**The Detection and Classification of Universal Emotions**

Austin Dase and Justine Caylor

Towson University

COSC 602 Computer Vision and Image Processing

Team Project 2

**Abstract.** Detecting and classifying emotions from static images is a challenge that is becoming popular and relevant in research conducted today. There are universal emotions that are expressed by individuals that can be identified worldwide, and these emotions are expressed by facial features such as eyebrows, eyes, and the mouth. This research aims to use image processing algorithms to identify the expressive features of the face. In addition, the features identified will be used for emotion classification of the six universal emotions (happiness, sadness, anger, fear, disgust, and surprise) using supervised learning algorithms Decision Tree, Linear Discriminant, Support Vector Machine (SVM), and k-Nearest Neighbor (KNN).

**Keywords:** Emotion Classification, Facial Expression, Object Detection, Edge Detection, Morphological Filtering

1. Introduction

Emotions are commonly identified by physical reactions and behavioral actions [1]. Without words, emotions can also be identified by just an individual’s face. Across the world, there are six universal emotions that can be identified: happiness, sadness, anger, fear, disgust, and surprise [2]. Facial expressions hold a lot of power in the ability to express emotions through the manipulation facial features [3]. The expressive features of the face are the eyebrows, the eyes, and the mouth; all of these features have a different movement and reaction when an emotion arises. For example, to express a surprised emotion: eyes usually widen, eyebrows typically are raised, and most often the mouth is opened. This expression can be seen differently than an individual who is sad. However, identifying and classifying emotion by computer applications is a challenging issue that of interest in research [4].

This research tackles this challenge by using image processing and analysis techniques to identify the expressive features of the face from static images. Once the expressive features are identified, calculations will be computed of which will be used for emotion classification. The paper is organized as the following: Section 2 highlights related worked that has also been conducted in emotion classification, Section 3 is a description of the dataset used in this research, Section 4 describes the system flow model in image processing and analysis, and what was done in this research, Section 5 and 6 highlights the image processing methodology utilizing various algorithms and features extracted, Section 7 is a description of the classification algorithms used to classify emotions, and concludes with Section 8.

1. Related Work

This section provides insight on the related work that has been done in this field on image processing to detect facial features and emotion classification.

Halder et al. focused on eye extraction, eyebrow extraction, and mouth extraction for features to be used for emotion recognition [5]. The eyes are extracted using Sobel operator and Roberts operator, and then the holes are filled using dilation. Eyebrows are extracted Sobel operator and then dilation to fill in holes. The mouth is extracted using Roberts operator and dilation is used to fill in holes. Harris corner point detection was used to detect the corner points of the selected features. The calculated features were used as inputs for a neural network for emotion classification.

Chow and Li proposed a model-based method to detect facial features from images, composed of face location, eyes, and mouth [6]. Morphological filtering was used for capturing face location. Eyes and mouths were confirmed and extracted using Hough transformation.

Lyons, Budynek, and Akamatsu proposed a method to automatically classify images based on labeled elastic graph matching, 2D Gabor wavelet representation, and linear discriminant analysis [7]. In addition to recognizing expression based on the six emotions, the proposed method will be able to classify “race” and gender.

Pal, Iyer, and Yantorno proposed a system that analyzes facial images and sounds of cries from infants to determine the reasons why infants are crying [8]. There are five reasons suggested that infants cry: sadness, anger, hunger, pain, and fear (some of which are similar to the six universal emotions). The facial features that were analyzed were mouth, eyes, and eyebrows. Each of the features were calculated as a state of either open/closed for the mouth or eyes, and up/down for eyebrow positions.

1. Description of the Dataset

The dataset in this study was obtained by The Japanese Female Facial Expression (JAFFE) Database [9]. This database is available as open-access for non-commercial research purposes. The dataset included 213 gray-scale images of 10 Japanese female models that expressed one of seven emotions (happiness, sadness, anger, disgust, surprise, fear, and neutral). Each image is of a TIFF file format. All of the models in the images have a pair of eyebrows, a pair of eyes, one nose, and one mouth.



Figure 1 - Six Universal Emotions

1. System Flow

In image processing, the analysis is described to be accomplished in three steps: preprocessing of the image, data reduction of the image, and feature analysis of the image.

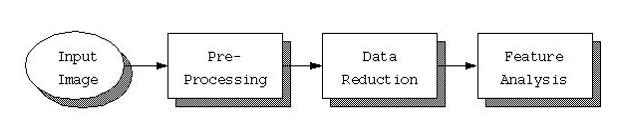


Figure 2 - System Flow Model

Preprocessing is focused on removing unnecessary information within images such as removing noise or reducing the spatial size. Data reduction consists of reducing the data in the spatial domain or transforming it into a frequency domain, and then extracting the desired features for analysis. Feature analysis is described as utilizing the features that resulted after the preprocessing and data reduction process are evaluated.

For this research, the JAFFE dataset will be used as the input for image analysis. The focus of this research is to analyze and classify the six different universal emotions from the images, therefore images that expressed the ‘neutral’ emotion was omitted. Images of individuals in low contrast or had difficult with facial detection was also omitted. Therefore, 111 images were used for feature analysis.

The remove noise and the background from the images, thresholding and filtering was used on the images. Emotions are detectable from the expressive parts of the face: the eyes, the eyebrows, and the mouth. Using segmentation and morphological filtering, only these expressive features were identified as objects to be used for feature analysis. The relevant objects identified are calculated in different ways that would be useful for differentiating between emotions and emotion classification. The calculations in this research was calculating the area, centroid, circularity, and perimeter.



Figure 3 - Objects of Relevant Expressive Features

1. Image Processing Algorithms

The process of extracting the desired features from the image was completed as a multi-step application of various feature extraction algorithms, using Microsoft Visual Studio. Initially, as in [10] an edge-based mask was used as the first step in feature extraction. Unlike [10], which used a canny edge detector to find relevant edges, a *Sobel filter* with a kernel size of 3 was applied, this method was similar to that described in [5]. After applying the Sobel edge detector, an automated thresholding algorithm as suggested in [11] was applied in order to achieve optimal thresholding results, see Fig. 5. A series of morphological filters similar to that described in [10] and [6] was then applied to the image in order to accentuate the regions of interest. The first morphological filter applied was a *dilation* with a structuring element of size 3 x 3, see Fig. 6. That same *dilation* was applied a second time immediately following the first, see Fig. 7, with a goal of creating a mask over the regions of interest. Next, in order to make sure objects identified in the image were not overlapping, a morphological *erosion* was applied to the image, Fig. 8. At this point the first attempt at object detection was attempted. The image, after applying the dilations as well as the erosion, was then processed by two object detection algorithms described in [11]. The first was the *mark8* algorithm used to identify distinct objects in the image, at this point the area of each object was also able to be calculated and stored to be used for later classification. After applying the mark8 filter to identify all the unique objects, the *chain8* algorithm was applied to extract perimeter information from each of the objects, the results of this process can be seen in Fig. 9. In Fig. 9 the pixel for each object is set to the value of its corresponding object number subtracted from 255 so that objects have greater contrast to the background. With the *area*, *diameter,* and *perimeter* for each object now extracted, the *circularity* could also be calculated. The circularity is given as the following formula.

Following the circularity calculation, the centroid for each object was found using the *centroid* detection method described in [11]. All values for centroid, area, perimeter and circularity were then stored for each object identified in the image for use in extracting and identifying the facial features.



Figure 4 Original Image

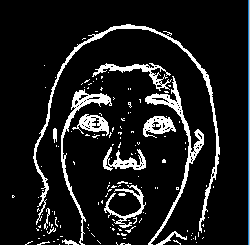


Figure 5 Sobel

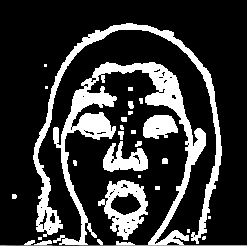


Figure 6 Dilation 1

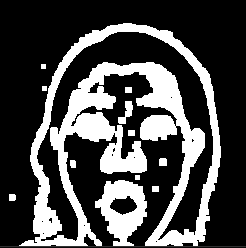


Figure 7 Dilation 2

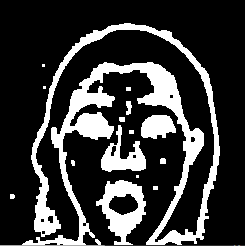


Figure 8 Erosion



Figure 9 Final Result

1. Facial Feature Identification and Extraction

The facial features were detected after the application of the afore described processing algorithms. After the centroid, circularity, area, and perimeter of each object identified in the image had been calculated, those calculations were used to identify the facial features. A method for identification and extraction was developed based on the observed characteristics of the analyzed test images. As was the case in [10] information about the relative position of facial features was used to help validate the results of facial feature identification. That is to say that it is known that the left eye should be to the left of the right eye, the mouth should be below the eyes, the nose should be between the eyes and the mouth and the eyebrows should be above the eyes. The process of feature identification was broken down into two stages, first an attempt at extracting the eyes. Their relative positioning was used to help determine the mouth and nose.

In the analyzed results of the test images, it was observed that the circularity of the eyes for each image tended to be greater than .45 an also tended to be greater than all other identified objects. Additionally, it was observed that all objects that were desired to be identified had areas above 200 pixels. These observations were used in the first step of identifying the eyes. First, all irrelevant objects were eliminated from the image by setting the pixel value for all objects with an area of less than 200 to 0. Next the circularities of all the remaining objects we’re compared. If the objects circularity was greater than .5 then it was tentatively labeled as an eye. Since all the scans of the image were done from left to right, it could be assumed that the first object with an area greater than 200 and a circularity greater than .5 was most likely the left eye. This was not always the case but provided a good starting point to work from, in the event that the right eye was mistakenly identified as the left, logic later on in the identification process was added to correct for this possibility. Continuing through the logic, the next object with an area greater than 200 and a circularity greater than .5 was considered the right eye. At this point, using the center point calculations for both objects identified as eyes, a check was made to ensure that the left eye was positioned to the relative left of the right eye and that the right eye was positioned to the relative right of the left eye. If the eyes were in the incorrect position, their designation as left or right was switched.

Next, further relative position checked we’re applied to ensure that the eyes were not mistaken for eyebrows. Each of the remaining unidentified but extracted objects of interest was compared using their relative position to the eyes, circularity and area to identify the remaining facial features. For each object, if the circularity was between .2 and .5 it was labeled as a possible mouth or nose; this range was determined by observing the analyzed results of the test images and finding that circularities for the mouth and nose tended to be the next highest to the eyes of all identified objects, and that the circularities for the mouth and nose tended to fall between .2 and .5. As the possible mouth and nose objects were extracted, the known relative positioning of the nose and mouth were used to ensure that the identified mouth was below the identified nose. After iterating through all the non-eye objects, the object with a centroid location below the eyes with the largest area and circularity was identified as the mouth, the next highest area and circularity, with a row value for the centroid falling in between the eyes and the mouth was considered the nose. After the nose, possible eyes and mouth had been identified, the know relative positions of all objects on a face were used to validate that the position of all identified objects was reasonable. If it was the case that an eye was detected with a centroid column within +- 15 pixels of another eye, the eye with the lower centroid row value was re-labeled as the corresponding eyebrow. Any facial features that were not able to be identified and labeled at this point were noted. This process of applying mark8, then chain8 and then identifying and labeling facial features was repeated until at least two eyes and a mouth had been labeled. For each iteration, if two eyes and a mouth had not been labeled, a morphological erosion and then a dilation was applied to the intermediary image in hopes of increasing the next iterations ability to correctly identify and label facial features. This process was limited to a maximum of four iterations in order to protect against infinite looping at this stage. The results of this process can be seen in Fig. 9. In Fig. 9 the object number 223 is the left eye, object 219 is the right eye, object 211 is the nose and object 205 is the mouth. The left and right eyebrows were not individually identified but are both part of object 240. The identification process was able to properly identify the left eye, right eye, mouth and right eyebrow in this case as see in Table 1.

Finally, after all facial features for the image were identified and labeled, a log file was generated with each object for that image and the corresponding centroid, area, perimeter and circularity. These log files were then used to prepare the features that were used for classification. The results for the sample image from Fig. 4 are shown below in Table 1 and Table 2.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Left Eye** | **Right Eye** | **Nose** | **Mouth** | **Left Eyebrow** | **Right Eyebrow** |
| **Identified Value** | 223 | 219 | N/A | 205 | N/A | 240 |
| **Actual Value** | 223 | 219 | 211 | 205 | 240 | 240 |

Table Identified vs Actual Facial Features

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Area** | **Circularity** | **Center Row** | **Center**  **Column** |
| **Left Eye** | 657 | 0.725234 | 127 | 95 |
| **Right Eye** | 532 | 0.713549 | 127 | 159 |
| **Mouth** | 1347 | 0.358636 | 210 | 125 |
| **Right Eyebrow** | 3635 | 0.047362 | 121 | 145 |

Table Measurements for Correctly Identified Features

1. Emotion Classification

The image processing and feature analysis as described in the previous sections (Section 5 and 6) resulted in the identified expressive features, all of which to be utilized as predictive features for emotion classification. Due to the challenges that researchers have faced with classifying emotions from static images, it was of interest to see if the image processing algorithms and calculations will be able to produce accurate classifications. This section describes the emotion classification approach and the results using the expressive features extracted.

The classification approach was completed using the MATLAB classificationLearner tool, which allows for models be trained and tested using supervised learning algorithms [12]. Our research focused on the six universal emotions, therefore the six emotions (happiness, sadness, anger, fear, disgust, and surprise) were used as the classifiers. The specific extracted feature calculations included are the left eye area, right eye area, left eye circularity, right eye circularity, mouth area, the distance between the left eyebrow and left eye, and the distance between the right eyebrow and right eye. The distance between left/right eyebrows and eyes was calculated using Euclidean distance. The supervised learning algorithms that will be used to classify images based on the extracted features are Decision Tree, Linear Discriminant, k-Nearest Neighbor (kNN), and Support Vector Machine (SVM); all of which are popular algorithms used in research for classification purposes. The reason four algorithms were chosen for this research instead of just implementing one was for comparison purposes to see what algorithms performed the best in terms of classification accuracy.

The dataset was split into 70% for training the classification model, and 30% was used to test out the performance and accuracy of the recently trained classification model on emotion classification. The determination of the dataset split percentage was based on the size of the dataset, as an attempt for enough data to successfully train the model and enough to properly test with, using a limited size dataset. A dataset split of 60/40 and 80/20 was also performed, however, results did not indicate any significant changes or improvements. In addition, a 10 cross validation methodology was applied to help prevent the over fitting of training data, and to asses the performance of the training model before the model is used for testing with the split dataset. Table 3 shows the results produced from the classification algorithms.



Table Emotion Classification Accuracy (Train Set)

The results in Table 3 indicate that the trained models were not as successful at classifying the emotions based on the features. It was observed that the low accuracy rate is attributed to that some of the emotions have similar feature calculations depending on the expressions. The similar feature calculation between emotions is what may have caused the model to interpret the emotion to be something else (i.e. happy and surprise emotions were commonly mistaken with each other due to their high area calculations and distances between eyebrows and eyes). Due to the limited size of the dataset, we were only able to extract information from a small subset of individuals in society. If the dataset was larger or more diverse, accuracy may be improved as the features may differentiate more. To check to see if the reduction of features would improve accuracy, Principal Component Analysis (PCA) was used. However, it results that it did not affect the accuracy scores. Despite the low accuracy results, the trained models were still tested using the test dataset. Results are shown in Table 4. It is observed that the test classification results are similar to the accuracy scores of the model in the train set.



Table - Emotion Classification Accuracy (Test Set)

While the results of the emotion classification using the extracted expressive image are disappointment to the low accuracy, it opens the possibilities for trying different image processing techniques and classification techniques for future work.

1. Conclusion

This research focused on the detection and classification of the universal emotions, and it allowed us to become familiarized with the image processing and techniques for analysis that was introduced to us this semester. The results of our work, while not as successful as we would like at determining emotions based on the features that we selected, provide a foundation for attritional research. In subsequent studies we would make the following recommendations based on our experience. In the stage of feature extraction we recommend the following: First we would recommend applying a noise reduction filter as the second step after the Sobel edge detector is applied, we feel that this could better separate the distinct objects, reduce noise and thereby reduce the occurrences where the eyes and eyebrows overlap with either each other or other facial objects, causing the algorithm to misclassify the impacted feature. We would also suggest the exploration of other edge detection techniques, transforming into the frequency domain with a Fourier transform and then performing an optimized band pass filtering, as suggested in [11] for fingerprint detection, could provide more accurate edge detection. Further, in the area of feature identification we would recommend that additional logic be introduced to more consistently accurately identify the eyes, mouth, nose and eyebrows. Finally, in the area of emotion classification we would suggest using more complex features as was done in [5] rather than the simple calculations that we have used. Additionally, to investigate emotion misinterpretation and hopefully mitigate against it happening, the classifiers that can be tested in future research is of emotions that stand out against each other (i.e. just classification of happy vs anger or positive vs negative emotion). We feel that our current work lays a good foundation for further research to build on, both in the areas of facial feature identification as well as in emotion detection in facial images.

References

[1] Design, S. (n.d.). The Ekmans' Atlas of Emotion. Retrieved from <http://atlasofemotions.org/#introduction/>

[2] Ekman, P., & Keltner, D. (1970). Universal facial expressions of emotion. California mental health research digest, 8(4), 151-158.

[3] Berdahl, J. L., & Martorana, P. (2006). Effects of power on emotion and expression during a controversial group discussion. European journal of social psychology, 36(4), 497-509.

[4] Chen, L. S., Huang, T. S., Miyasato, T., & Nakatsu, R. (1998, April). Multimodal human emotion/expression recognition. In Automatic Face and Gesture Recognition, 1998. Proceedings. Third IEEE International Conference on (pp. 366-371). IEEE.

[5] Halder, R., Sengupta, S., Pal, A., Ghosh, S., & Kundu, D. (2016). Real Time Facial Emotion Recognition based on Image Processing and Machine Learning. International Journal of Computer Applications, 139(11).

[6] Chow, G., & Li, X. (1993). Towards a system for automatic facial feature detection. Pattern Recognition, 26(12), 1739-1755.

[7] Lyons, M. J., Budynek, J., & Akamatsu, S. (1999). Automatic classification of single facial images. IEEE transactions on pattern analysis and machine intelligence, 21(12), 1357-1362.

[8] Pal, P., Iyer, A. N., & Yantorno, R. E. (2006, May). Emotion detection from infant facial expressions and cries. In Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on (Vol. 2, pp. II-II). IEEE.

[9] http://www.kasrl.org/jaffe.html

[10]S. Ioannou, G. Caridakis, K. Karpouzis, and S. Kollias, “Robust feature detection for facial expression recognition,” *Eurasip J. Image Video Process.*, vol. 2007, no. May 2014, 2007.

[11]Scott E Umbaugh, *Digital image processing and analysis: human and computer vision applications with CVIPtools*. CRC press, 2010.

[12] https://www.mathworks.com/help/stats/classificationlearner-app.html