<u>How do you *really* feel?</u>
<u>Benchmarking automatic techniques for micro-expression recognition</u>

Abstract

A main goal in the field of affective computing is to train machines to sense and recognize emotions from image data using techniques from Computer Vision. In Psychology, a distinction in studying facial human affect (or emotional expression) is made between macro-expressions and micro-expressions. In contrast to macro-expressions, micro-expressions are short and involuntary expressions of emotions which cannot be concealed, revealing reactionary emotions. The application of Computer Vision methods to recognizing macro-expressions is a well-defined space in which features to extract and machine learning algorithms to apply are still evolving, but are already being applied generally. In contrast, the study of micro-expressions using Computer Vision methods is an emerging field for which methodologies and datasets are newly being contributed addressing the short and involuntary characteristics of this type of facial movement. Challenges in the automatic recognition of micro-expressions include the spontaneous nature which needs to be present in data collected, as well as the short duration of expression. Researchers address the short duration of expression by adding temporal resolution or granularity by using higher frame rates for data collection or using temporally interpolated frames. In parallel, more sophisticated machine learning algorithms requiring fewer training examples are proposed. The two primary research groups addressing these challenges have proposed datasets and methodologies for the task, but have failed to show that micro-expression recognition needs methods distinct from macro-expression recognition. Our primary research objectives are to validate the claim that micro-expression and macro-expression automatic recognition need distinct methods. I aim to test the following hypothesis: classification models using only texture-based image features will have significantly lower recognition rates when applied to micro-expression datasets than methods providing temporal resolution. This hypothesis will be tested by implementing micro- and macro-expression methodologies available in the literature, and applying them to currently available datasets.

1. Introduction

1.1 *Psychology motivation*

In psychology, an area of interest is being able to detect one of the 6 basic human emotions from facial expressions: anger, contempt, disgust, fear, enjoyment, sadness and surprise. [1] For studies of social phenomena involving emotions, researchers record interactions among participants and manually label or annotate each frame according to Ekman and Friesen's Facial Action Coding System (FACS). [2] Observers first annotate frames with Action Units (AU) corresponding to contracted muscles and from the labelled AUs from which emotion-based inferences can be made. [3] For example, happiness (or enjoyment) is represented by AUs 6 (cheek raised) and 12 (lip corner pulled). [4] After each frame of a video (observing some phenomenon of interest) is annotated, the patterns of expressed emotions can be analyzed for meaning. For example, in [5], videos of guided conversations between spouses are coded using FACS (as well as using voice affect coding and physiological data) and aggregated as counts of positive and negative emotional expression. From these encodings, intermediate factors are constructed and a model is built capable of predicting divorce.

1.2 *Macro vs. Micro-expressions*

"Discovered" by Ekman in [6], micro-expressions are distinguished from macro-expressions (or commonly just *facial expressions* in the literature) by two characteristics: spontaneity and duration. In contrast to macro-, micro-expressions are thought to be the result of automatic, uncontrollable biological response. [1] Both types of facial expressions are thought to be universally displayed and understood (across cultures), and both can be identified by experts using FACS frame-by-frame coding. The main differences noted are the shorter duration of micro-expressions as well as fragmented or more subtle manifestations of AUs. [1] From a psychology standpoint micro-expressions can reveal true emotions or deception. In contrast, macro-expressions are learned and can be posed. Refer to [1] for a more complete review of the psychological research and implications of micro-expressions.

1.3 *Automation of facial expression recognition*

The task of recognizing and annotating frames with intermediate AUs or emotions is well-suited for automation through Computer Vision and machine learning. Decades of work in psychology has provided a visual dictionary of AUs (used for human training in this task) and an increasing number of datasets have become available suitable for this task. Automatic techniques can be used more often and at a lesser cost than experts, especially given the expense and time required to complete the training necessary. Also, the proliferation of video data in both public domain (YouTube) as well as private domain (business laws regulating retention of video-conferencing, investment in closed-circuit security) provides many application opportunities.

Human experts are also unreliable at this task. When applying FACS to video, they label AUs as well as label the frames with the onset and offset of an expression. Example of a sequence of temporal labels include: neutral face, onset frames, peak frame, offset frames and neutral face. When using FACS, a single rater could be wrong as often as 20-30% [7] of the time, an issue compounded by adding a second rater. Inter-observer agreement traditionally measured for multiple raters performing this task include: occurrence/non-occurrence of AUs, temporal precision (similarity in labelling of onset/peak/offset frames), intensity, and aggregates (agreement on emotional event inferred from AUs). [3] Cohen's kappa (a measure modified to account for chance of random agreement) of occurrence/non-occurrence of specific AUs ranges from 0.44 to 0.73. [3] These issues are thought to be even more dramatic when identifying micro-expressions as their intensity is less pronounced.

1.4 *Computer Vision*

One goal of Computer Vision is to train a machine to acquire specific visual perception capabilities with inputs such as sensor data, images or video (an ordered sequence of images). The learned visual perception can be a spatial 3D model of a scene or image understanding (resulting in an ability to tag images with labels according to a specific task.) The traditional approach to image understanding in Computer Vision has four main stages: data collection, image processing, feature extraction, and machine learning/classification. The quality of the learned task is affected by methods at each stage of this process.

Data Collection → Image → Image Processing → Feature Extraction → Machine Learning → Prediction

Figure 1 Computer Vision Sequence

During the data collection phase, videos are acquired pertinent to the task at hand and labelled by experts according to the phenomenon under study. For micro-expression recognition, the video collected must represent spontaneous (vs. acted) expression so that the facial expressions will be short in duration and fragmented. For macro-expression recognition, the participants under study can simply act out facial expressions. Experts from psychology then view each frame labelling AUs, onset/peak/offset frames and emotion inference. During the image processing phase images are prepared in such a way as to limit the amount of noise present that is unrelated to the task performed. In facial expression research, images are typically converted to grayscale, faces are then located according to robust, existing techniques, and images are cropped according to a standard size. More sophisticated processing techniques can be applied such as interpolating of frames or normalization of facial features according to an average face. During the feature extraction phase, features are extracted to capture the size, shape, intensity or texture features of the image (a 2-D matrix of intensity values) such that each feature has one value per image. In facial expression recognition, texture features are commonly used as they represent elements of human visual perception (such as corners, lines and edges). Features for each image are then supplied to a machine learning algorithm, which learns which labels (in our case, emotions) to apply to which images according to their feature representation.

1.5 *Overview of the proposed research*

The field of facial expression recognition has been studied since 1991 [8], with many techniques proposed for each stage of this process [9]. Datasets available for the recognition of these tasks are comprised of acted expressions, meaning only macro-expressions can be identified from them. [10, 11] Researchers specifically identify micro-expression limitations in their work, motivating the work of two independent research groups emerging in 2011 with micro-expression focus. [12, 13] These groups simultaneously proposed datasets and methods, claiming to address the primary challenge presented by micro-expressions of short duration of emotional expression. From a machine learning perspective, the short duration of a micro-expression translates to fewer training frames representing that expression. This challenge is compounded by the multiple possible target classes models need to identify (6 basic emotions as well as the neutral face). This problem is one of temporal resolution: having enough granularity in time to observe expressions for long enough to provide sufficient training examples for a model to distinguish emotions. The proposed approaches in [12, 13] deal with the temporal resolution problems using two different techniques (details are in section 2), but never motivate the need for their sophisticated methods by benchmarking mature macro-expression techniques against their new micro-expressions datasets. Furthermore, since their micro-expression techniques were simultaneously proposed, no comparisons exist between their two distinct contributions.

We propose to validate the claim that distinct, temporally sensitive methods are required for the automatic recognition of micro-expressions by comparing temporally-aware methods and their counterparts on the most recent and complete micro-expression dataset. The main contribution of this work will be in validating the need for temporally-aware micro-expression methods. Secondary benefits will include a comparison of current micro-expression techniques as a performance benchmark for the field. The methods will be compared using paired t-tests of significance on the recognition rate performance of the algorithms for each emotion present in the dataset.

The rest of the proposal is organized as follows: section 2 presents a common macro-expression recognition methodology as well as two competing micro-expression methodologies, section 3 details the research design and methodology, section 4 summarizes and concludes.

2. Related Work

In this section, we briefly present the approaches of the micro-expression recognition foundational papers from the University of Oulu [12] and the Chinese Academy of Science [13]. Both papers attempt to identify micro-expressions from videos, but apply different techniques to different datasets, but used different techniques to provide more temporal resolution or granularity to provide more training examples to the classifiers chosen.

2.1 *Data sets*

Newer datasets such as the Spontaneous Micro-expression Corpus (SMIC) and Chinese Academy of Sciences Micro-expression database (CASME2) use higher frame rates (to provide more samples of micro-expressions). Each dataset has different target labels with varying levels of granularity; while SMIC only has positive vs. negative labels, CASME2 offers a larger range of emotion labels including happiness, disgust, surprise, repression and an uncertain ``other'' category. CASME2 also provides AU codings, representing distinct active muscle contractions which human annotators traditionally use as an intermediate step to determining emotions (according to Ekman's FACS methodology [14]). In Table 1, we outline the datasets used in the papers, including older sets that were used for macro-expression recognition ("acted"). In the case of datasets with extensions, we show them in the same line but only record the details for the extended dataset. It's important to note that samples (n) represent counts of sequences of images.

Name	Proposed	Applied	N	Frame Rate	Labels
YorkDDT (acted)	2009 [10]	[12]	18	25 fps	Truth and deception Emotion and not-emotion
CK, CK+ (acted)	2000 [15], 2010 [11]	N/A, [13]	327	30 fps	AUs + happiness, disgust, surprise, sadness, fear, anger and contempt
SMIC	2011 [12]	[12]	77	100 fps	Positive vs. negative
CASME, CASME2	2011 [13], 2014 [16]	[13] / Us	247	200 fps	AUs + happiness, disgust, surprise, repression and "others"

Table 1 Overview of datasets in related work

2.2 *Image Processing*

Both methodologies require basic image pre-processing in order to standardize the images for the machine learners. For each dataset, each image in a sequence is converted to 8-bit grayscale from color (RGB). Face detection is applied in order to crop images according to a specific size containing only the face. In [12], eyes are detected using an AdaBoost ensemble cascade of classifiers trained on Haar-like features that model sharp changes in contrast (such as edges) [17, 18]. In [13], faces are detected using a Support Vector Machine (SVM) trained one a reduced set vector approximation of the 2-D image matrix [19]. Both methods are performant, though recognizing eyes first allows for better standardization of the dimensions of the face region across individuals using the distance between eyes as a base proportion from which a crop region can be calculated.

In [12], further processing is performed to address low temporal resolution by interpolating frames in between captured frames. To limit the variation of faces, normalization is performed by extracting 68 feature points from an Active Shape Model (ASM) [20] and transforming each individual face according to the mean face using a Local Weighted Mean (LWM) [21]. From these normalized images, for each micro-expression sequence, intermediate (non-captured) frames are interpolated using a Temporal Interpolation Model (TIM) using a graph embedding technique.

2.3 *Feature Extraction*

Both methodologies use common texture features to represent the images. Texture features are commonly used in facial expression recognition and closely approximate facets of human perception. The proposed work will compare the performance of Local Binary Pattern (LBP used in [12]) and Gabor features (used in [13]).

2.4 *Machine Learning*

Support Vector Machine (SVM) [22] is a classifier that determines the optimal hyperplane separating classes in a high-dimensional space [23] linearly or after a kernel transformation. It has been used with Gabor features or LBP for facial expression recognition [24] with accuracies up to 92.2 % for recognizing the six basic emotions in a macro-expression task. It was also used in [12] to recognize positive vs. negative emotions in the micro-expression SMIC dataset with an accuracies of 54.2% and 62.8% (when used with TIM). SVMs perform well with a high-dimensional feature space and can handle unbalanced datasets with low representation of a certain class, making them useful for our application with few examples of micro-expression frames. Also, using SVMs is common in both micro- and macro-expression literature, motivating our use for it in our benchmarking task.

Boosting classifiers such as AdaBoost [25, 26, 27] or GentleBoost [28] generate multiple classifications opinions using individual weak learners named iterations. In these methods, emphasis is placed on mis-classified examples between iterations by increasing the sampling weights of mis-classified cases (and symmetrically decreasing weights of correctly classified ones). The algorithm is geared towards placing more weight or emphasis on patterns that are difficult to classify [29] and can offset the representation problem of micro-expressions due to their short duration. In [13], GentleBoost performed recognition using Gabor features with an accuracy of 86.5% (average of 6 emotions and neutral recognition).

3. Research Design and Methodology

3.1 *Data*

The most complete micro-expression dataset available is the CASME2 [16]. The video sequences will be separated into training (60%) and testing (40%) sets stratified by the distribution of emotion labels.

3.2 *Computer Vision Techniques*

The following techniques will be applied to each set in combinations proposed in section 3.4.

3.2.1 Image pre-processing: grayscale conversion, face detection and size normalization

All configurations require the conversion of each video frame to gray scale as well as face detection using a cascade of boosting machine learners trained on Haar-like features extracted from faces [30]. This detector can also be trained to find eyes in an image, allowing to crop an image according to a region of fixed size about the eyes. Once eyes are located, the image is cropped according to a function of the distance between eyes.

3.2.2 Active Shape Model Normalization using Local Weighted Means

An Active Shape Model (ASM) fits an image representation of an object with a simple prototypical representation it. In our work, we will fit each face frame using a 68-point standard model that is iteratively deformed to fit the features as they appear in the face, matching points to image locations using standard template matching as in [20, 12]. From all frames in a micro-expression sequence, a "mean-face" is determined by calculating the mean location of each point of the ASM. Frames within a sequence are normalized according to this "mean-face" using a Local Weighted Means (LWM) transformation [21] in order to reduce sources of variation not relating to micro-expressions. LWM is an image registration technique that determines a function $F = f(x, y)$ capable of mapping each point from one ASM representation to another. In this work, we will determine the mapping function between the "mean-face" of a sequence and its corresponding first neutral face. After this transformation is determined, it will be applied to all other frames in a given sequence, normalizing the ASMs according to the "mean-face".

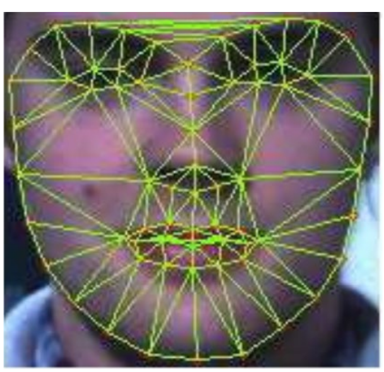

Figure 2. Example of 68 points of ASM from [12].

3.2.3 Temporal Interpolation Model

In order to provide more training examples, we will produce temporally interpolated frames ξ_i as in [12]. The model assumes that each observed frame in a sequence is sampled from a curve, represented as a continuous function in a low dimension manifold. Frames are interpolated using an approximation of the curve $\mathcal{F}^n(t)$ and the Singular Value Decomposition of weights providing a mapping between adjacent frames (**U**). To interpolate a single frame, we use:

$$\boldsymbol{\xi} = \boldsymbol{U}\boldsymbol{M}\mathcal{F}^n(t) + \bar{\boldsymbol{\xi}}$$

3.2.4 Gabor texture features

Gabor filters will be applied with 9 scales (v) and 8 orientations (u) according to [13]:

$$\Psi_{u,v}(z) = \frac{\left\|k_{u,v}\right\|^2}{\sigma^2} e^{-\frac{\left\|k_{u,v}\right\|^2 \|z\|^2}{2\sigma^2}} \left(e^{ik_{u,v}z} - e^{-\frac{\sigma^2}{2}}\right)$$

For each pixel intensity $z = (x, y)$, $\Psi_{u,v}$ is calculated with $k_{u,v} = \frac{k \max e^{i\phi}}{f^v}$

Parameters are $\sigma = 2\pi$, $\phi_u = \frac{u\pi}{8}$, $u \in [0,1]$, $k_{max} = \frac{\pi}{2}$, $f = \sqrt{2}$

3.2.5 Local Binary Pattern texture feature

Local Binary Pattern (LBP) proposed in [31] is a popular texture-based feature representation for faces studied in [32]. It is shown to be ``highly discriminative'', computationally efficient and invariant to ``monotonic graylevel change'' (possibly eliminating the need for image-enhancement pre-processing). LBP is a filter of size $p \times p$ with a distance d centered around pixel p_c. The intensity value of p_c becomes a threshold value for the filter, with neighboring pixels (d away from p_c being replaced by values of 0 or 1 (binary) depending on whether their value is less or greater than the threshold. After the filter is applied to a single neighborhood, it is represented by string collected by traversing the neighborhood in a clockwise fashion. When this filter is applied to the entire image, it is represented by a histogram of counts for each unique binary pattern. These binary patterns are thought to represent texture primitives such as spots, end of lines, edges and corners.

3.3 Machine Learning

Once images are pre-processed and features are extracted, classification models capable of determining which emotion is expressed in each sequence will be learned using two techniques: Support Vector Machines and GentleBoost. SVM is a standard learning method in which the hyperplane that bests separate the classes in the given feature space is determined [22]. The GentleBoost classifier algorithm is detailed in [13], and is an ensemble classifier in which generated weak learners are compared and discarded if it is too similar are previous iterations according to the Mutual Information (MI) measuring entropy. GentleBoost as proposed increases the sampling weights of misclassified cases on previous iterations (while symmetrically decreasing weights of correctly classified cases). However, this process is resource intensive and not efficient. To correct for this, Dynamical Weight Trimming (DWT) will be performed, in which cases with weights less than a threshold (determined by a minimum percentile) will be thrown out for training (for the iteration being performed).

3.4 Research Design

According the literature, we have determined three primary research questions:

1. Given the mature state of macro-expression research, are temporally-sensitive methods (unique to micro-expression recognition) required to identify micro-expressions?
2. In terms of feature representation, which texture features (LBP or Gabor) perform better?
3. Given the two micro-expression specific methods in [12, 13], which performs better when applied to the same dataset?

Using these questions, we have generated a table of different combinations of computer vision techniques to apply to the CASME II dataset, from which pairs of configurations can be selected to isolate the impact of the selected computer vision techniques. Paired t-test will be run on the emotion recognition accuracy results in order to determine whether there is a significant difference between the emotion recognition accuracy between two techniques.

	Macro-techniques		***Micro-techniques***			
	Method 1	***Method 2***	***Method 1***		***Method 2***	
	A	***B***	***A [12]***	***B***	***A***	***B [13]***
Image processing	None	None	TIM	TIM	None	None

Feature extraction	LBP	Gabor	LBP	Gabor	LBP	Gabor
Machine learning	SVM	SVM	SVM	SVM	GentleBoost	GentleBoost

Table 2. Combinations of Computer Vision techniques

3.4.1 Question 1: Which performs better: Macro- or micro- techniques?

The general hypothesis can be stated as follows: h_0: $\mu_{micro} = \mu_{macro}$, h_1: $\mu_{micro} > \mu_{macro}$

In these comparisons we isolate the temporally sensitive methods(Temporal Interpolation Model (TIM) and the use of GentleBoost) and compare their performance against methods that do not use them. TIM interpolates unobserved frames to provide more training samples. Using GentleBoost increases the weights of mis-classified cases, producing more samples of troublesome training examples. In order to isolate which temporally aware method performs better we use the following head-to-head comparisons (the subscript of the population mean determines which combination from Table 2 is used):

Utility of Temporal Interpolation Model		*Utility of GentleBoost*	
LBP features	Gabor features	LBP features	Gabor features
h_0: $\mu_{micro1A} = \mu_{macro1A}$ h_1: $\mu_{micro1A} > \mu_{macro1A}$	h_0: $\mu_{micro1B} = \mu_{macro2B}$ h_1: $\mu_{micro1B} > \mu_{macro2B}$	h_0: $\mu_{micro2A} = \mu_{macro1A}$ h_1: $\mu_{micro2A} > \mu_{macro1A}$	h_0: $\mu_{micro2B} = \mu_{macro2B}$ h_1: $\mu_{micro2B} > \mu_{macro2B}$

3.4.2 Question 2: Which performs better: LBP or Gabor features?

In the literature of both macro- and micro-expression research, Local Binary Pattern (LBP) or Gabor features are commonly extracted. In order to determine which feature representation is best, we will pair up techniques in such a way that the only difference between them is the type of feature extracted. The general hypothesis can be stated as follows: h_0: $\mu_{LBP} = \mu_{Gabor}$, h_1: $\mu_{LBP} > \mu_{Gabor}$

The head-to-head comparisons to perform are:

Macro-techniques	Micro- method 1	Micro- method 2
h_0: $\mu_{macro1A} = \mu_{macro2B}$ h_1: $\mu_{macro1A} > \mu_{macro2B}$	h_0: $\mu_{micro1A} = \mu_{micro1B}$ h_1: $\mu_{micro1A} > \mu_{micro1B}$	h_0: $\mu_{micro2A} = \mu_{micro2B}$ h_1: $\mu_{micro2A} > \mu_{micro2B}$

3.4.3 Question 3: Which performs better: [12] or [13]?

Finally, since only two micro-expression techniques are currently used in the literature (that fit the traditional Computer Vision recognition methodology), we compared these two techniques head-to-head:

$$h_0: \mu_{micro1A} = \mu_{micro2B}$$
$$h_1: \mu_{micro1A} > \mu_{micro2B}$$

An example summary table for the recognition results can be composed as per Table 3 (with recognition rate as RR).

Emotion	***Macro-techniques***		***Micro-techniques***			
	Method 1	***Method 2***	***Method 1***		***Method 2***	
	A	***B***	***A [12]***	***B***	***A***	***B [13]***
Happiness	RR	RR	RR	RR	RR	RR
Disgust	RR	RR	RR	RR	RR	RR
Surprise	RR	RR	RR	RR	RR	RR
Repression	RR	RR	RR	RR	RR	RR
Others	RR	RR	RR	RR	RR	RR
Neutral	RR	RR	RR	RR	RR	RR
All (as average)	RR	RR	RR	RR	RR	RR

Table 3. Sample Summary Results Table

Superscripts such as ^{1,2,3} will be used to show the significance of a technique relative to its head-to-head comparisons for each research question. For example:

Emotion	***Macro-techniques***		***Micro-techniques***			
	Method 1	***Method 2***	***Method 1***		***Method 2***	
	A	***B***	***A [12]***	***B***	***A***	***B [13]***
Happiness	RR^{2}	RR	RR^{1,2,3}	RR^{1}	RR^{2}	RR

This table can be read as:

1. Using the Temporal Interpolation Model to recognize micro-expressions leads to significant recognition improvement for happiness when both LBP and Gabor features are used; Micro1A is significantly better at a 5% significance level than Macro1 and Micro1B is significantly better at a 5% significance level than Macro2B. Using the GentleBoost method for temporal resolution shows no significant improvement upon methods not using it.
2. Models using LBP features significantly outperform those using Gabor for the recognition of happiness; Macro1A is significantly better at a 5% significance level than Macro2B, Micro1A is better than Micro1B, and Micro2A is better than Micro 2.
3. The proposed methodology of [12] outperforms at a 5% significance level the methodology of [13] for the recognition of happiness.

Improvements upon this visualization will be explored.

4. Plan of Work

The work is mostly implementation. Each objective corresponds to an implementation as well as writing pertinent sections of a paper. Each objective also corresponds to a Computer Vision technqiu

Week	Step	Objective
1 & 2	1	Implement image normalization: Active Shape Models (ASM) and Least Weighted Means (LWM)
2 & 3	2	Implement Temporal Interpolation Model (TIM)
4	3	Apply steps 1 and 2 to image data to generate normalized interpolated images
5 & 6	4	Implement Local Binary Pattern (LBP) and Gabor feature extractors
7	5	Apply step 4 to images, perform emotion classification using SVMs
8	6	Implement GentleBoost and apply to features extracted
9	7	Perform t-tests according to evaluation method
10		Writing: synthesis, discussion, connections

The final deliverable will include a complete API that can be applied generally to other datasets as well as a manuscript for publication that will be of interest to researchers in the field of micro-expression recognition.

5. Conclusions

The improper application of the scientific method to the field of micro-expression recognition leads to invalidated claims of the requirement of new techniques for the field (compared to that of macro-expression recognition). Furthermore, competing techniques have yet to be properly compared. Creating the benchmarks proposed will lead to more informed decision making for researchers aiming to study social phenomena relying upon the identification of both macro- and micro-expressions.

Bibliography

[1] M. G. Frank and E. Svetieva, “Microexpressions and deception,” in *Understanding Facial Expressions in Communication*. Springer, 2015, pp. 227–242.

[2] E. Friesen and P. Ekman, “Facial action coding system: a technique for the measurement of facial movement,” *Palo Alto*, 1978.

[3] J. F. Cohn, Z. Ambadar, and P. Ekman, “Observer-based measurement of facial expression with the facial action coding system,” *The handbook of emotion elicitation and assessment*, pp. 203–221, 2007.

[4] D. Matsumoto, D. Keltner, M. N. Shiota, M. Oâ€™Sullivan, and M. Frank, “Facial expressions of emotion,” *Handbook of emotions*, vol. 3, pp. 211–234, 2008.

[5] J. M. Gottman and R. W. Levenson, “A two-factor model for predicting when a couple will divorce: Exploratory analyses using 14-year longitudinal data*,” *Family process*, vol. 41, no. 1, pp. 83–96, 2002.

[6] P. Ekman and W. V. Friesen, “Nonverbal leakage and clues to deception,” *Psychiatry*, vol. 32, no. 1, pp. 88–106, 1969.

[7] J. F. Cohn, “Foundations of human computing: facial expression and emotion,” in *Proceedings of the 8th international conference on Multimodal interfaces*. ACM, 2006, pp. 233–238.

[8] M. Kenji, “Recognition of facial expression from optical flow,” *IEICE TRANSACTIONS on Information and Systems*, vol. 74, no. 10, pp. 3474–3483, 1991.

[9] T. Wu, S. Fu, and G. Yang, “Survey of the facial expression recognition research,” in *Advances in Brain Inspired Cognitive Systems*. Springer, 2012, pp. 392–402.

[10] G. Warren, E. Schertler, and P. Bull, “Detecting deception from emotional and unemotional cues,” *Journal of Nonverbal Behavior*, vol. 33, no. 1, pp. 59–69, 2009.

[11] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, “The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression,” in *Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on*. IEEE, 2010, pp. 94–101.

[12] T. Pfister, X. Li, G. Zhao, and M. Pietikainen, “Recognising spontaneous facial micro-expressions,” in *Computer Vision (ICCV), 2011 IEEE International Conference on*, Nov 2011, pp. 1449–1456.

[13] Q. Wu, X. Shen, and X. Fu, “The machine knows what you are hiding: an automatic micro-expression recognition system,” in *Affective Computing and Intelligent Interaction*. Springer, 2011, pp. 152–162.

[14] P. Ekman and E. L. Rosenberg, *What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS)*. Oxford University Press, 1997.

[15] T. Kanade, J. F. Cohn, and Y. Tian, “Comprehensive database for facial expression analysis,” in *Automatic Face and Gesture Recognition, 2000. Proceedings. Fourth IEEE International Conference on*. IEEE, 2000, pp. 46–53.

[16] W.-J. Yan, X. Li, S.-J. Wang, G. Zhao, Y.-J. Liu, Y.-H. Chen, and X. Fu, “Casme ii: An improved spontaneous micro-expression database and the baseline evaluation,” *PloS one*, vol. 9, no. 1, p. e86041, 2014.

[17] P. Viola and M. Jones, “Rapid object detection using a boosted cascade of simple features,” in *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, vol. 1. IEEE, 2001, pp. I–511.

[18] P. I. Wilson and J. Fernandez, “Facial feature detection using haar classifiers,” *Journal of Computing Sciences in Colleges*, vol. 21, no. 4, pp. 127–133, 2006.

[19] W. Kienzle, M. O. Franz, B. Schölkopf, and G. H. Bakir, “Face detection—efficient and rank deficient,” in *Advances in Neural Information Processing Systems*, 2004, pp. 673–680.

[20] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, “Active shape models-their training and application,” *Computer vision and image understanding*, vol. 61, no. 1, pp. 38–59, 1995.

[21] A. Goshtasby, “Image registration by local approximation methods,” *Image and Vision Computing*, vol. 6, no. 4, pp. 255–261, 1988.

[22] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.
[23] M. A. Hearst, S. T. Dumais, E. Osman, J. Platt, and B. Scholkopf, "Support vector machines," *Intelligent Systems and their Applications, IEEE*, vol. 13, no. 4, pp. 18–28, 1998.
[24] C. Shan, S. Gong, and P. W. McOwan, "Robust facial expression recognition using local binary patterns," in *Image Processing, 2005. ICIP 2005. IEEE International Conference on*, vol. 2. IEEE, 2005, pp. II–370.
[25] Y. Freund, R. E. Schapire *et al.*, "Experiments with a new boosting algorithm," in *ICML*, vol. 96, 1996, pp. 148–156.
[26] J. Zhu, H. Zou, S. Rosset, and T. Hastie, "Multi-class adaboost," *Statistics and its Interface*, vol. 2, no. 3, pp. 349–360, 2009.
[27] R. E. Schapire, "Explaining adaboost," in *Empirical inference*. Springer, 2013, pp. 37–52.
[28] J. Friedman, T. Hastie, R. Tibshirani *et al.*, "Additive logistic regression: a statistical view of boosting (with discussion and a rejoinder by the authors)," *The annals of statistics*, vol. 28, no. 2, pp. 337–407, 2000.
[29] L. Rokach, "Ensemble-based classifiers," *Artificial Intelligence Review*, vol. 33, no. 1-2, pp. 1–39, 2010.
[30] R. Lienhart and J. Maydt, "An extended set of haar-like features for rapid object detection," in *Image Processing. 2002. Proceedings. 2002 International Conference on*, vol. 1. IEEE, 2002, pp. I–900.
[31] T. Ojala, M. Pietikäinen, and D. Harwood, "A comparative study of texture measures with classification based on featured distributions," *Pattern recognition*, vol. 29, no. 1, pp. 51–59, 1996.
[32] T. Ahonen, A. Hadid, and M. Pietikainen, "Face description with local binary patterns: Application to face recognition," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 28, no. 12, pp. 2037–2041, 2006.

Appendix

Budget: $5,000 (10 week's salary)