Determining Mood from Facial Expressions

CS 229 Project, Fall 2014

Matthew Wang
mmwang@stanford.edu

Spencer Yee
spencery@stanford.edu

I Introduction

Facial expressions play an extremely important role in human communication. As society continues to make greater use of human-machine interactions, it is important for machines to be able to interpret facial expressions in order to improve their authenticity. If machines can be trained to determine mood to a better extent than humans can, especially for more subtle moods, then this could be useful in fields such as counseling. This could also be useful for gauging reactions of large audiences in various contexts, such as political talks.

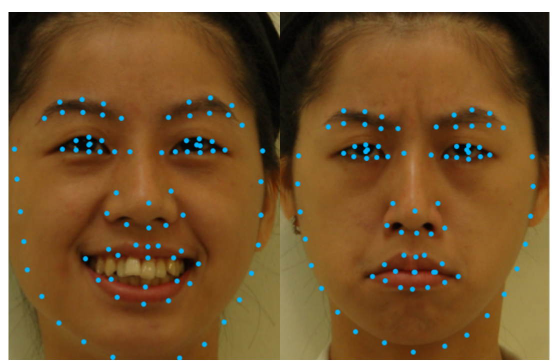
The results of this project could also be applied to recognizing other features of facial expressions, such as determining when people are purposefully suppressing emotions or lying. The ability to recognize different facial expressions could also improve technology that recognizes to whom specific faces belong. This could in turn be used to search a large number of pictures for a specific photo, which is becoming increasingly difficult, as storing photos digitally has been extremely common in the past decade. The possibilities are endless.

II Data and Features

2.1 Data

Our data consists of 1166 frontal images of people's faces from three databases, with each image labeled with one of eight emotions: anger, contempt, disgust, fear, happiness, neutral, sadness, and surprise. The TFEID [1], CK+ [2], and JAFFE [3] databases primarily consist of Taiwanese, Caucasian, and Japanese subjects, respectively. The TFEID and JAFFE images are both cropped with the faces centered. Each image has a subject posing with one of the emotions. The JAFFE database does not have any images for contempt.

Figure 1



Happiness Anger

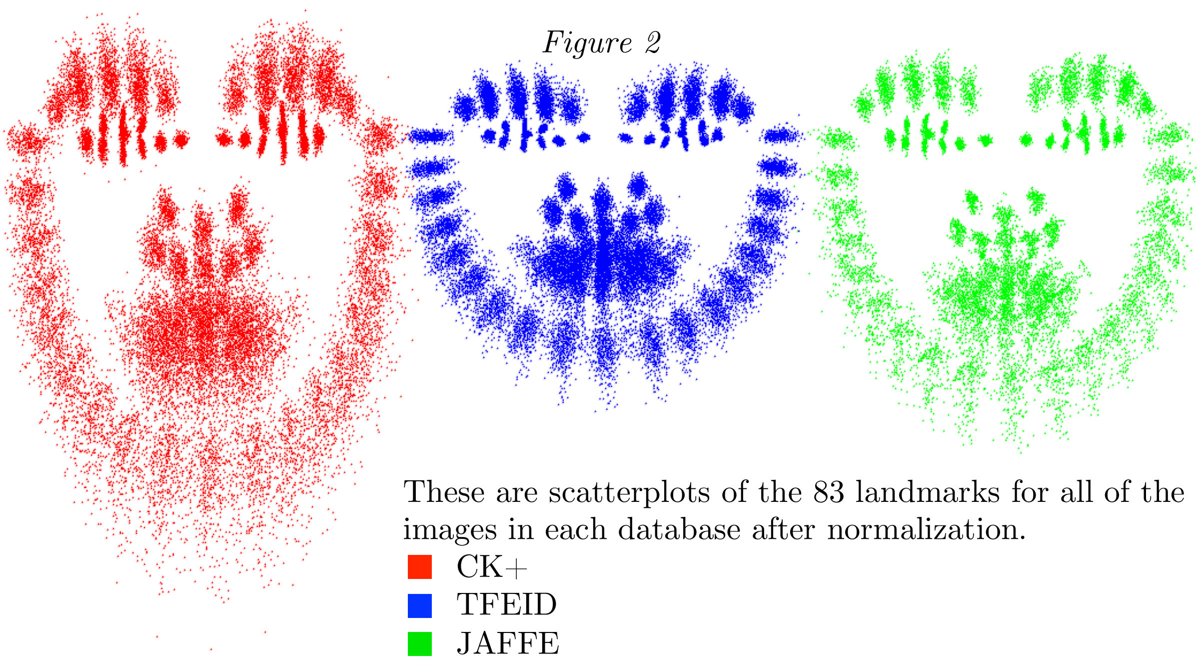
2.2 Features

On each face, there are many different facial landmarks. While some of these landmarks (pupil position, nose tip, and face contour) are not as indicative of emotion, others (eyebrow, mouth, and eye shape) are. To extract landmark data from images, we used

Face++ [4], a publicly available facial recognition API. Since Face++ face detection can only take a URL as a reference to the image, we used ImageShack [5] to host the images from our database. We used 167 of the features given by Face++, which include a smiling metric and the x- and y-coordinates of 83 facial landmarks, shown in Figure 1. Since each image can have differently sized faces at arbitrary locations within the image, we normalized each image as follows: we translated each face in order to center the eyes around the origin by using the x- and y-coordinates of the center of each eye, and then scaled each image to fix the distance between the centers of the eyes to a constant.

Upon normalizing and cross-testing databases to see how well a model trained on one database could classify another database, we realized that, since each database was homogenous in terms of race, faces from one database were consistently differently shaped than faces from another database (see Figure 2). Furthermore, using only positions of individual landmarks results in missing information, because the positions are not independent.

Figure 2



These are scatterplots of the 83 landmarks for all of the images in each database after normalization.

- CK+
- TFEID
- JAFFE

Therefore, from these facial landmarks, we derived 37 more features from angles between certain landmarks that we decided varied among emotions. For example, the angle formed by the two corners of an eyebrow and the center of the eyebrow shows the eyebrow's shape and how much it is raised. Likewise, the angle formed by the corners of the nose and the nose tip is a good measure of how much the nose is scrunched. Furthermore, using angles can scale the intensity of emotions more accurately than a simple difference in y-coordinates between certain landmarks. For example, the difference in y-coordinates for the mouth's landmarks between a neutral expression and a smile is similar to the difference in y-coordinates between a smile and a wider smile, but the former is clearly more significant. Thus, we calculated 37 angles that we thought would vary with emotion and used these as additional features, for a total of 204 features.

III Models

3.1 Softmax Regression

We used the MATLAB built-in function with a feature set of the 37 angles we selected, because our entire feature set was too high of a dimension to efficiently be calculated with softmax regression. The parameters of the model are those that maximize the log-likelihood function:

$$\sum_{i=1}^{m} \log \prod_{l=1}^{k} \left(\frac{e^{\theta_l^T x^{(i)}}}{\sum_{j=1}^{k} e^{\theta_j^T x^{(i)}}} \right)^{1\{y^{(i)}=l\}}$$

3.2 Multiclass Support Vector Machine

We used the LibSVM library [6] to train our data with a multiclass Support Vector Machine. Specifically, we used C-Support Vector Classification with a Gaussian radial basis kernel function. We experimented with different values for the parameters γ (in the kernel function) and C (the regularization parameter) and chose the values γ = 0.0005 and C = 2.5 for the kernel equation:

$$K(x_i, x_j) = e^{-\gamma \|x_i - x_j\|^2}, \gamma > 0$$

IV Results

We used all 1166 of our images to run tests. Accuracy is calculated as the percentage of images that were classified correctly, and precision and recall are calculated as the average of the precision and recall for each emotion.

Test	Softmax Regression			Multiclass SVM		
	Accuracy	Precision	Recall	Accuracy	Precision	Recall
Training	78.30%	76.20%	76.44%	98.89%	98.76%	98.88%
2-fold cross validation	66.21%	63.33%	63.35%	80.70%	79.43%	78.70%
5-fold cross validation	68.52%	66.34%	66.27%	84.56%	83.82%	83.09%

The SVM model clearly fit our data better than the softmax model did, so we chose this as our final classifier. Here are some more detailed results for the SVM 5-fold cross validation test:

Confusion Matrix		Predicted emotion							
		Anger	Contempt	Disgust	Fear	Happiness	Neutral	Sadness	Surprise
Actual emotion	Anger	120	2	6	1	0	11	4	2
	Contempt	0	82	3	2	4	12	4	1
	Disgust	11	1	146	3	1	3	2	2
	Fear	4	4	2	101	4	7	3	8
	Happiness	0	3	0	0	177	0	0	1
	Neutral	8	5	1	3	0	113	5	0
	Sadness	10	1	4	6	1	12	80	2
	Surprise	0	1	0	8	1	0	1	167

Emotion	Anger	Contempt	Disgust	Fear	Happiness	Neutral	Sadness	Surprise
Precision	78.43%	82.83%	90.12%	81.45%	94.15%	71.52%	80.81%	91.26%
Recall	82.19%	75.93%	86.39%	75.94%	97.79%	83.70%	68.97%	93.82%

As expected, happiness and surprise were more easily expressed and identified than the other emotions, probably because certain characteristics, such as a smile or a wide open mouth, are very distinguishable. Interestingly, sadness, quite a common emotion, ended up being misclassified most often. Looking at the confusion matrix, however, we can see that it usually gets confused for anger or neutral. This relationship also works the other way, with most improperly classified sadness images actually being anger or neutral. There are several other pairs that our classifier often gets confused. The most notable ones are anger and neutral, anger and disgust, contempt and neutral, and fear and surprise. This is not surprising, since the two emotions in each pair tend to produce similar facial features. For example, people commonly express both fear and surprise with lifted eyebrows, wide eyes, and an open mouth.

For the poster presentation, we developed a live demonstration in which people could take a picture in Photo Booth and have their emotion classified by our SVM model. Upon testing our algorithm on other people, we noticed that different people may express the same emotion in different ways, and that not all of these expressions were captured by our model. Some people are also much less expressive than others, and some are often unsure of how to express a certain emotion. In particular, most people were confused when asked to try contempt. Emotion can be very subjective in and of itself, so it is probably difficult to achieve a significantly higher accuracy. This is already observable in our databases, which tend to be homogenous within themselves. Some of the similarities among images within a database can be attributed to general facial features of particular races, but it is worth noting that all of the subjects in the TFEID database express contempt with a twist of the mouth to one side. Therefore, it is possible that the subjects in each database were told to consider certain facial expressions while simulating each emotion. Finally, it is important to keep in mind that facial expressions are more complicated than pure expressions of exactly one emotion.

V Future Work

As we refine our algorithm, it is important that we obtain access to a much larger and much more diverse database to make our model more robust to different people. Currently, we only have 1166 training examples from three different databases that tend to use models of the same race. Expanding this number to around ten or twenty thousand training examples could help the algorithm classify each emotion more accurately.

Furthermore, we could focus on pairs of emotions that tend to get confused and identify features that would specifically help distinguish between those emotions. This would lessen confusion between specific emotions and improve overall accuracy.

Another set of features that we could add is the orientation of each of the 37 angles on the face. For example, even though we know the angle between the corners of the eyebrows and the top of the eyebrow, we do not know how the eyebrow is positioned; it could be slanted outward to convey sadness or inward to convey anger. This would also help distinguish between different mouth and eye positions, both of which could affect classification. Taking this factor into account would give the algorithm a better representation of the face and help it differentiate between emotions.

It is also important that we begin to move away from using the 83 landmarks directly as features since position of features is largely dependent on the innate shape of the face. We could add an intermediate step between landmark identification and feature selection. Based on the landmarks that the face detection API gives us, we can train another model to identify certain facial structures and feed a representation of these facial structures to a multiclass SVM to classify emotion. For example, we could combine information about the position of landmarks of the nose along with the angle between some of these positions in order to determine whether or not the nose is scrunched.

It might also be worth investigating how mirroring a face affects the emotion classification. For emotions that are generally accompanied by asymmetric faces, it might help to normalize the faces so that the side with a certain characteristic, such as a wink, is always on the same side.

Lastly, as our algorithm currently only takes into account frontal images; rotating a face in any direction would render it ineffective. Taking into account facial rotation about all axes could help our algorithm identify emotions of rotated faces more accurately.

VI References

[1] Li-Fen Chen and Yu-Shiuan Yen. (2007). Taiwanese Facial Expression Image Database. Brain Mapping Laboratory, Institute of Brain Science, National Yang-Ming University, Taipei, Taiwan.
[2] Lucey, P., Cohn, J. F., Kanade, T., Saragih, J., Ambadar, Z., & Matthews, I. (2010). The Extended Cohn-Kanade Dataset (CK+): A complete expression dataset for action unit and emotion-specified expression. Proceedings of the Third International Workshop on CVPR for Human Communicative Behavior Analysis (CVPR4HB 2010), San Francisco, USA, 94-101.
[3] Michael J. Lyons, Shigeru Akemastu, Miyuki Kamachi, Jiro Gyoba. *Coding Facial Expressions with Gabor Wavelets*, 3rd IEEE International Conference on Automatic Face and Gesture Recognition, pp. 200-205 (1998).
[4] http://www.faceplusplus.com
[5] https://imageshack.com/
[6] Chih-Chung Chang and Chih-Jen Lin. (2001). *LIBSVM – A Library for Support Vector Machines*. National Taiwan University, Taipei, Taiwan. http://www.csie.ntu.edu.tw/~cjlin/libsvm/