A Method for Converting a Smiling Face to a Neutral Face with Applications to Face Recognition

Mahesh Ramachandran, Shaohua Kevin Zhou, Divya Jhalani and Rama Chellappa
Centre for Automation Research
University of Maryland College Park
College Park, Maryland, 20742

Abstract

The human face displays a variety of expressions like smile, sorrow, surprise etc. All these expressions constitute non-rigid motions of various features of the face. These expressions lead to a significant change in the appearance of a facial image which leads to a drop in the recognition accuracy of a face-recognition system trained with neutral faces. There are other factors like pose and illumination which also lead to performance drops. Researchers have proposed methods to tackle the effects of pose and illumination; however, there has been little work on how to tackle expressions.

In this work, we attempt to address the issue of expression invariant face-recognition. We present pre-processing steps for converting a smiling face to a neutral face. We expect that this would in turn make the vector in the feature space go closer to the correct vector in the gallery, since we are performing appearance-based face recognition. This conjecture is supported by our recognition results which demonstrate that indeed the accuracy goes up if we include the expression-normalization block.

1. Introduction

Pose, illumination and expression variations are three important issues to be dealt with in the area of face-recognition. An image of a person taken under a different state of pose, lighting and expression than the one in the gallery is likely to be classified wrongly because of the increase in distance (measured with an appropriate metric) from the image in the gallery corresponding to the correct identity. Therefore, for robust face-recognition, it is imperative that we compensate for the effects of the three issues mentioned above.

Researchers have suggested methods to tackle the effects of pose and illumination [1, 2]. In this paper, we attempt to study the effects of expressions on face-recognition performance and also try to compensate for these expressions.

2. Motivation

In order to robustly tackle the issues mentioned above, the best way would be to somehow separate the identity from the expression/lighting condition in the facial image. However, we do not have any expression-invariant metrics for face-recognition. So we will try to develop some pre-processing tools to convert faces with expression into neutral faces before using them for recognition. We expect this to bring the resulting image closer to the image in the gallery of faces with neutral expressions.

We note that such processing will help only to a certain extent; robust neutralization of expression calls for a generative 3D model of the face that models the geometric and textural changes corresponding to the various expressions accurately. This means that we will be able to correct for the changes in geometric location of the various facial features and we will also be able to remove artifacts. However, subtle expressions involving outward projection of the cheek, pointed nose etc are difficult to correct using a single image.

3. Previous work

For many decades now, researchers have been trying to understand and characterize facial expression. One of the early works in this area has been the attempt by Ekman [6] to describe facial expressions in terms of basic motions of various parts of the face such as lips, eyebrows etc. Later, this idea was formalized in terms of the Facial Animation Parameters (FAP) in the MPEG-4 Facial Animation Coding System.

There has been a lot of work in the areas of expression recognition, synthesis of expression sequences etc. For classifying expression in video sequences, Yaser and Davis [3] have used optical flow-field vectors as feature points for a radial basis network classifier.

Works on generation of synthetic images with facial expression have traditionally relied on building 3D facial models of the person using a set of images under different poses. After accurate matching of a 3D morphable model to a face, Vetter et al. [4] have changed the weights of some 'expression-eigenvectors' to check if this leads to a change in expression on the synthesized image. However, there is no strong justification for a linear model for expressions. There is little work on a generic conversion technique to convert a single image into the neutral expression for the purpose of aiding recognition.

4. Problem Statement

Assume that we have a still-image face-recognition system trained with images of people having a neutral expression, with frontal pose, and with similar illumination conditions. Assume that the recognition system is based on one of the existing classifiers like PCA, LDA or EGBM [9]. Such classifiers work reasonably well when images of the people with same illumination/pose/expression conditions are presented to it. Our goal is to boost the recognition-accuracy on faces with expression by including an expression-normalization block.

The expressions that lead to the greatest change in the appearance of a face are the those that involve the motions of the lips and jaws. Therefore, we consider only those expressions which involve different configurations of the mouth such as in smile, surprise, etc. We do not consider other subtle expressions involving tightening of nose muscles because (i) they do not affect appearances so much and (ii) they are difficult to correct without an accurate 3D model.

5 Mesh based transformation

Our approach to expression normalization would be to change the appearance of a face using a triangular mesh model of generic human face. We used CANDIDE [7], a generic face model (Refer Fig. 1) with support for the various basic motions of the facial features. In this work, we considered the following motions: (i) upper lip raiser, (ii) jaw dropper, (iii) lip stretcher, (iv) lip corner depresser and (v) lip presser.

5.1 Global affine warp

We register the mesh on the facial image accurately such that the various features of the mesh coincide with the features in the image. Active appearance models [8] enable automatic registration by attempting to find the optimal shape and appearance parameters so that the model best describes the face.

Once the mesh is registered to the given image, we change the geometry of the mesh to bring it to the neutral state. In this process, we will also need to do the texture mapping from the original mesh to the new mesh. We tackle

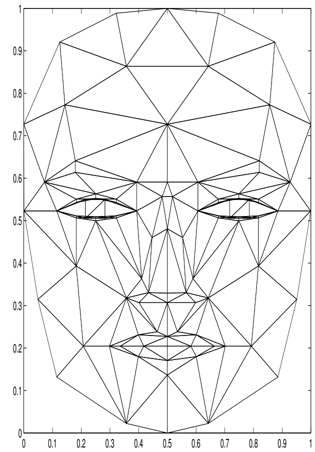

Figure 1: CANDIDE - the generic face model

this problem by doing a piecewise affine warping of the texture. The transformation of the texture from a face with expression to a neutral image is definitely more complicated than an affine warping. However, for expressions involving simple displacements of various facial features, we can assume that illumination of moving points do not change too much. So an affine mapping is a good approximation.

5.2 Correction of local artifacts

The resulting face will have all its features at the correct locations but there will be some artifacts because the affine warping will preserve the wrinkles due to the facial expression. We need to remove these artifacts for the images to look realistic and increase the recognition accuracy.

Since the mesh is registered to the face, we know the locations in the image where we expect to find wrinkles. We observe that most of the wrinkles can be modeled by a smoothly varying texture surface that has prominent peaks or valleys at the position of the artifacts. This is illustrated in Fig. 2, where we plot the texture as a 2D surface. Note the valley corresponding to the prominent wrinkle in the facial image.

If we think of such textures as thin plates, we observe that wrinkles cause a lot of bending of these plates. We would expect that if we remove the valleys in such surfaces and interpolate them smoothly, this will remove the wrinkle from the image while still preserving the exact variation of the texture near the wrinkle. The next question is how to model the surface and how to remove the wrinkle. We use the idea of thin-plate splines to model the texture variations in the area near a wrinkle.

The steps of the algorithm are listed below. We use the term 'surface' to refer to the texture surface in a small portion of an image.

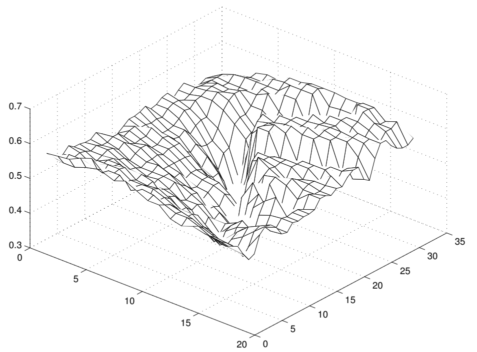

Figure 2: Texture surface near a wrinkle

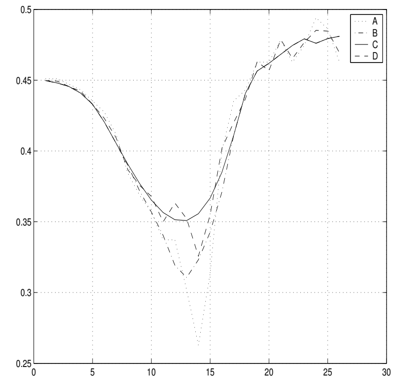

Figure 3: Profile of the texture near a wrinkle. (A) shows the original profile S, (B) shows the profile after averaging S_{avg}, (C) shows the profile after wrinkle removal $S_{smoothrec}$ and (D) shows the profile S_{newtex} after the smooth variations have been added

1. We register the triangular mesh to the given image and move the triangles so as to change the configuration of the face. Then we generate the new image corresponding to the new configuration by means of a piecewise affine warp.

2. For areas of the image where there are wrinkles, we approximate the texture surface S by means of a sum of functions of the form $a_1 + xa_x + ya_y + \sum_{i=1}^{p} w_i U(|P_i - (x,y)|)$ where $U(r) = -r^2 log(r^2)$ is as defined in [5]. First we use an averaging filter to smoothen the surface S and get S_{avg}. We also choose a threshold for approximation.

3. We find the error between the actual surface S_{avg} and the approximated (using thin-plate splines) surface S_{rec}. To the list of points P_i currently used to approximate the surface, we add the point giving the maximum error in reconstruction. We initialize this process by choosing a random set of 5 points for approximation.

4. We terminate the above process when the maximum error of reconstruction is within the threshold. Denote the set of points that reconstruct the surface S_{avg} by the set A.

5. We now find the edges in the portion of the image that we are analyzing. We expect to see prominant edges if there are any peaks/valleys in the surface. If the region is relatively smooth, we will (i) be able to approximate the intended surface by lesser number of points and (ii) we will not see prominent edge patterns. We compute the Euclidean distance transform corresponding to this edge-map.

6. To unconstrain the surface from the peak or valley, we adopt the following approach: For each point in the set A, look up the distance transform (DT) calculated above. If this is below a threshold, then it means the point is near the wrinkle. Therefore we would like to release the texture constraint on this. Hence, move only those points whose DT is above the threshold into a set B.

7. Reconstruct the surface using the set of points in B to get $S_{smoothrec}$.

8. Add $S - S_{avg}$ to $S_{smoothrec}$ to get S_{newtex}. This step is because we want to preserve the fine variations in texture. Recall that we ignored these fine variations when we averaged S to get S_{avg}.

9. Assign the texture of the points in S to S_{newtex}.

10. Repeat the above process for every region in the face where we expect to find wrinkles.

The nice thing about this method is that we are efficiently able to preserve the texture variation surrounding a wrinkle while smoothly interpolating over the wrinkle. This leads to a visually pleasing variation of the texture without unwanted edges. In addition, we are not drastically changing the texture variation in other regions of the face thereby preserving identity.

To illustrate the algorithm, we include profiles of the texture variation at various stages of the algorithm in Fig. 3.

6. Results

We now give the results of our experiments on evaluation of recognition accuracy. We used the images in the FERET database that were in the frontal pose and with uniform illumination. The database contains neutral and smiling faces satisfying the above requirement. In addition, we looked at

Figure 4: Examples of expression normalized faces. The first column has neutral faces (ground truth), the second has smiling faces and the third column has neutralized faces obtained from the smiling faces

the JAFFE database which contains images of 10 subjects under different expressions.

For face recognition, we implemented a PCA+LDA [10] classifier. We tested for recognition by randomly picking a subset of the set of images and using them in the recognizer. We found that when faces with different expressions related to the mouth were fed into the classifier as such, recognition accuracy was 67.5%. But when the faces were expression normalized, the recognition accuracy jumped to 73.8% indicating that expression normalization indeed helps in making the face vector move closer to the image in the gallery corresponding to the correct identity. This is also expected since we are doing appearance-based face-recognition; so if two images look almost similar, it is expected that they will be grouped together by appearance-based classifiers.

For the images in the JAFFE database (which contains 10 subjects), we obtained a recognition accuracy of 12/20 for the surprise images and 13/20 for the corresponding neutralized images.

In Fig. 4, we have included some images that have been converted into neutral expression. The visual quality of these images looks reasonably good, better than the results of the statistical learning technique that we implemented.

7. Summary and Conclusions

In summary, we have demonstrated that accuracy of a face-recognition system can be improved by converting the facial expressions to neutral expressions before feeding it to a face-recognition system. We have developed a simple technique that will normalize a face with expressions, involving simple displacements of the various facial features. We have obtained neutral faces with good visual quality as well as higher recognition accuracy.

We also note that assuming linear models and using statistical learning techniques may not be the best way to deal with expressions; they are definitely not the most intuitive.

To handle more complex expressions calls for an accurate generative 3D model of the face that models the exact texture mapping from one expression to another.

Acknowledgments

References

[1] S. Zhou, R. Chellappa and D. Jacobs, "Characterization of human faces under illumination variations using rank, integrability, and symmetry constraints," *European Conference on Computer Vision (ECCV)*, May 2004.

[2] Zhou and Chellappa, "Illuminating light field: Image-based face recognition across illuminations and poses," *IEEE International Conference on Automatic Face and Gesture Recognition (FG)*, May 2004.

[3] Mark Rosenblum, Yaser Yacoob and Larry Davis, "Human Expression Recognition from Motion using a Radial Basis Function Network Architecture," *IEEE Transations on Neural Networks*, Vol. 7, No. 5, September 1996.

[4] V. Blanz, C. Basso, T. Poggio, T. Vetter, "Reanimating Faces in Images and Video," *Proceedings of EUROGRAPHICS 2003*, Granada, Spain, 2003.

[5] Fred L. Bookstein, "Principal warps: Thin-Plate Splines and the Decomposition of Deformations," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 11, No. 6, June 1989.

[6] P. Ekman and W. Friesen, "Facial Action Coding System," *Consulting Psychologists Press Inc.*, 577 College Avenue, Palo Alto, California, 94306, 1978.

[7] M. Rydfalk, "CANDIDE, a parameterized face, Report No. LiTH-ISY-I-866," *Consulting Psychologists Press Inc.*, Dept. of Electrical Engineering, Linkping University, Sweden, 1987

[8] T.F.Cootes, G.J. Edwards and C.J.Taylor, "Active Appearance Models," *Proc. European Conference on Computer Vision 1998*, Vol. 2, pp. 484-498, Springer, 1998.

[9] W. Zhao, R. Chellappa, P. J. Phillips and A. Rosenfeld, "Face Recognition: A Literature Survey," *ACM Computing Surveys*, Vol. 35, No. 4, December 2003, pp. 399-458.

[10] Wenyi Zhao, Arvindh Krishnaswamy, Rama Chellappa, Daniel L. Swets, John Weng, "Discriminant Analysis of Principal Components for Face Recognition," *Proc. of Intl. Conf. on Automatic Face and Gesture Recognition*, pp. 14-16, 1998.