Preprocessing and Descriptor Features for Facial Micro-Expression Recognition

Chris House and Rachel Meyer
Department of Electrical Engineering, Stanford University
{chouse12, r3m3y3r}@stanford.edu

June 5, 2015

Abstract

Facial micro-expressions contain significant information about how people feel, even when they are attempting to conceal their emotions. Previously, very little research has been done to detect and recognize micro-expressions using computer vision methods. In this paper, detection and classification of micro-expressions from the Spontaneous Micro-Expression database were implemented, following preprocessing and cropping of raw images using Haar features, using local binary patterns on three orthogonal planes (LBP-TOP) and local gray code patterns on three orthogonal planes (LGCP-TOP) as descriptors and support vector machines (SVMs) as detection and recognition algorithms. Results show accuracy comparable to other work.

1 Introduction

Facial expressions contain significant information about how a person feels. Humans are adept at recognizing macro-expressions that occur for time periods on the order of a few seconds, but humans are able to manipulate these expressions to hide their true emotions. Micro-expressions, on the other hand, are short (lasting less than half a second), involuntary expressions that reveal true feelings even when someone is attempting to conceal their emotions. These expressions are extremely difficult to detect due to their short time scales and subtlety. Micro-expressions were first reported in psychological literature in the 1960's and have been studied since then[1, 2]. Recently, a few research groups have attempted micro-expression detection and recognition using computer vision techniques. Trained humans can recognize micro-expressions accurately about 47% of the time [3], but perhaps computer vision methods can achieve higher accuracy. The applications of technology that can successfully detect and recognize micro-expressions are varied. It would be valuable in law enforcement interrogations to detect deceit, in marketing to detect how humans respond to advertising, and in general psychology and artificial intelligence research to study human emotion.

2 Background

Previous work in micro-expression analysis has taken mostly a psychological analysis form. The psychologists Haggard and Isaacs first proposed the idea of micro-expressions, or short, involuntary reactions that are immediately retracted in an attempt at suppression, in a 1966 study [4]. Paul Ekman, noted for his work in facial expression analysis, published his seminal work in 1968 regarding "Nonverbal Leakage and Clues to Deception," in which he detailed his findings in detecting deception. Ekman used microexpression analysis in conjunction with his previous work regarding emotion classification based off of facial patterns to further understand why humans lie and how they display deceit nonverbally [5]. In 2002, John Gottman, a famous psychologist for his work in predicting divorce rate based off of a couples' interaction, used micro-expression analysis from short video clips to enhance his understanding of couple interaction and his divorce prediction ability [6]. All of these psychological experts certainly gained an ability to detect and recognize micro-expressions. However, while all of this work helped increase understanding of micro-expressions from a psychological standpoint, none of the work sought to gain a generalized ability to detect or recognize micro-expressions. Thus, in the last few years, several groups have attempted to create general computer vision methods that can detect and recognize specific facial micro-expressions. Until recently, the biggest challenge in attempting to create a general method for micro-expression detection and recognition had been a lack of useful data. However, in 2013, two databases were developed and released for public use in research with micro-expressions: the SMIC database (Spontaneous Micro-Expression Database) and CASME II. These databases were developed under similar premises: test subjects were filmed using high speed cameras while being shown various video clips that were intended to induce emotional reactions. The test subjects were also instructed to suppress an outward reaction so as to properly induce micro-expressions [7, 8]. Other databases have also been constructed but these databases are generally less desirable for analysis since the micro-expressions were obtained through posed expressions (test subject told to display facial expression corresponding to an emotion, but with low muscle intensity) or while the test subject was talking or moving their head (which makes computer vision analysis more difficult as it essentially introduces noise) [7]. Other more general facial expression databases include CK, MMI, JAFFE, RUFACS, MULTI-PIE, and Oulu-CASIA. A number of different approaches have been proposed and implemented for detection and recognition of micro-expressions [9]. Some groups have used Gabor features and algorithms such as SVM, Adaboost, and other variations of those to recognize micro-expressions [10]. Others have divided the face into sub-regions and analyzed facial strain in these regions for detection [11], or used 3D gradient descriptors [12]. Temporal interpolation models and multiple kernel learning have also been explored [3]. However, the most common method by far has been Local Binary Patterns on Three Orthogonal Planes (LBP-TOP) combined with various learning techniques for classification [7, 8]. In the last several years, several groups have achieved detection accuracies on the order of 60-70% (with 50% as the baseline chance detection) and recognition accuracy between 3 classes of emotions on the order of 50% (3 classes, so 33% as the baseline chance detection) [7]. Another newly developed descriptor, Local Gray Code Patterns (LGCP), has claims of improving upon the deficiencies of the descriptors listed above, most notably the memory requirements of Gabor features and the inability to handle scene illumination variance and noise for LBP [13].

2.1 Our Contributions

Our project evaluates previous research by implementing similar descriptors and algorithms. Our main contributions include a more thorough analysis of the results and implementing a newly developed feature descriptor, LGCP-TOP.

3 Methods

For this project, we used the Spontaneous Micro-Expression Database (SMIC). The database contains both raw images and preprocessed and cropped images. The raw images were cropped using Haar features to detect the subject's face. The descriptors used were local binary patterns on three orthogonal planes (LBP-TOP) and local gray code patterns on three orthogonal planes (LGCP-TOP). Support vector machines (SVMs) were used for detection and recognition.

3.1 SMIC Database

The SMIC database, developed by the research group led by Xiaobai Li at University of Oulu, was created specifically for work with micro-expression recognition. In the SMIC dataset, the test subjects' facial expressions were filmed while watching 16 various movie clips designed to induce strong emotions. Each subject was instructed to try to conceal their reaction to each clip as much as possible, as the testers would attempt to guess their reactions to each clip afterwards. As an incentive for the test subjects, the testers added a penalty of having to complete a long, boring questionnaire if they were to accurately guess the subjects' emotions. In this way, the data obtained was a more accurate representation of how humans produce micro-expressions. The database contains 164 micro-expression video clips elicited from 16 participants. These micro-expressions were classified into positive, negative, and surprise categories. We utilized the high speed video (100 fps) data from the dataset (other options included video sequences at 25 fps and near-infrared images). Table 1 shows the composition of the dataset.

Table 1: Composition of the SMIC database. Note that these images are available both as raw images and cropped images.

Subjects	Micro-Expressions				Non-Micro-Expressions
	Positive	Negative	Surprise	Total	Total
16	51	70	43	164	164

3.2 Preprocessing of Images

Using the raw images from the SMIC database, we implemented the CascadeObjectDetectorSystem from the MATLAB computer vision toolbox. This object detector has several built-in object detectors, including three face detectors (two using Haar features, one using LBP), eye detectors, a nose detector, and a mouth detector. We implemented the face detector using Haar features for facial features, which are then classified using a classification and regression tree (CART). The output of this function was a bounding box for the given image. Using the bounding box, the image was then cropped (to a uniform size) and then resized in order to allow for faster processing in obtaining the feature descriptors later on. In comparison to the preprocessed images in the SMIC database, our cropping method achieved a similar result.

Figure 1: Process outlining the cropping of the raw images [3].

3.3 LBP-TOP

Choosing the features for micro-expressions is a difficult problem, as both spatial and temporal information should be encoded into them. Additionally, micro-expressions are very subtle and last for a short period of time. The local binary patterns on three orthogonal planes (LBP-TOP) descriptor was selected after an extensive survey of relevant literature [7, 8]. The local binary pattern (LBP) descriptor, from which LBP-TOP was derived, was developed in 1994 and is used commonly in computer vision applications. Considering just one pixel in the image, a LBP is computed by comparing the pixel value with that of its neighbors, and then converting the binary string representing the local texture into a decimal number. A feature vector can be formed by calculating a histogram of LBPs over an entire image. The LBP method is effective for describing 2D textures of static images, but to analyze time-dependent textures (i.e. changing expressions in video), this model needs to be extended. To do so, LBP histograms are calculated in three orthogonal planes. For a video with time length T, LBP are computed for the XY-, XT-, and YT-planes. The XY-plane describes the spatial changes, while the XT- and YT-planes describe the spatial-temporal change in each respective dimension. These histograms are then concatenated to form the final LBP-TOP feature vector. Thus, LBP can be extended to our videos of facial micro-expressions. It is difficult to observe noticeable patterns in micro-expressions data because micro-expressions are reactions which one deliberately tries to suppress, which causes facial expressions to have very low intensity. As a result, additional modifications are usually implemented in micro-expression recognition. Uniform binning is used to reduce dimensionality and improve information quality. Some groups also implement temporal interpolation models (TIM) to achieve more stable results [3]. In our experiments, we adjusted the radius of the descriptor and number of neighboring points, and we also investigated uniform/non-uniform binning and bi-linear interpolation. All of the results reported used eight neighboring points and bi-linear interpolation. Temporal interpolation was not implemented yet, but could be worth incorporating in the future as it can improve accuracy.

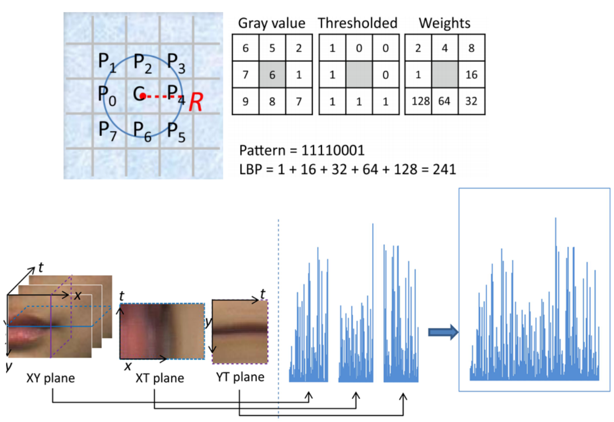

Figure 2: Example showing LBP calculation (top) and concatenation (bottom) to form an LBP-TOP descriptor [8].

3.4 LGCP-TOP

Like LBP-TOP, Local Gray Code Patterns in Three Orthogonal Planes (LGCP-TOP) is a descriptor used for computer vision applications [13]. Developed in 2013, LGCP involves using Robinson Compass Masks convolved with a 3x3 region to get 8 edge directions for the center pixel. Then, using bitwise comparison between neighboring pixels in a clockwise orientation, the edge direction magnitudes are converted to gray code, as shown in Figure 2. With the whole image split into sub-blocks (typically 81 sub-blocks in a 9x9 grid), the gray code values for each pixel are then used to form a histogram for each sub-block. When computing the histogram, unused bins are discarded, allowing the dimensionality of the feature vector to be reduced from 256 bins to 20 bins. At this point, an optional weighting can be performed to give bins with higher "term frequency" more importance in the feature vector. Following this, the histogram of each sub-block is then concatenated to form a feature vector for the whole image of size 1215 (81x15). While LGCP is a new feature descriptor, promising results have been witnessed. According to the authors, LGCP is particularly good at handling noisy images, as shown in Figure 3. In terms of classification accuracy for facial expressions, LGCP was able to achieve improvements in facial expression recognition for both the JAFFE database and CK+ database (databases with 7 classes of macro-expressions, and 326 images and 213 images, respectively) over other feature descriptors, including LBP, Gabor features, and LPQ (local phase quantization) [13].

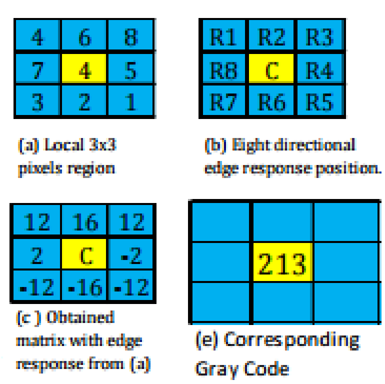

Figure 3: Example showing LGCP calculation [13].

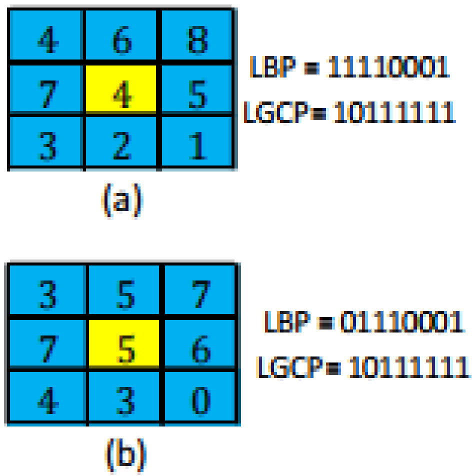

Figure 4: Example showing LGCP's invariance to noise, as opposed to LBP's inability to handle the noisy image (a is original image, b is noisy image) [13].

3.5 Detection and Recognition Algorithms

Previous researchers have used Support Vector Machines (SVM), Random Forests (RF), and Multiple Kernel Learning (MKL) as detection and classification algorithms. SVMs with RBF kernels were chosen as a baseline algorithm to compare to other work. The dataset was split by subject into training (75% of the data) and testing sets (25% of the data) by subject in an effort to test how well these algorithms would generalize to unseen people. Leave-one subject-out cross-validation was performed on the training set to optimize parameters for LBP-TOP and the SVM. The detection and recognition were done in separate experiments so that we could isolate the effects of each stage separately.

4 Results

4.1 LBP-TOP

As shown in Table 2, detection accuracy achieved almost a 19% improvement over chance detection (50%) and recognition accuracy achieved an almost 16% improvement over chance recognition (33%). These results are comparable to those reported by Li and Pfister [7], who achieved 65.5% detection accuracy and 49.3% recognition accuracy using LBP-TOP and SVM. Those results are not directly comparable to ours, however, as they reported accuracies based on leave-one-subject-out cross validation as opposed to accuracies obtained from testing their model on a separate test group. We also examined the difference in results achieved while varying the parameters used in the LBP-TOP descriptors. We computed LBP-TOP descriptors without uniform binning and spatial radius of one, descriptors with uniform binning and spatial radius of one, and descriptors with uniform binning and spatial radius of three. The results were interesting, as the descriptors without uniform binning achieved the best detection accuracy but worst recognition accuracy. Perhaps uniform binning reduces useful information in terms of discriminating between micro-/non-micro expressions, as it drastically reduces the number of bins in the histogram. In the case of recognition, uniform binning might eliminate some of the noise that occurs in expressions and retain the useful information that allows the SVM to discern between different micro-expressions. It also seems that finding an optimal spatial radius can improve the recognition accuracy as the descriptors with uniform binning and a spatial radius of three achieved the best recognition accuracy. Finding optimal LBP-TOP spatial radius can help find the right level of localization and changes in features, which allows for better discrimination between different types of micro-expressions. Detection and classification of micro-expressions is extremely challenging. Figure 2 shows two examples of micro-expressions from the database. The small changes in expression that these subjects show is difficult to represent in a descriptor. We noticed that small changes in the descriptor and machine learning algorithms drastically affected performance. It makes sense that detection accuracy was better than recognition accuracy (19% above chance vs. 16% above chance), as when detecting micro-expressions the classifier is looking for short periods of movement. This task is much easier than recognition, where the classifier must discern between differ-

Figure 5: Example of two subjects displaying micro-expressions (positive on the left - note the slight curl of the lip, negative on the right - note the slight furrow of the brow) [3].

ent types of subtle movements that may be somewhat different between subjects.

4.2 LGCP-TOP

The results for LGCP-TOP were less encouraging than those of LBP-TOP, shown below in Table 2. The top detection accuracy observed with LGCP-TOP descriptors using SVM's as a classifier was about 61% and the top recognition accuracy was about 48%. These accuracies still represent improvements over chance (11% and 15%) yet these accuracies still fall short of the the results witnessed with LBP-TOP. There are several possibilities for these discrepancies. One is that the feature vectors from LGCP-TOP were far too big for SVM's to classify without overfitting. In an attempt to counter this, we performed PCA through a low-rank SVD approximation to reduce the feature vectors from dimensionality greater than 4000 to less than 1000. This helped improve the recognition accuracy (from 38% to 48%) but did not improve detection accuracy significantly (57% to 61%). Another possibility is that the sub-block structure used in computing the LGCP-TOP descriptors was not optimized for recognition or detection. With LBP-TOP, changing the spatial radius of the neighbor points and binning for the descriptor had large effects on the detection and recognition accuracy, as increasing the spatial radius from 1 to 3 improved recognition accuracy, while not using uniform binning (reducing dimensionality from 256 to 59 by discarding repeated bins) with a spatial radius of 1 improved detection accuracy. Our tests of LGCP-TOP block structures were not exhaustive although several tests seemed to indicate that block structure only played a minimal role in the results obtained. However, despite the lower accuracy, LGCP-TOP was advantageous over LBP-TOP as it was approximately 15x faster than LBP-TOP in computing the feature descriptors. This has to do with the block structure utilized with LGCP, whereas LBP simply iterates over every pixel in each image sequence. The faster computation time allowed for testing with different block structures and easier debugging, although the block structures did not have a significant effect in improving accuracy.

Table 2: Final results for each descriptor.

Method	Detection Accuracy	Recognition Accuracy
LBP-TOP (Li, et al.)	70.3%	52.5%
LBP-TOP	68.9%	48.6%
LGCP-TOP	61.2%	48.1%

5 Conclusion

A method for detecting and identifying facial micro-expressions was implemented. Using the SMIC database, preprocessing of raw images was implemented using Haar features to identify a facial structure and the images were cropped based off of the resulting bounding box. Two feature descriptors, LBP-TOP and LGCP-TOP were implemented and tested using SVM's as a classifier. LBP-TOP achieved results comparable to Li's group, while LGCP-TOP fell short in detection accuracy but close with recognition. LGCP-TOP also had the advantage of much shorter computing time to compute the feature descriptors.

6 Future Work

Future work will seek to improve detection and recognition accuracy in several manners. One possibility for the poorer performance of LGCP-TOP was that its feature dimensionality was too large to classify without overfitting. As noted previously, LGCP-TOP descriptor dimensionality was reduced using PCA. Other possibilities for decreasing dimensionality could involved performing weighting over the image block structure used in computing LGCP descriptors. Blocks close to the center of the face would most likely be more important than blocks towards the edges of the image and would thus be weighted more heavily. Other possible improvements could come from face alignment. Face alignment would be performed to align each image to a canonical position, such as aligning the position of each subject's nose or eyes. Finally, improvements could be witnessed with different micro-expression databases. Other databases, such as CASME II which contains spontaneous micro-expression image sequences, might have better image quality. This could certainly help improve performance, particularly with LBP-TOP, given that LBP is somewhat more susceptible to noise and illumination changes.

Acknowledgements

We would like to thank Huizhong Chen as our project mentor and the entire EE368 teaching staff for the work over the entire quarter.

References

[1] Haggard, Ernest A., and Kenneth S. Isaacs. "Micromo- mentary facial expressions as indicators of ego mechanisms in psychotherapy." Methods of research in psychotherapy. Springer US, 1966. 154-165.

[2] Ekman, Paul, and Wallace V. Friesen. "Nonverbal leakage and clues to deception." Psychiatry 32.1 (1969): 88-106.

[3] Pfister, Tomas, et al. "Recognising spontaneous facial micro-expressions."Computer Vision (ICCV), 2011 IEEE International Conference on. IEEE, 2011.

[4] Isaacs, K. S., et al. "Micro-momentary facial expressions as indicators of ego mechanisms in psychotherapy." Methods of Research in Psychotherapy, 1966. 154-165.

[5] Ekman, Paul, et al. "Nonverbal Leakage and Clues to Deception." Psychiatry: Interpersonal and Biological Processes. 1969. 88-109.

[6] Gottman, John, et al. "A Two-Factor Model for Predicting When a Couple Will Divorce: Exploratory Analyses Using 14-Year Longitudinal Data." Family Process. 83-86.

[7] Li, Xiaobai, et al. "A spontaneous micro-expression database: Inducement, collection and baseline." Automatic Face and Gesture Recognition (FG), 2013 10th IEEE International Conference and Workshops on. IEEE, 2013.

[8] Yan,Wen-Jing,etal."CASMEII: An improved spontaneous micro-expression database and the baseline evaluation." PloS one 9.1 (2014): e86041

[9] Zeng, Zhihong, et al. "A survey of affect recognition methods: Audio, visual, and spontaneous expressions." Pattern Analysis and Machine Intelligence, IEEE Transactions on 31.1 (2009): 39-58.

[10] Wu,Qi, Xunbing, Shen,and Xiaolan, Fu."The machine knows what you are hiding: an automatic micro-expression recognition system." An Effective Computing and Intelligent Interaction. Springer Berlin Heidelberg, 2011. 152-162.

[11] Shreve, Matthew, et al. "Macro-and micro-expression spotting in long videos using spatio-temporal strain." Automatic Face & Gesture Recognition and Workshops(FG2011), 2011 IEEE International Conference on. IEEE, 2011

[12] Polikovsky, Senya, Yoshinari Kameda, and Yuichi Ohta. "Facial micro-expressions recognition using high speed camera and 3D-gradient descriptor." (2009): 16-16

[13] Islam, Mohammad Shahidul, et al. "Local Gray Code Pattern (LGCP): A Robust Feature Descriptor for Facial Expression Recognition." In-

ternational Journal of Science and Research (IJSR), India Online ISSN: 2319-7064. 2013.