of the decision making process, we build a multimodal system that integrates features extracted from three different modalities. Features are extracted from the linguistic re-

ICMI '14, November 12-16 2014, Istanbul, Turkey

sponses of the subjects, thermal recordings of the face, and physiological measures obtained from several sensors. We report the individual performance of each modality as well as the performance of various modality combinations. To our knowledge, this is the first attempt to detect deceptive behaviors by integrating these modalities.

This paper is organized as follows. Section 2 surveys some of the related work. Section 3 describes the data collection process and our experimental design. Section 4 illustrates the feature extraction process utilized for each modality. Section 5 discusses our experimental results. Finally, concluding remarks and future work are provided in Section 6.

2. RELATED WORK

While most of the earlier research on the deception detection task focused on analyzing physiological responses such as skin conductance, blood pressure pulse, and respiration rate, there have been important efforts on identifying deceptive behavior in scenarios where contact-based measurements are not available. For instance, researchers have studied verbal behaviors exhibited by people while deceiving [5, 23]. Speaking rate, energy, pitch, range as well as the identification of salient topics have been found useful to distinguish between deceptive and non-deceptive speech [4]. Linguistic clues such as self references or positive and negative words have been used to profile true tellers from liars [8]. Other work has focused on analyzing the number of words, sentences, self references, affect, spatial and temporal information associated with deceptive content [16].

Several efforts have also been presented with the use of non-invasive measurements such as thermal imaging. Researchers investigated the relation between measurements extracted from the subjects' faces and states of deception. Pavlidis et al. [10] developed a high definition thermal imaging method to detect deceit from the facial area claiming to reach a performance close to that of the polygraph tests. Warmelink et al. [25] used thermal imaging as a lie detector in airports. They extracted the maximum, minimum, and average temperatures from the thermal images to detect deceit. With 51 participants, their system was able to detect lying participants with accuracy above 60%. However, interviewers outperformed the system with above 70% accuracy. Pavlidis and Levine [11, 12] applied thermodynamic modeling on thermal images to transform the raw thermal data from the periorbital area to blood flow rates to detect deception. Merla and Romani [7] investigated the correlation between different emotional conditions, such as stress, fear, and excitement, and facial thermal signatures using thermal imaging.

In order to specify which thermal areas have higher correlation to deceptive behaviors, facial regions of interest were specified. Rajoub and Zwiggelaar [17] analyzed thermal faces by creating two regions of interest by manually identifying the corners of the eyes and tracking these regions over the recorded video frames. Their deception detection system performed well on within-subject data but not on across-subject scenarios. Pollina et al. [15] extracted the minimum and maximum temperatures from video frames recorded of subjects in states of deceptions and truthfulness. They focused on the eyes region of the face and reported a significant change in the surface skin temperature between the two states. Jain et al. [6] employed a thermal camera with face detection, tracking, and landmark detections systems to detect and track landmarks on the regions of interest in the face area. The method calculated the average temperature of the 10% hottest pixels of a window that includes both tear ducts. Pavlidis et al. [13] observed distinct non-overlapping facial thermal patterns resulting from different activities and detected an increase in the blood flow around the eyes when subjects act deceptively. Tsiamyrtzis et al. [21] used tandem tracking and noise suppression methods to extract thermal features from the periorbital area without restriction on the face movements of the subjects in order to improve deception detection rates. Sumriddetchkajorn and Somboonkaew [20] introduced an infrared system to detect deception by converting variation in the thermal periorbital area to relative blood flow velocity and by deducing the respiration rate from the thermal nostril areas. Park et al. [9] employed a functional discriminant analysis to separate between deceptive and non-deceptive behaviors using the average maximum temperature of the periorbital region of the subjects' faces during their answers to specific questions. To be able to only extract measurements from the periorbital area, subjects were not allowed to move or tilt their faces.

3. EXPERIMENTAL SETUP

We build upon previous work on deception detection, in particular we use protocols similar to those previously used in psychology and physiology. The hypothesis, verified in these earlier works and in ours, is that as a person acts/speaks deceptively, there will be subtle changes in his or her physiological and behavioral response. Additionally, during our experiments, we avoided any bias towards certain topics by using two different topics. Furthermore, we did not interfere with the responses of the participants. The sole purpose of the interviewer was to hand the subjects the topic they had to discuss in a deceptive or truthful manner. Our scenarios are more difficult for the detection of deception as compared to other question-based scenarios.

3.1 Equipment and data acquisition

We acquired measurements using a thermal camera FLIR Thermovision A40 with a resolution of 340x240 and a frame rate of 60 frames per second, as well as four biosensors including: blood volume pulse (BVP sensor), skin conductance (SC sensor), skin temperature (T sensor), and abdominal respiration (BR sensor). Each session was also video recorded (with audio) using a Logitech Web Cam.

During each recording session we obtained thermal measurements of participants' faces using the software provided with the thermal camera, namely, the Flir ThermaCam Researcher. Also, four sensors were attached to the non-dominant hand of the participants. Two skin conductance electrodes were placed on the second and third fingers whereas the skin temperature and blood volume blood volume sensors were placed at the thumb and index fingers respectively. The respiration sensor was placed comfortably around the thoracic region. The output of each sensor was obtained from a multimodal encoder connected to the main computer using an USB interface device. We recorded the combined output using the Biograph Infinity Physiology suite[1], which allowed us to visualize and control the data acquisition process.

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

3.2 Participants

The human subjects consisted of 30 graduate and undergraduate students. The sample consisted of 5 female and 25 male participants, all expressing themselves in English, from several ethnic backgrounds (Asian, African-American, Caucasian, and Hispanic), with ages ranging between 22 and 38 years.

3.3 Collecting truthful and deceptive responses

The participation in the study consisted of sitting at the recording station while connected to the physiological sensors. Thermal recordings and physiological measurements were obtained during each part of the experiment.

First, we described the experimental system and procedure to be followed by the participants and instructed them to respond either truthfully or deceptively, depending on the topic being run. Then, we attached the sensors to the non-dominant hand and setup the thermal camera. Participants were asked to avoid any excessive movements with their head or hands in order to keep to a minimum measurement uncertainties. The reason of movement restrictions was to obtain high quality data from the cameras and reduce the amount of possible disturbances with the physiological sensors. This is particularly important for the temperature and the skin conductance measurements, since they are obtained using sensors that need to be in permanent contact with participant's skin. For this study we have decided to use wired sensors due to their cost, reliability and robustness but wireless sensors can also be used.

In order to elicit deceptive and truthful responses, we designed the following two topics, for which participants were asked to speak freely for about 2-3 minutes.

Abortion (AB) In this experiment participants were asked to provide a truthful and a deceptive opinion about their feelings regarding abortion. Participants were asked to imagine a topic where they took part in a debate on abortion. The experiment session consisted of two independent recordings for each case, when the participant was either telling the truth or lying. In the first part of the experiment, the participant had to defend his or her point of view regarding abortion, while in the second part the participant was asked to lie about what she or he really thinks about abortion.

Best Friend (BF) In this experiment participants were asked to provide both a true description of their best friend, as well a deceptive description about a person they cannot stand. Thus, in both cases, a person was described as the participant's best friend, but in only one of the cases the description was truthful. The experiment session consisted of two independent recordings for each case, when a given participant was either telling the truth or lying.

4. METHODOLOGY

After the data collection, we obtained a total of 30 truthful and 30 deceptive observations for each topic to form a total of 120 responses, including their corresponding audio/visual, thermal and physiological sensors recordings. The deception detection process involves two steps: discriminant feature extraction and classification.

4.1 Multimodal features

The raw data was processed to obtain a set of features to represent each modality.

4.1.1 Physiological features

We obtained physiological features by processing the raw signal from each sensor. We used the Biograph Infiniti Physiology suite to obtain physiological assessments for temperature, heart rate, blood volume pulse, skin conductance, and respiration rate. More specifically, using the raw output from each sensor, we calculated the mean, standard deviation, and power mean. We also obtained the same descriptors for each two-seconds epoch length. In addition to this, we obtained features derived from inter beat intervals (IBI) measurements such as minimum and maximum amplitude and their intervals. The final set consists of 60 physiological features.

4.1.2 Linguistic features

We obtained linguistic features which represent the participant's use of language when they are either telling the truth or lying. To represent the linguistic component we obtained manual transcriptions of the recorded statements. We then extracted two different sets of features.

First, we used a bag-of-words representation of the transcripts to derive unigram counts, which are then used as linguistic features. We started by building a vocabulary consisting of all the words occurring in the transcripts. We then removed those words that have a frequency below 10 (value determined empirically on a small development set). The remaining words represent the unigram features, which are then associated with a value corresponding to the frequency of the unigram inside each video transcription. Second, in order to obtain features that represent psychological processes occurring while people are providing truthful or deceptive statements, we opted for using the Linguistic Inquiry and Word Count (LIWC) lexicon, which is a resource developed for psycholinguistic analysis [14] and has been widely used to aid deceit identification in written sources [8, 18]. In particular, we used the 2001 version of the dictionary, which contains about 70 word classes relevant to psychological processes (e.g., emotion, cognition), which in turn are grouped into four broad categories[2] namely: linguistic processes, psychological processes, relativity, and personal concerns. We extracted frequency counts of words occurring in the transcripts belonging to the different lexicon word classes. We performed separate evaluations using each of the four broad LIWC categories, as well as using all the categories together. The classification results for truthful and deceptive responses, obtained with a decision tree classifier implemented as described in Section 5 are shown in Table 1.

Motivated by the results shown in Table 1, which indicate that the combination of the Linguistic Processes and unigrams attained the best overall accuracy for the best friend and abortion topics, we decided to keep only these features for the remaining analysis. Note that we also attempted to use unigrams and higher order n-grams (bigrams and trigrams) in combinations with the different LIWC categories but evaluations did not show any improvements over the use of unigrams. The final feature set for both topics consisted of 214 linguistic features.

[2] http://www.liwc.net/descriptiontable1.php

Features	Topic AB	BF	AB+BF
LIWC:Linguistic Processes	55.00%	66.66%	**63.33%**
LIWC:Psychological Processes	40.00 %	58.33%	57.50%
LIWC:Personal Concerns	58.33%	41.66%	49.16%
LIWC:Relativity	48.33%	65.00 %	60.00%
LIWC:All	53.33%	65.00%	60.83%
Unigrams	56.66%	65.00%	59.16%
Ling.Proc.+Uni	**56.66%**	**70.00%**	62.50%

Table 1: Deceptive and truthful statements classification using LIWC categories and unigrams. Results are obtained using leave-one-out cross-validation

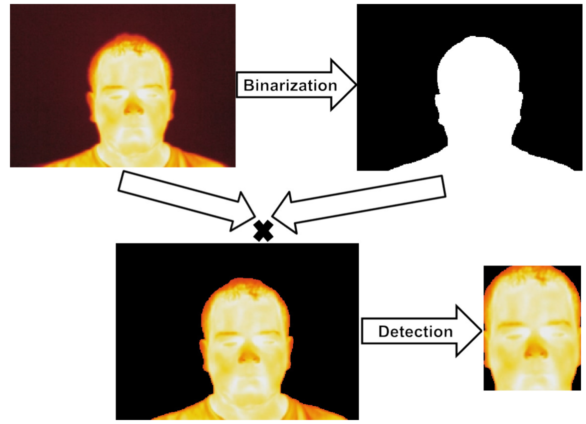

Figure 1: An overview of the thermal face detection process.

4.1.3 Thermal features

We decided to use the whole facial area to extract meaningful features that can discriminate between deceptive and non-deceptive behaviors. In order to do this, we first detected the facial areas of the subjects, followed by creation of heat maps of the detected faces. We tried the Viola-Jones face detection algorithm [22] to detect the faces directly from the thermal images but were not satisfied with the results due to a large number of undetected faces. Hence, we automatically isolated the subjects' faces from the background using image binarization given that the heat emitted from the background was always assumed to be of lower intensity compared to the skin temperature. Higher values of the frames' pixels indicate higher temperatures, which corresponds to lighter colors in the frames. The binarization process converts the pixels of each frame into black and white by thresholding the values of the pixels into either (0) for black or (1) for white using Otsu's method [19]. The thresholding process relies on the pixels intensities to minimize the intra-class variance of the white and black pixels.

The binarization process results in a holistic shape of white pixels of the upper body including face, neck, and shoulder areas. The binarized image is then multiplied by the original image to eliminate the background, i.e., the upper body is retained when multiplied by 1 while the background pixels are blackened. Using relative measurements, we were able to locate the neck area assuming that it always has the least width of non-zero pixels. Hence, we were able to detect the facial area while completely eliminating the background including the corners found in the cropped face images. All the eliminated parts have black (zero-valued) pixels. The face detection process is shown in Figure 1.

Following the face detection process, we extracted the maximum pixel value corresponding to the highest face temperature, the minimum pixel value corresponding to the lowest face temperature, the average of the pixels values of the face, the maximum/minimum pixels range which measures the difference between the maximum and minimum temperatures, and histogram of 120 bins over the pixels values in the facial area to form a total of 124 thermal features for each image. We uniformly sampled 200 frames from each video response. The extracted features were averaged over this number of frames for each subject. These measurements along with the histogram create a complete heat map that defines the heat distribution on the face. We aim at detecting variations in this heat map when any of the subjects acts deceptively.

Different subjects can have varying skin temperatures in normal conditions. These variations can manipulate the thermal maps and negatively affect the deception detection performance. To treat different subjects fairly, we used the first 50 seconds of each video recording as the thermal baseline for each subject. During this time, each subject was sitting on a chair without performing any activity or verbal responses. The same set of 124 features were extracted and averaged over this time frame. Hence, thermal correction is created by dividing the extracted features from the subjects' responses by their thermal baseline. This normalization process indicates whether there is a shift in the thermal map towards higher temperatures, i.e. higher pixel values, when the subjects respond deceptively regardless of the inter-personal temperature variations.

4.2 Deception classification

The feature extraction process generates a feature vector for each modality for each subject. Feature-level fusion is then employed by concatenating the features extracted from the three modalities for each subject. The concatenated feature vectors are used to train a decision tree classifier as recommended in [16] using the statistical toolbox in Matlab R2013a in order to detect deceptive instances. The classification process employs a leave-one-out cross validation scheme and the average overall and per class accuracies are reported. To clarify whether the integration of features from different modalities in fact improves the performance, we conducted our experiments using features from individual modalities as well as in combination. Additionally, we also report the performance of an across-topic learning scheme, where the classifier is trained with features extracted from one topic while tested on the other. The across-topic analysis is conducted to explore the capability of detecting deception when the training and test data is drawn from different topics.

5. EXPERIMENTAL RESULTS

Given our set of a total of 120 instances for each of the three modalities (two per subject per topic), we started by

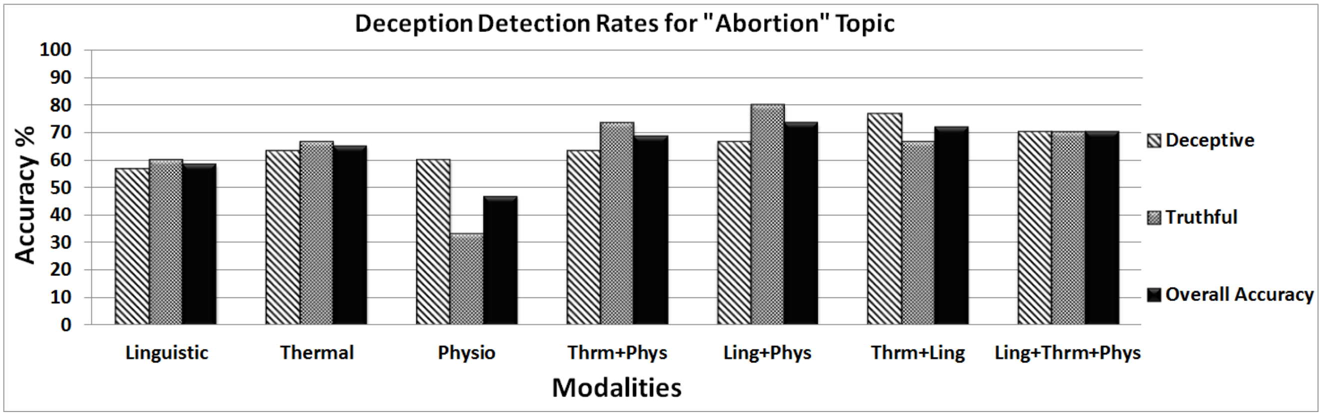

Figure 2: Deception, truthfulness, and overall accuracy percentages for individual and integrated modalities using features extracted from the abortion topic.

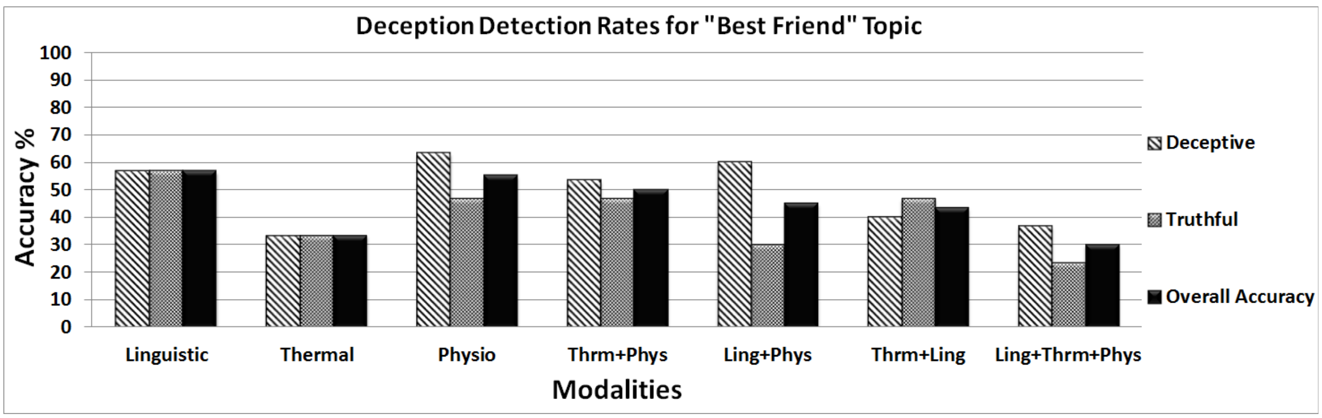

Figure 3: Deception, truthfulness, and overall accuracy % for individual and integrated modalities using features extracted from the "Best Friend" topic.

evaluating the performance of the features extracted from each topic, followed by the the evaluation of the features of both topics combined. We compared the performance of individual modalities to their combinations. This is followed by evaluating the performance of the across-topic learning process to investigate whether a trained model can identify deceit in different domains.

5.1 Individual and integrated modalities

Figure 2 shows deception and truthfulness detection rates in addition to the overall accuracy using different modalities for the abortion topic. The figure indicates that overall the combination of different modalities improves the performance compared to individual ones. The thermal modality achieved the highest performance of the individual modalities with above 60% for each class as well as for the overall accuracy. The linguistic features exhibit close performance while the physiological features achieved the lowest performance. While there is an improvement in the deception detection using the physiological features, the performance of the truthful class is deteriorated significantly.

A clear improvement in performance can be seen for several modality combinations. In particular the combination of all three modalities achieves a consistent 70% accuracy among all classes. The combination of linguistic and thermal features exceeds 70% accuracy for the deceptive class as well as for the overall accuracy. The combination of linguistic and physiological features leads also to a similar overall performance with a drop and an improvement in the performances of the deceptive and the truthful classes, respectively. The improvement in the overall accuracy using this combination of modalities results in an error rate reduction of 53% as compared to the use of one modality at a time.

The performance of the features extracted from the best friend topic is significantly lower than the first topic using different modalities as can be seen in Figure 3. In the abortion topic, subjects had a conclusive opinion whether to support it or not. Therefore, their responses were clearly either deceptive or truthful. However, the subject's responses on the best friend topic were mixed with positive and negative statements. For example, while truthfully describing their best friend, subjects sometimes mentioned things they don't like about their best friend such as "He is kind of lazy." Our analysis indicates that inclusion of some negative statements and/or memories in the truthful responses affect the quality of the extracted features, especially the thermal and physiological features. This is supported by the improved performance of the deceptive class compared to the truthful one for most modalities.

Additionally, Figure 3 shows that different combinations

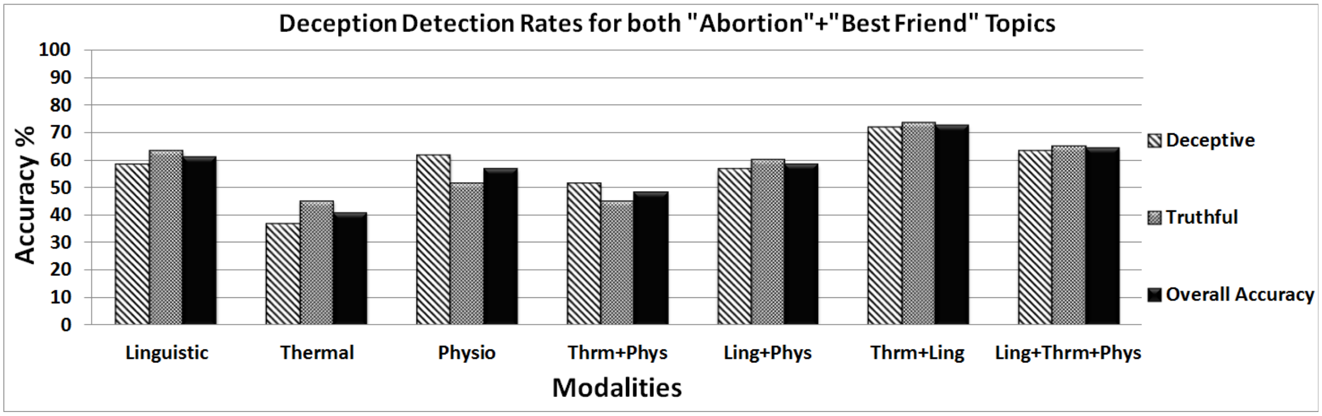

Figure 4: Deception, truthfulness, and overall accuracy percentages for individual and integrated modalities using features extracted from both the abortion and best friend topics.

of modalities achieve an accuracy that is equal to or lower than that of binary random guessing. The performance of individual modalities for this topic is sometimes better than their combinations as seen with the linguistic features. The deception detection rate using physiological and linguistic features is able to rise above 50%, but it does not improve over the use of linguistic features by themselves.

Figure 4 illustrates the performance of the features extracted from both topics together for all modalities. The linguistic and physiological features exhibit improved performance compared to their per-topic performance. This indicates that as the number of instances increases, even with data from different topics, the trained model shows considerable increase in its capability of discriminating between deceptive and truthful instances. However, although the thermal features here demonstrate improved performance compared to the best friend topic, their overall accuracy does not exceed that of random guessing and does not reach that of the abortion topic. Given the obvious difference in the performance of thermal features for both topics, the increase in the number of instances is not able to reconcile this large difference. On the other hand , the difference in performance using the linguistic and physiological features for both topics is not significant and hence, integrating features from both topics further improve the accuracy.

The use of multimodal features further enhances the classification accuracy. In particular, the integration of all three modalities together in addition to the integration of the thermal and linguistic features obtain higher accuracy in comparison to all other combinations as well as all individual modalities. Although the best performing single modalities are linguistic and physiological, the combination of thermal and linguistic modalities exceeds 70% for both classes and for the overall accuracy. The decision tree model was able to select its nodes and classification rules using a combination of thermal and linguistic features, which was more discriminative between the deceptive and truthful classes as compared to features obtained from the individual modalities. Furthermore, the improvement in accuracy obtained by using the linguistic-thermal combination is statistically significant (t-test $p< 0.05$) over the use of single modalities, i.e linguistic and thermal. The overall accuracy of this combination is additionally better than the accuracy of any individual or combined modalities for the individual topics. These results suggest two important remarks. First, enlarging the data size with features from different deceptive topics is useful. Second, integrating features from multiple modalities can significantly improve the performance. In fact, the improvement obtained from the thermal and linguistic features can move us towards automated non-invasive deception detection methods.

5.2 Across-topic learning

Figures 5 and 6 illustrate the deceptive and truthful detection rates and the overall accuracy for the across-topic learning process using individual and combined modalities. In this learning scheme, the classifier is trained using features from one topic and then tested on the other topic. For example, the classifier is trained using the best friend features and tested using abortion features in Figure 5 and viceversa for Figure 6. In both cases, it can be noticed that the linguistic modality creates a large imbalance between the detection rate of the deception and truthfulness classes. While one class has a significantly improved performance, the other class suffers a deteriorated performance, which indicates the failure of the learning process. Moreover, the same trend is observed whenever the linguistic features are added to features extracted from other modalities. This trend is not observed with the thermal and physiological individual modalities as well as the combined modalities that exclude the linguistic features.

The disposition of the results can be explained with the dependency of the linguistic features on the corresponding topic. For instance, the unigrams extracted from abortion depend on words that are mostly related to this particular topic such as “illegal,” “health,” “baby,” etc. These features are not related to the other topic and, hence, does not support the learning process.

Surprisingly, the across-topic learning scheme improves the detection rates for the thermal and physiological modalities except when the abortion topic is tested in Figure 5 with the combination of the thermal and physiological features. This can be attributed, as discussed earlier, to the abortion features having higher quality compared to the features obtained for the best friend topic. The deception detection rate is improved and the truthful class accuracy is deteriorated using thermal features in Figure 5, however, the overall accuracy remains the same at 65%. On the other hand,

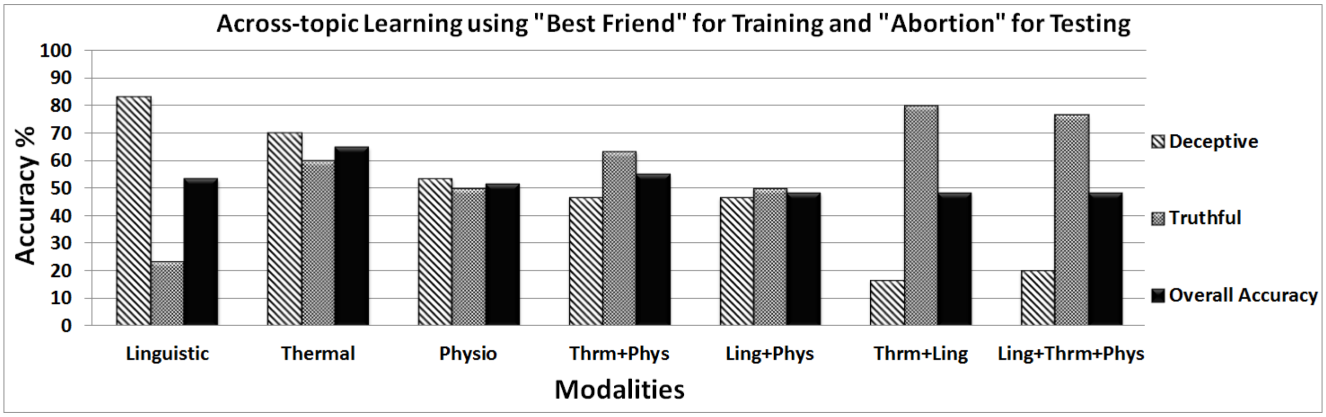

Figure 5: Deception, truthfulness, and overall accuracy percentages for individual and integrated modalities using across-topic learning. best friend features are used for training and abortion features are used for testing.

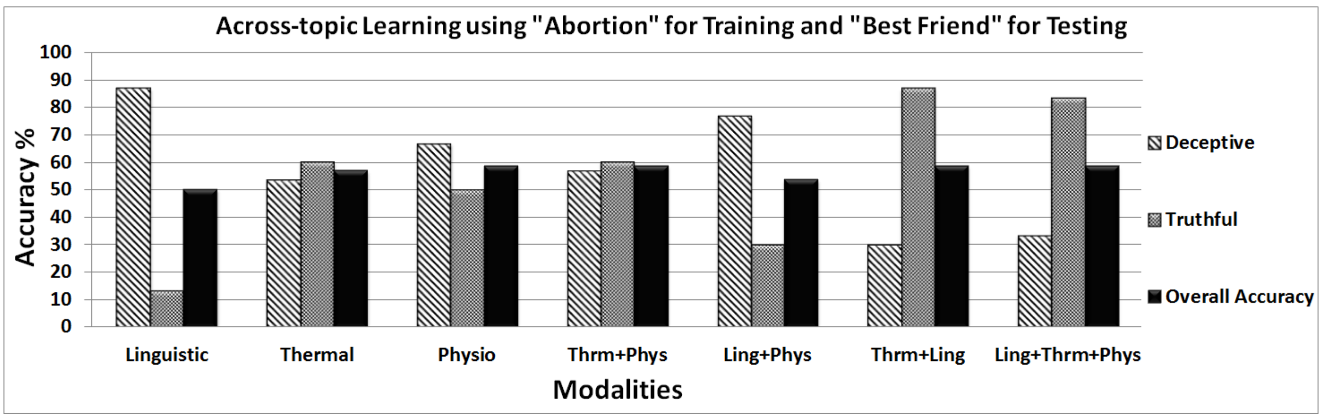

Figure 6: Deception, truthfulness, and overall accuracy percentages for individual and integrated modalities using across-topic learning. Abortion features are used for training and best friend features are used for testing.

training the classifier with the abortion thermal features in Figure 6 significantly improves the performance of the best friend topic to 70.1% overall accuracy.

6. CONCLUSIONS AND FUTURE WORK

This paper introduced a novel multimodal system to detect deceptive behaviors by integrating features from linguistic, thermal, and physiological modalities. The paper also contributed a new dataset consisting of deceptive and truthful responses. The proposed method is a step towards a completely automated non-invasive deception detection process.

Experimental results suggested that features extracted from linguistic and thermal modalities can potentially be good indicators of deceptive behaviors. Moreover, creating a multimodal classifier by integrating features from different modalities proved to be superior compared to learning from individual modalities. In particular, the integration of thermal and linguistic features resulted in a significant performance gain.

Additionally, our experiments showed that the quality of the extracted features is topic-related. In some topics, emotions and memories could negatively affect the quality of the features and reduce their capability of discriminating between deceptive and truthful behaviors. A disadvantage of using linguistic features existed when the system attempted to detect deceit on a completely new topic that was not used to train the classifier. However, this is not a problem for the thermal and physiological features, which can improve the performance of detecting deceit in a new topic if high quality features were provided for the training process. These trends indicate that a larger dataset needs to be collected using multiple topics and events in order to further improve the performance.

In the future, we are planning to collect a larger dataset with a variety of new topics and scenarios. Additionally, we will investigate the integration of additional modalities such as visual features. We also plan to extract more sophisticated features from different modalities. For instance, we will divide the thermal faces into regions of interest to analyze which parts of the face provide the highest discriminatory features. For the linguistic modality, we plan to explore the use of syntactic stylometry in order to capture deeper syntax patterns associated with deceptive statements. Furthermore, we are investigating the use of automatically extracted action units, which will be used to identify facial behaviors associated with deception.

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] C. F. J. Bond and B. M. DePaulo. Accuracy of deception judgments. *Personality and Social Psychology Review*, 10:214–234, 2006.

[2] B. M. Depaulo, B. E. Malone, J. J. Lindsay, L. Muhlenbruck, K. Charlton, H. Cooper, B. M. Depaulo, B. E. Malone, D. O. Psychology, J. J. Lindsay, L. Muhlenbruck, and K. Charlton. Cues to deception. *Psychological Bulletin*, pages 74–118, 2003.

[3] P. A. Granhag and M. Hartwig. A new theoretical perspective on deception detection: On the psychology of instrumental mind-reading. *Psychology, Crime & Law*, 14(3):189–200, 2008.

[4] J. Hirschberg, S. Benus, J. Brenier, F. Enos, S. Friedman, S. Gilman, C. Gir, G. Graciarena, A. Kathol, and L. Michaelis. Distinguishing deceptive from non-deceptive speech. In *In Proceedings of Interspeech 2005 - Eurospeech*, pages 1833–1836, 2005.

[5] D. Howard and C. Kirchhübel. Acoustic correlates of deceptive speech: an exploratory study. In *Proceedings of the 9th international conference on Engineering psychology and cognitive ergonomics*, EPCE'11, pages 28–37, Berlin, Heidelberg, 2011. Springer-Verlag.

[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] A. Merla and G. Romani. Thermal signatures of emotional arousal: A functional infrared imaging study. In *29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS.*, pages 247–249, Aug 2007.

[8] M. Newman, J. Pennebaker, D. Berry, and J. Richards. Lying words: Predicting deception from linguistic styles. *Personality and Social Psychology Bulletin*, 29, 2003.

[9] 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.

[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] I. Pavlidis and J. Levine. Monitoring of periorbital blood flow rate through thermal image analysis and its application to polygraph testing. In *Proceedings of the 23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society.*, volume 3, pages 2826–2829, 2001.

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

[13] I. Pavlidis, J. Levine, and P. Baukol. Thermal imaging for anxiety detection. In *IEEE Workshop on Computer Vision Beyond the Visible Spectrum: Methods and Applications.*, pages 104–109, 2000.

[14] J. Pennebaker and M. Francis. Linguistic inquiry and word count: LIWC, 1999. Erlbaum Publishers.

[15] D. Pollina, A. Dollins, S. Senter, T. Brown, I. Pavlidis, J. Levine, and A. Ryan. Facial skin surface temperature changes during a "concealed information" test. *Annals of Biomedical Engineering*, 34(7):1182–1189, 2006.

[16] 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.

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

[18] V. Rubin. On deception and deception detection: Content analysis of computer-mediated stated beliefs. *Proceedings of the American Society for Information Science and Technology*, 47(1):1–10, 2010.

[19] M. Sezgin and B. Sankur. Survey over image thresholding techniques and quantitative performance evaluation. *Journal of Electronic Imaging*, 13(1):146–168, 2004.

[20] S. Sumriddetchkajorn and A. Somboonkaew. Thermal analyzer enables improved lie detection in criminal-suspect interrogations. In *SPIE Newsroom: Defense & Security*, 2011.

[21] P. Tsiamyrtzis, J. Dowdall, D. Shastri, I. Pavlidis, M. Frank, and P. Eckman. Lie detection - recovery of the periorbital signal through tandem tracking and noise suppression in thermal facial video. In *SPIE Conference on Sensors and Command Control Communications and Intelligence Technologies for Homeland Security and Homeland Defense IV*, pages 555–âĂŞ566, 2005.

[22] P. Viola and M. Jones. Robust real-time face detection. *International Journal of Computer Vision*, 57(2):137–154, 2004.

[23] A. Vrij, P. Granhag, and S. Porter. Pitfalls and Opportunities in Nonverbal and Verbal Lie Detection. *Psychological Science in the Public Interest*, 11(3):89–121, Dec. 2010.

[24] A. Vrij, S. Leal, P. Granhag, S. Mann, R. Fisher, J. Hillman, and K. Sperry. Outsmarting the liars: The benefit of asking unanticipated questions. *Law and Human Behavior*, 33(2):159–166, 2009.

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