



## Eyeglass Localization for Low Resolution Images

Earl Arvin Calapatia<sup>1</sup>

<sup>1</sup>*De La Salle University*

<sup>1</sup>*earl\_calapatia@dlsu.ph*

**Abstract:** Facial data is a necessity in facial image processing technologies. In the context of facial expression recognition, there are a number of approaches taken to extract the features of the face during the pre-processing phase of the facial data. However, there are instances when there are objects and artifacts that compromise the reliability and usability of data such as eyeglasses. Active shape models and active appearance models are commonly used in facial recognition and biomedical image segmentation. Active shape models and active appearance models are trained with webcam images to detect eyeglass objects in facial data to be used for preprocessing such as the removal of eyeglasses. The basic approaches are followed where a set of images with varying lighting and angles is marked with landmark points and then used as training sets for Principal Component Analysis. For efficiency and minimization of unnecessary processing, a new feature was added, the nose. Comparison of the two approaches is based on the accuracy of unseen test data, the performance with respect to the output shape and which approach performed better with fewer features. The mean square error (MSE) over the results of the proposed methods is used to evaluate the performance. The performance with respect to output shape is measured through the mean square error of the predicted localization to the actual contours of the test set. The active appearance model performed well with satisfactory results and improved speed. While training sets that included additional features had better performance in use compared with eyeglass only sets. The addition of the nose proved to have contributed to the effectiveness of the localization method.

**Key Words:** Image reconstruction; Face feature Extraction; Object Localization; Active Shape model; Active Appearance Model

### 1. INTRODUCTION

Faces play an important role in imparting information. Some forms of communication are done through expressions and micro-expressions. Faces exhibit characteristics that allow computers to see what a person is feeling (Wu, Liu, Shum, Xy & Zhang, 2004). The face holds a vast amount of data given the correct approach to collect and process it. Many techniques in image processing and machine learning have already been applied to extract data and features from the face. However, facial data can

be compromised through various obstructions such as eyeglasses.

### 2. PREVIOUS WORK

There are already efforts to handle occlusions found on the face. The most obtrusive are the eyeglasses. Several researchers (Vijayanandh & Balakrishnan, 2010; Wu et al., 2004) have done research on finding unobtrusive ways to remove glasses through the use of image processing. The common framework is first, the glasses are



automatically detected, the localization step employs a model will locate the exact points on the glass frames, the frames are removed and replaced with a new texture. The work of Jiang & Chen (2008) addressed several facial reconstruction tasks in their work (Wu et al., 2004). These include, red eye removal and eyeglass removal. In their set up, several cameras take in the facial data. A classification tool detects eyeglasses and maps its location on the images. The contour of the glasses is extracted with the use of a 3D deformable model. The 3D representation provides more features for a more effective localization. However, this 3D feature is traded off with significantly high computational effort. The general framework in the work of Vijayanandh et al. (2010) is similar, albeit with some differences. For detection and classification, a training sample is aligned by triangulation and compared to samples that have glasses and samples that don't. Active shape models and deformable contour models were set to localize the glasses. Sample based reconstruction was used to patch the frame regions (Vijayanandh et al., 2010).

### 3. RESEARCH PROBLEM

Little is mentioned in the literature with regards to the actual implementation of the eyeglass localization. Thus, this research is also an attempt to replicate and improve upon the previous works. As an extension to the pervious works, this research is conducted to address eyeglass localization for images with low resolution and handle them in real time. There are several limitations to such a goal including computational resources. However, the hypothesis of this research is that the goal can be achieved if the appropriate features are selected and used with the appropriate method. Also, because the dataset is low on instance and resolution, computation will be light enough that it will allow the method to process in real-time.

### 4. METHODOLOGY

As described in the paper of Cootes, Edwards & Taylor (2001), training images are manually labeled to determine their contours. The mean shape is taken and the data undergo Principal Component Analysis. Then a test set composed of unseen facial data will be used to localize eyeglasses. Performance is evaluated through measuring the

mean square error between predicted localizations and manually labeled contours on the test set. Performance with respect to computational efficiency is evaluated by performance over iterations towards convergence.

#### 4.1 Active Shape Model

The active shape model is a technique in model-based vision that makes use of manually labeled objects in training data to learn how to localize an object of interest in unseen data (Cootes, Taylor, Cooper, & Graham, 1995). Active shape models are distinct from active contour models, having contours that are only as deformable as imposed by the limits of the contours in the training set. Similar to active contours, energy minimization is the main method of search done by the active shape models. A shape is considered to have minimal energy when the boundaries lie on lines and edges. During search, the mean shape will deform to search for the best fit on the edges until convergence.

#### 4.2 Active Appearance Model

Active appearance models differ from active shape models with respect that they are not edge dependent but appearance and shape dependent instead (Cootes et al., 2001). Active shape models begin process by aligning the features of the image through the use of the mean shape through triangulated warping. For the implementation of this research, Delaunay triangulation was used. After alignment, shape models and appearance models are made through PCA. Shape and appearance models are computable with the functions:

$$x = x' + P_s b_s, \quad g = g' + P_g b_g \quad (\text{Eq. 1})$$

where:

- $x$ , = a vector representing an instance of a shape
- $x'$  = mean shape
- $P_s$  = the orthogonal set or principal components
- $b_s$  = serves as parameter for  $P_s$ .
- $g$  = instance of a transformed image.
- $g'$  = mean of the transformed images

The models are combined to consider the correlation between appearance and shape. The parameter for shape,  $b_s$ , the parameter for

appearance,  $b_g$  and a set of weights  $W$  are concatenated in a vector  $b$ .

The final equation can be derived forming the final equations:

$$x = x' + P_s W_s Q_s c, \quad g = g' + P_g Q_g c \quad (\text{Eq. 2})$$

where:

$Q$  = shape appearance eigenvector set.

$W$  = set of weights

The variable  $x$  is an instance of a shape transformed by the eigenvectors in  $P$  and  $Q$  with parameter  $c$ . The variable  $g$  is an instance of an image transformed from the eigenvectors.

Triangulating the image  $g$  with shape  $x$  will yield the necessary image for comparison with the unseen image. It is this principle of combining shape and appearance that contributes to the effectiveness of the active appearance model.

### 4.3 Dataset

9 images from 3 people (3 images each) are used for the training set. 2 of the volunteers are male and 1 is female. All volunteers are facing in differing directions for each instance. Images are taken at different lighting conditions and low-resolution web cameras, with a resolution of 340 x 230 pixels. 7 images are used for testing. 4 images come from another set of volunteers and the other 3 come from random images of people from the Internet. All images are manually labeled with an eyeglass only contour and an eyeglass with nose contour.

### 4.4 Features

Vijayanandh et al. (2010) opted to use all the features such as the lip line and the cheeks, however, features should be minimized since additional area to the test set will put a strain on computation. Several features were considered besides the glasses such as nose, mouth and ears. Because this research is founded on the premise that the method will serve as a preprocessing step to facial expression recognition, it became significant that the feature to be selected would have little variation in movement and will have to be in close proximity to the eyeglass. Hence, it is conjectured that the nose is the most suitable for the task because it does not move and it has distinctive

features such as highlight on the tip and darkening gradients on the edges (Cootes et al., 1995).

### 4.5 Evaluation

Three aspects are considered in the comparative evaluation. First is if the localization was at least able to determine the proximity of the frames. Second is evaluating which among the methods had the best localization result. Third is to prove whether the nose as a feature, contributes to the localization. The first evaluation will use a qualitative approach, where the images will be observed. The second evaluation will make use of the testing set's manual labels and extract the "correct" patch within the image and perform mean square error.

## 5. RESULTS AND ANALYSIS

Any further experimentation with active shape models were discontinued in favor of active appearance models due to its ineffectiveness despite being trained and tested with the same data.



Fig 1. Frontal, Eyeglass Only. Manual (Green), Localized (Blue and Red)



Fig 2. Oblique, Eyeglass Only. Manual (Green), Localized (Blue and Red)

Based on qualitative observation, the eye glass only set was not able to align properly with the frames, however, they are still in close proximity to their ideal locations. Table 1 shows the MSE per image from the two test sets. Set 1 refers to images with the nose feature while Set 2 refers to the images with glasses only. Figure 3 shows a plotted version of the MSE per image.

Table 1. MSE after convergence

Set	im1	im2	im3	im4	im5	im6	im7
Set1	.03	.04	.03	.03	.22	.04	.01
Set2	.09	.09	.06	.08	.07	.05	.07

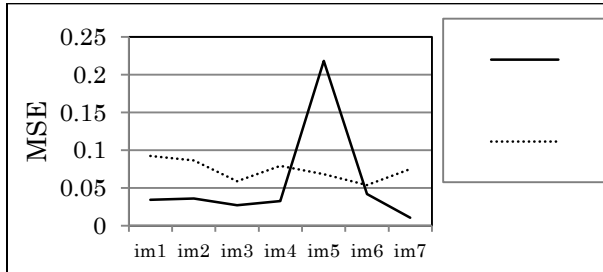


Fig 3. MSE after convergence

For the performance of localization, the most of the sets with the nose performed better compared to the items belonging to the no-nose set. Also, the with nose set was capable of localizing glasses that are tilted and adjust the contour according to the angle of the face.



Fig 4. Oblique, With Nose. Manual(Green), Localized (Blue and Red)



Fig 5. Frontal, With Nose. Manual(Green), Localized (Blue and Red)

This confirms the prior analysis that the nose can be an effective feature in localization. However, one of the testing instances scored very poorly. Compared to the rest of the data, the anomalous instance had large glasses and a very thin nose. This is in stark contrast to the rest of the set which had moderately sized glasses and wide noses. This exposes the limitation of the current model since there are no instances of training data that can accommodate the glass to feature proportions of the image.

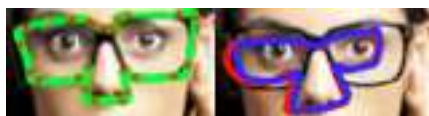


Fig 6. The Anomalous Image. Manual(Green), Localized (Blue and Red)

For speed and computational efficiency, the webcam only set was recorded to show simulation of actual use. It has been found that the set with the nose converges rapidly compared to the eyeglass only set. This leads to the conjecture that the added feature of the nose also aids in speeding up the search for the error minima. Figure 7 is a plot showing the error history from the first iteration towards convergence.

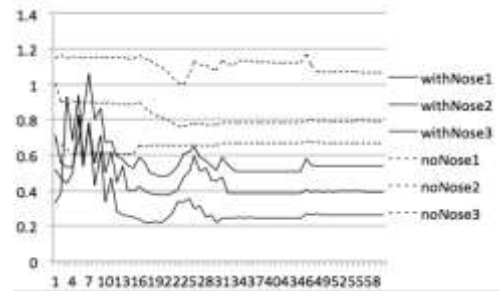


Fig 7. Convergence history

However, the number of frames it took to process reached up to 61 and higher. While the proposed method is relatively fast, it may still not be fast enough for applications that require immediate response. However, it can still be used by applications that run in real time but don't necessarily require it with a rate of the millisecond scale.

## 6. CONCLUSIONS

A number of features were considered for the experiment until the nose was identified as the most suitable feature for the localization of eyeglasses in low-resolution images. The nose feature effectively improved the speed and performance of the localization method. Despite the small dataset, the localization method was still able to perform well on a range of sets in varying lighting conditions and in low resolution. Another factor attributed to the success of the feature is the use of the active appearance model rather than the usual active shape model. The active appearance model was able to maximize the use of the appearance of the features and gave satisfactory results on convergence



## 7. RECOMMENDATIONS

The dataset is small compared to the ones used in literature and it would do well to take into account larger data sets and test for robustness over a wider span of conditions. Other factors that should be considered are shape, frame thickness, eyeglass to feature ratios and whether the glasses are full rimmed. Despite the role of the nose feature as aid to the speed and accuracy, the speed is still not enough to support fast response systems and will need further optimization. Future research can also extend this work towards complete eyeglass removal for low-resolution images.

## 8. ACKNOWLEDGMENTS

Gratitude to Sir, Macario Cordel for giving advice and attentive guidance in the endeavor for this research. Also to Dirk-Jan Kroon, who authored the open source implementation of Active Shape Models and Active Appearance Models in Matlab.

## 9. REFERENCES

Cootes, T. F., Taylor, C. J., Cooper, D. H., & Graham, J. (1995). Active shape models-their training and

application. *Computer vision and image understanding*, 61(1), 38-59.

Cootes, T. F., Edwards, G. J., & Taylor, C. J. (2001). Active appearance models. *IEEE Transactions on pattern analysis and machine intelligence*, 23(6), 681-685.

Vijayanandh, R., & Balakrishnan, G. (2010, December). Human face detection using color spaces and region property measures. In *Control Automation Robotics & Vision (ICARCV), 2010 11th International Conference on* (pp. 1605-1610). IEEE.

Wu, C., Liu, C., Shum, H. Y., Xy, Y. Q., & Zhang, Z. (2004). Automatic eyeglasses removal from face images. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 26(3), 322-336.

Jiang, X., & Chen, Y. F. (2008). Facial image processing. In *Applied Pattern Recognition* (pp. 29-48). Springer Berlin Heidelberg.