Image Processing Pipeline for Facial Expression Recognition under Variable Lighting

Ralph Ma, Amr Mohamed
ralphma@stanford.edu, amr1@stanford.edu

Abstract

Much research has been done in the field of automated facial expression recognition because of the importance of facial expressions to understanding human interactions and emotions. While several systems have achieved positive results using either facial model based classification or feature based classification, most of these systems have been tested on subjects in constant lighting conditions. These systems may thus be susceptible to lighting changes since illumination contribute much more to image variation than facial features. In this report, we augment the BU-4DFE dataset by adding different lighting conditions to 3D images of subjects performing different facial expressions. Then we develop an image processing pipeline to rectify the effects of illumination on the images, hoping to preserve high classification rate even in harsh lighting conditions. Then we test our pipeline on two measurement: classification accuracy based on a LDA model and SIFT keypoint repeatability. For our results, we found that our image processing pipeline helped improve classification accuracy when performing LDA to identify images in dark lighting conditions. We did not find significant improvement in keypoint detection.

1. Introduction

Expression recognition is used in many different settings. Medical researchers use expression recognition device to provide therapy to children and adults with Autism. Companies and government can use expression detection to gauge response to products and changes. Thus research into this field can have tremendous impact. Many studies addressing this subject use images and datasets that are created in a lab with uniform lighting conditions[1]. This is understandable because it allows for accurate evaluation of the recognition algorithm. However, for most practical applications, the emotions recognition task is done in real-world conditions where the lighting is diverse and far from being uniform. In this study, we aim to study the effects of different lighting/shadowing on the emotions recognition task and find the best techniques to improve emotions recognition under variable lighting.

Illumination changes have huge effects on facial classification tasks. In fact, differences due to varying illumination can be much larger than difference due to varying emotions and even larger than differences among faces [2]. Thus it is important for expression classification systems to take lighting into account.

This poses a problem even to the most sophisticated computer vision techniques. For instance, convolutional neural networks are one of the most successful classification tools so it is natural to wonder why they are not being used to solve the problem of emotion detection. One of the main reasons is that only a small amount of labeled training data is available. Convolutional neural networks usually require a large amount of training data in order to avoid overfitting. A common technique is to train the network on a larger data set from a related domain. Once the network parameters have converged an additional training step is performed using the in-domain data to fine-tune the network weights. This allows convolutional networks to be successfully applied to problems that do not have large labeled datasets. For our purposes, however, unavailability of labeled data was a deterrent from using convolutional neural networks.

We can address the problem of lighting through preprocessing of our training/test sets or through feature engineering. We decide to focus on image preprocessing, because different classification systems vary widely in their feature selection techniques. However, all classification systems can incorporate a preprocessing module into their pipelines

without having to change any other modules.

2. Related Works

Most recent work on lighting normalization techniques for facial images has been focused on the problem of image recognition. These techniques range from illumination modeling that needs training datasets to universal image processing that can be performed without prior training. Because we lack a big dataset, we focus on image processing techniques.

Several techniques have been proposed to remove lighting or to equalize the effects of lighting. Wang [1] suggests the self-quotient face method which entails dividing an image by the low-pass filtered version of the same image. Wang achieved ¿90% recognition accuracy on the Yale B Dataset. Tan and Triggs [3] recommended a pipeline that performed dynamic range correction, difference of Gaussian filtering, and contrast equalization . Using their pipeline, the authors were able to improve a facial recognition system's accuracy from 41.6% to 79.0%.

Both of these works suggest that low-pass filters are effective in removing light which are usually encoded in the low frequencies of an image. However, removing too much data from the lower frequency may remove important shading from the face. Furthermore, high-pass filter can help eliminate noise and the effects of aliasing. However, high-pass filters can eliminate details and finer edges, such as those around the eye and lips. Those facial features may not be as important for facial recognition, which uses more face structure and feature locations. However, in expression recognition, the eyes and mouth region carry a huge amount of information.

3. Image Processing Pipeline

Our pipeline consists of four stages. We assume that we are given an image with the background cut out and only the face. For our training images, we use the Viola-Jones Haar Cascade Method implemented in OpenCV [4] to do so). We base our pipeline on the TanTriggs preprocessing [3] with adjustments suiting the facial expression recognition task. Our pipeline consists of gamma correction, selective filtering, and finally contrast equalization.

3.1. Gamma Adjustment

Both low lighting and high lighting conditions tend to wash out features of the face. We found that a gamma correction of $\gamma = 1.6$ is optimal to increase the dynamic range and thus, increase the contrast in the image. This highlights the edges on the face which can be useful for expression recognition.

3.2. Selective Filtering

Tan and Triggs apply a Difference of Gaussian filter in order to remove the illumination variation in low frequencies and noise/aliasing in the high frequencies. We found in experiments that the bandpass filter does not preserve well the details in the eye and mouth region. To counter against this problem, we sharpen both the eye and mouth before applying the bandpass filter. For the eyes, we apply the Viola-Jones technique [4] for eye detection and filter out false positives by only taking the two largest detected regions. For the mouth, Viola-Jones does not work very well. Instead we use the location heuristic that the mouth tends to be in the lower center region of the face. For a 200 by 200 pixel image, we take a 40 by 100 pixel long rectangle in the center bottom of the image. After finding the boxes that contain the eyes and mouth, we sharpen each region by the Unsharp Masking technique of subtracting low-pass filtered version of the images from the image to be sharpened. A Guassian filter with $\sigma = 1.2$ is used to create the low-pass filtered image. After sharpening only the eye and mouth, we apply a Difference of Gaussian filter to the entire image using $\sigma_1 = 1$ and $\sigma_2 = 2$. The Difference of Gaussian filter helps to remove noise and is also a good edge detector due to its similarity to the Laplacian Filter.

3.3. Contrast Equalization

Contrast equalization is done according to methodology suggested by Tan-Triggs. Contrast equalization is done in two stages. For a given image $I(x, y)$, we perform the following

$$I(x,y) \leftarrow \frac{I(x,y)}{(\text{mean}(|I(x',y')|^{\alpha}))^{1/\alpha}} \tag{1}$$

$$I(x,y) \leftarrow \frac{I(x,y)}{(\text{mean}(\min(\gamma, |I(x',y')|)^{\alpha}))^{1/\alpha}} \tag{2}$$

α prevents the mean from being affected negatively by large outlier values. γ is also a hard threshold used to truncate large values in the normalization.

Figure 1 and Figure 2 show results of preprocessing.

4. Methods

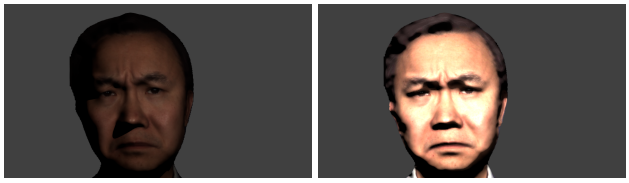
In this section, we discuss the data generation and preprocessing as well as the metrics we use to evaluate and improve our pipeline. We test our pipeline on two fronts. We measure improvements in restoring the SIFT keypoints that are lost after changing illumination as well as improvements in the classification accuracy under varying lighting conditions.

Figure 1. Face before preprocessing

Figure 2. Face after preprocessing

4.1. Data Generation

The starting point is the BU-4DFE dataset which contains 3D video data of 101 subjects (43M, 58F) performing six different emotions: Anger, Disgust, Sadness, Happiness, Fear, and Surprise[5]. For each subject and each video, we pick a frame representative of the subject performing the emotion. To create different lighting conditions, we load the frame into the graphics software Blender and augment the 3D image by placing a point light in the space with the 3D image and vary the illumination of the light. We placed the light 45 degrees away from the plane of the camera on the left side of the face and used 6 different energy settings for the point light (units dictated by Blender): 0, 1, 3, 5, 10, 15. In this project, we use the light intensity of 5 as the control illumination and other light intensities as test illuminations. We then project the frontal rendering of the face into a two dimensional image used for testing. Therefore, for 101 subjects with 6 different emotions and 6 different lighting conditions, we produced a dataset of 3636 images. Generated images are shown on top of this page



4.2. SIFT Keypoints

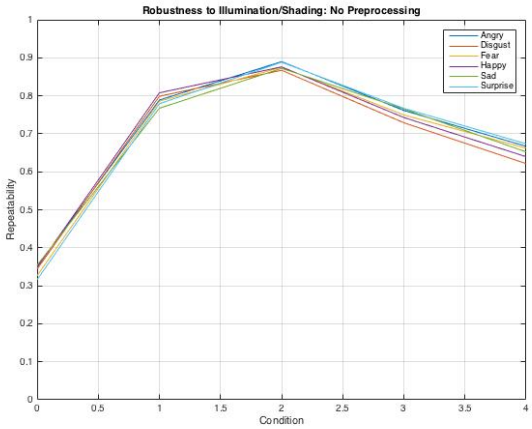


Figure 3. Repeatability of Sift Keypoints without preprocessing

We evaluate the robustness of SIFT keypoints detector to illumination changes. We compare the SIFT keypoints found for our control illumination to the SIFT keypoints found for our test illuminations. For every keypoint found in the images with control illumination, we take the location (x, y) of that keypoint and search for a keypoint in the image of a same subject with test illumination that has a location within an Euclidean distance of 2 to (x, y). If we find such a keypoint, we declare a match. For each comparison of test illumination to control illumination of the same subject, we measure the percentage of matches relative to the number of keypoints found in the image with control illumination. We are hoping to measure the repeatability of the SIFT keypoint detector in changing illumination conditions.

4.3. Fisherfaces

We evaluate the effects of preprocessing pipeline on a classifier that uses Fisherface. Fisherface is very sensitive to illumination changes, because lighting variations is much more significant than facial feature variations. We form a baseline by performing Fisherface on unprocessed images. Then, we measure improvements in the classification accuracy to multiple different light intensity after applying the processing pipeline to the images. Our initial classification produced poor accuracy which led us to investigate our dataset. After examination, we decided to ask human

subjects to do the classification task. Many human subjects misclassified anger, disgust, and fear. The human accuracy for the dataset is 0.53. We concluded that the dataset was not great for 6 class classification and decided to limit the classification to only 2 classes that are easier to classify: Happy and Sad.

For our classification task, we first divide the 101 subjects into a test set of 30 subjects and a training set of 71 subjects. We find the Fisherfaces by performing LDA on all training subjects with illumination 5. Then we find a threshold for the Fisherface projection score that best classifies between happy and sad subjects. Finally, we test the 30 subjects in all 6 illuminations using the model that is trained. The results are shown in Figure 4.

5. Results

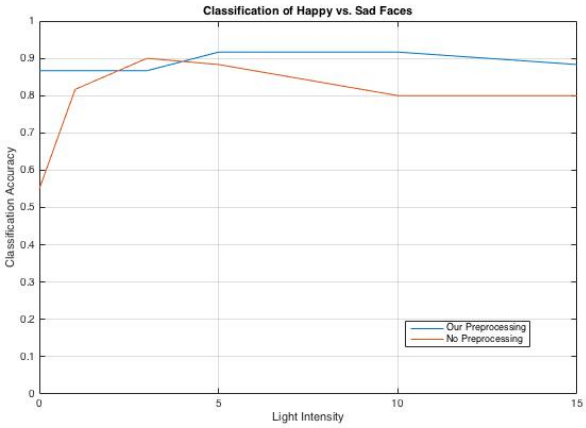


Figure 4. Results for FisherFace Classification

Figure 4 shows the the accuracy of Fisherface classifier for varying light intensities. We see that without preprocessing, Fisherface classifier loses accuracy when the light intensity is too low. At intensity of 0, no preprocessing Fisherface (np-Fisherface) achieves accuracy of .55 and FisherFace with preprocessing (p-Fisherface) achieves accuracy of .8667. Furthermore, we see that p-Fisherface achieves higher results in higher lighting conditions (10, 15).

Figure 5 shows the results for SIFT repeatability testing. Comparing Figure 3 with figure 5, we see that our preprocessing did not result in better Sift Keypoint detection.

6. Discussion

In order to potentially gain insights on how to design our processing pipeline, we examined the matched and missed SIFT keypoints. In figure 6, the green points are the matched SIFT keypoints of the control illumination and the red points are the missed keypoints. From figure 6, we

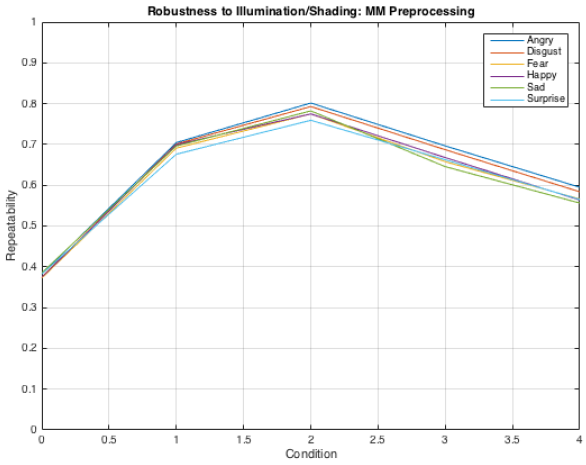


Figure 5. Results for FisherFace Classification

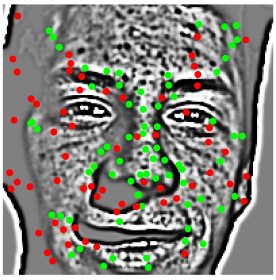


Figure 6. SIFT Keypoints

can see that all the keypoints that are on the left hand side of the face are missed. This is because the shadow created by the angle of the light caused a loss in some facial features. This explains the low accuracy in emotion classification under dim lighting conditions. This directed us towards improving the gamma correction portion of our processing pipeline as well as sharpening the eyes and mouth because they tell us the most about emotions and are rarely affected by extreme lighting/shadows.

Our preprocessing technique improved the results for classification based on Fisherfaces. The Fisherface used for classification is shown in Figure 7 and Figure 8. Figure 7 captures the lip shape of a smile and has cheek lines formed by the lips pushing upwards. Furthermore, we see the eyes are smaller and slanted upward. Both the eyes and mouth are features that are sharpened in our image preprocessing.

Figure 7. Fisherface with preprocessing

Figure 8. Fisherface without preprocessing

7. Conclusion

We were able to improve the classification accuracy for dim lighting conditions by studying the missed keypoints and missed features and adding preprocessing modules to fix them. A future extension of this project would be to extend the improvements to a range of complex human emotions, such as stress, wonder, pride, etc. Another future extension that can be done is to improve the robustness of the preprocessing module by including more varied lighting and shadows from different angles and from more than one source to create complex lighting conditions.

8. Work Division

Ralph Ma: Simulated lighting conditions in blender, Implemented Fisherfaces, Incorporated Tan Triggs

Amr Mohamed: Implemented SIFT Keypoints detector, Tuned processing parameters

9. Acknowledgements

We like to thank the fall 2015 staff of EE368. We also want to thank Azar Fazel for helping us with Blender.

References

[1] Yangsheng Wang Haitao Wang, Stan Z Li. Face recognition under varying lighting conditions using self quotient image. *Proceedings of the Sixth IEEE International Conference on Automatic Face and Gesture Recognition*, 2004.

[2] Beat Fasel and Juergen Luettin. Automatic facial expression analysis: A survey. *Pattern Recognition*, 36(1), 1999.

[3] Xiaoyang Tan and B. Triggs. Enhanced local texture feature sets for face recognition under difficult lighting conditions. *IEEE Transactions on Image Processing*, 19(6).

[4] Michael Jones Paul Viola. Rapid object detection using a boosted cascade of simple features. *Computer Vision and Pattern Recognition*, 2001.

[5] Lijun Yin; Xiaozhou Wei; Yi Sun; Jun Wang; Matthew J. Rosato. A 3d facial expression database for facial behavior research. *7th International Conference on Automatic Face and Gesture Recognition*, 2006.