Three-dimensional Face Shape by Local Features Prediction

Rubén García-Zurdo *rubengarcia@cu-cisneros.es*
Psychology Division
Colegio Universitario "Cardenal Cisneros"
Madrid, 28006, Spain

Abstract

A method is presented to estimate the 3D face shape from a frontal image using a multivariate linear regression model between intensity and depth features: block based discrete cosine transform (DCT), block based principal component analysis (PCA) and modular PCA (eigenfeatures) coefficients. After noting that between-illumination coefficients variances are smaller than between-subjects coefficients variances, we try to correct illumination variations by discarding the first coefficient in the DCT method but not in the PCA methods, as the between-illumination coefficients variances are distributed over many of them. The lowest fractional error in depth prediction is obtained by using a low number of coefficients and a high overlap degree between blocks. Modular PCA produces the best results when the test image is frontally illuminated as in the training phase. DCT and local PCA are more robust across point source horizontal angle and ambient illumination variations.

Keywords: Shape from Shading, Three-dimensional Face Shape, Eigenfeatures, Multivariate Regression.

1. INTRODUCTION

Statistical shape from shading aims to detect regularities between global variations in intensity and depth images, usually with faces, the goal being able to build a 3D face shape from its intensity image. Principal component analysis (PCA) is used [1] as a dimensionality reduction technique over a 3D head sample along with an optimization method to predict the coefficients that allow reconstruction of a 3D test image shape. In [2] PCA is applied separately to shape and texture after applying a dense one-to-one correspondence to an internal reference model, allowing reconstruction of a 3D face shape and even generate new face shapes and expressions as a linear transformation of the analysis extracted models. In [3] PCA is applied to a coupled intensity-depth model capturing the joint maximum variation directions between intensity and depth and allowing to implicitly recover shape from intensity.

Some of the 3D face shape reconstruction techniques developed in the face recognition area have not always been applied globally. PCA allows us to discover joint maximum variation direction in a sample of images from the eigenvectors calculated over the covariance matrix, allowing reconstruction of every original image from a reduced number of coefficients as a weighted sum of eigenvectors and to identify it by using the Euclidean distance between the coefficients vector and another reference vector [4]. In modular PCA [5] images are divided into a specific number of sections and individual PCAs are applied for each section, which yield better results in variable illumination and facial expression conditions, as local variations don't affect every section. Therefore, using local representations improves the representation of facial features in variable illumination and expression situations. Eigenfeatures systems are modular PCA systems applied to main facial anatomic features: eyes, nose and mouth. In [6] they are applied with a recognition rate next to global PCA under varying expression and transparent glasses conditions, although not in partial occlusion.

The discrete cosine transform (DCT) has been used as an alternative to PCA facial analysis and recognition [7]. 2D DCT decomposes an image into a weighted sum of cosine basis functions with different orientation and spatial frequency. The first coefficients, corresponding to the lower frequencies and occupying the top positions in a zigzag pattern in the spectral representation, are sufficient to achieve an acceptable reconstruction and perform recognition with good results. Compared to PCA, DCT has the advantage of being based on a fixed set of basis functions that are not to be calculated beforehand. Discarding the first DCT coefficients increases recognition in images with large illumination variations because they contain information about the illumination changes [8]. Scaling all but the the first DCT coefficients according to a global maximum and minimum produces better recognition without the loss of information which is ruled out by discarding some of the initial coefficients [9]. It has been found that integration of global and local DCT features improves recognition [10], and the use of local features from DCT and local PCA in face verification is more robust in pose variations than global features [11].

As a way to improve recognition results by overcoming pose variations, some methods have been developed to construct different virtual pose views by using linear prediction models from local representations. In [12], [13] non frontal views of a face are predicted from a frontal image using multivariate linear regression (MLR) on the DCT coefficients of 8x8 size blocks. In the training phase, a linear transformation matrix is calculated for each pair of blocks occupying the same position in a pair of images of the same subject in different views. A transformation matrix of acceptable size is obtained for a moderate number of image samples by using dimensionality reduction techniques on each block. Blocks with different degrees of overlap are used and results are averaged in order to reduce prediction errors. Best results are obtained with higher overlap degree (87.5%) and using a reduced number of DCT coefficients (25%), as the dimensionality of the representation is kept low relative to the number of training elements. A method is presented in [14] to predict frontal appearance from a non-frontal pose image by calculating a linear operator relating views in different poses, both globally and based on different sized blocks between views whose correspondence is estimated with the help of an average 3D reference model.

The goal of this research is to show that depth prediction from intensity using local features may provide acceptable results. We use MLR to predict the depth corresponding to a frontal intensity image by calculating the linear transformation matrices that relate intensity and depth block pairs for each subject in the training phase. Then we are able to study whether depth predictions from out-of-training images are acceptable. We study the effect produced by the horizontal and vertical overlap degree between blocks, the dimensionality or number of coefficients used and the effect of variable illumination on the predictions of three different local representations: block based DCT, block based local PCA with a common eigenvectors base and modular PCA applied to bigger blocks corresponding to anatomic face features (eigenfeatures).

2. MULTIVARIATE LINEAR REGRESSION FOR DEPTH PREDICTION FROM LOCAL FEATURES

Let I be an intensity image, Z its corresponding depth image belonging to an N size sample of intensity and depth images, $I_{(x,y)}$ a fixed size block from I starting at (x,y) coordinates and $Z_{(x,y)}$ its corresponding depth block from Z. After performing a dimensionality reduction to D coefficients on both blocks for all N pairs we define a linear regression model:

$$B_{(x,y)} = A_{(x,y)} \quad W_{(x,y)}$$

$$\begin{bmatrix} Z_1^T \\ Z_2^T \\ \vdots \\ Z_N^T \end{bmatrix} = \begin{bmatrix} 1 & I_1^T \\ 1 & I_2^T \\ & \vdots \\ 1 & I_N^T \end{bmatrix} \begin{bmatrix} W_{1,1} & \cdots & W_{1,D} \\ W_{2,1} & \cdots & W_{2,D} \\ \vdots & \vdots & \vdots \\ W_{D+1,1} & \cdots & W_{D+1,D} \end{bmatrix} \quad (1)$$

Where B is the target vectors matrix, A is the extended source vectors matrix and W is a transformation matrix. In order to get W, we turn to the solution of the corresponding linear equations system using the pseudo-inverse matrix:

$$W_{(x,y)} = \left(A_{(x,y)}^T A_{(\mathrm{x,y})}\right)^{-1} A_{(x,y)}^T B_{(\mathrm{x,y})} \quad (2)$$

Every intensity vector may be then transformed into a depth vector by applying equation (1), and an intensity image may be transformed into a depth image by performing the same calculation for every block:

$$\{\mathrm{I}_{(\mathrm{x}_1,\mathrm{y}_1)}, \mathrm{I}_{(\mathrm{x}_2,\mathrm{y}_1)}, \dots, \mathrm{I}_{(\mathrm{x}_1,\mathrm{y}_2)}, \dots, \mathrm{I}_{(\mathrm{x}_\mathrm{n},\mathrm{y}_\mathrm{n})}\}$$

If we use overlapping blocks, an averaging operation is performed at the end to estimate depth at each pixel by counting the number of blocks contributing to that pixel.

3. METHOD

We used a sample of 104 pairs of face and depth images provided by the Laboratory for Image and Video Engineering (LIVE) at the University of Texas at Austin, Austin, TX [15]. All images corresponded to a neutral expression frontal view and were aligned by an ordinary procrustes analysis using five landmarks. We selected 65 pairs of images as the training set and the remaining 39 as the test set.

The MLR method of depth prediction from intensity was applied for three different kinds of local features. For DCT we used an 8x8 block size. For local PCA an 8x8 block size and a previously built common eigenvectors base was also used, from all the horizontal and vertical overlapping blocks from 2 images (15 conditions with a total of 3984 blocks). For modular PCA two 32x32 size block grids were used to build a specific eigenvectors base for every block. As shown in Figure 1, both block grids capture important facial features in an overlapping manner. An initial training phase for every condition was performed using the originally illuminated intensity images to build the specific W transform matrixes for every horizontal and vertical overlapping blocks in the intensity and depth images. The first coefficient in the DCT condition was discarded as it captures every block average intensity, and was substituted for every sample average block depth in predictions. We subsequently performed experiments in order to ascertain the effect of different overlap degree, the dimensionality or number of coefficients used in the MLR, and two specific illumination variations, point source horizontal angle and ambient illumination proportion, over the reconstructed depth map fractional error percent:

$$\left(\sum \frac{|\mathrm{Z}(x,y) - \mathrm{Z}'(x,y)|}{\mathrm{Z}(x,y)}\right) \cdot 100$$

Where Z(x,y) represents the true depth at each point, and Z'(x,y) represents the predicted depth at each point.

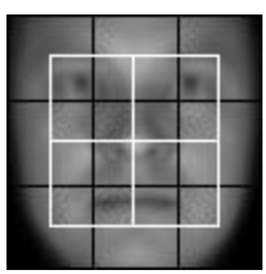

FIGURE 1: The two 32x32 block grids used in the modular PCA condition superimposed to the average intensity image: (a) beginning at (0,0) in black, and (b) beginning at (16,16) in white. Both block grids capture important facial features.

4. NUMBER OF FEATURES AND OVERLAP DEGREE

Figure 2 shows that for all predictors, fractional error increases when increasing the number of coefficients, while it decreases with the degree of overlap between blocks. In the two-way ANOVAs between number of coefficients and overlap degree, both factors proved to be significant for DCT ([$F(2,76) = 96.3$, $p < .001$, $\eta_p^2=.72$] , [$F(2,76 = 99.99$, $p < .001$, $\eta_p^2=.73$]), local PCA ([$F(2,76) = 338.6$, $p < .001$, $\eta_p^2=.9$] , [$F(2,76 = 118.63$, $p < .001$, $\eta_p^2=.76$]) and for number of coefficients in modular PCA[$F(2,76) = 19.83$, $p < .001$, $\eta_p^2=.34$], with assumed sphericity in all cases. When comparing fractional error between different features for a number of 16 coefficients and an overlap of 87.5%, results yield an advantage of local and modular PCAs over DCT, with respective marginal means of 9.64, 9.67 and 10.62%. Figure 3 shows that the difference error under original illumination conditions tends to be lower for modular PCA than for DCT and local PCA.

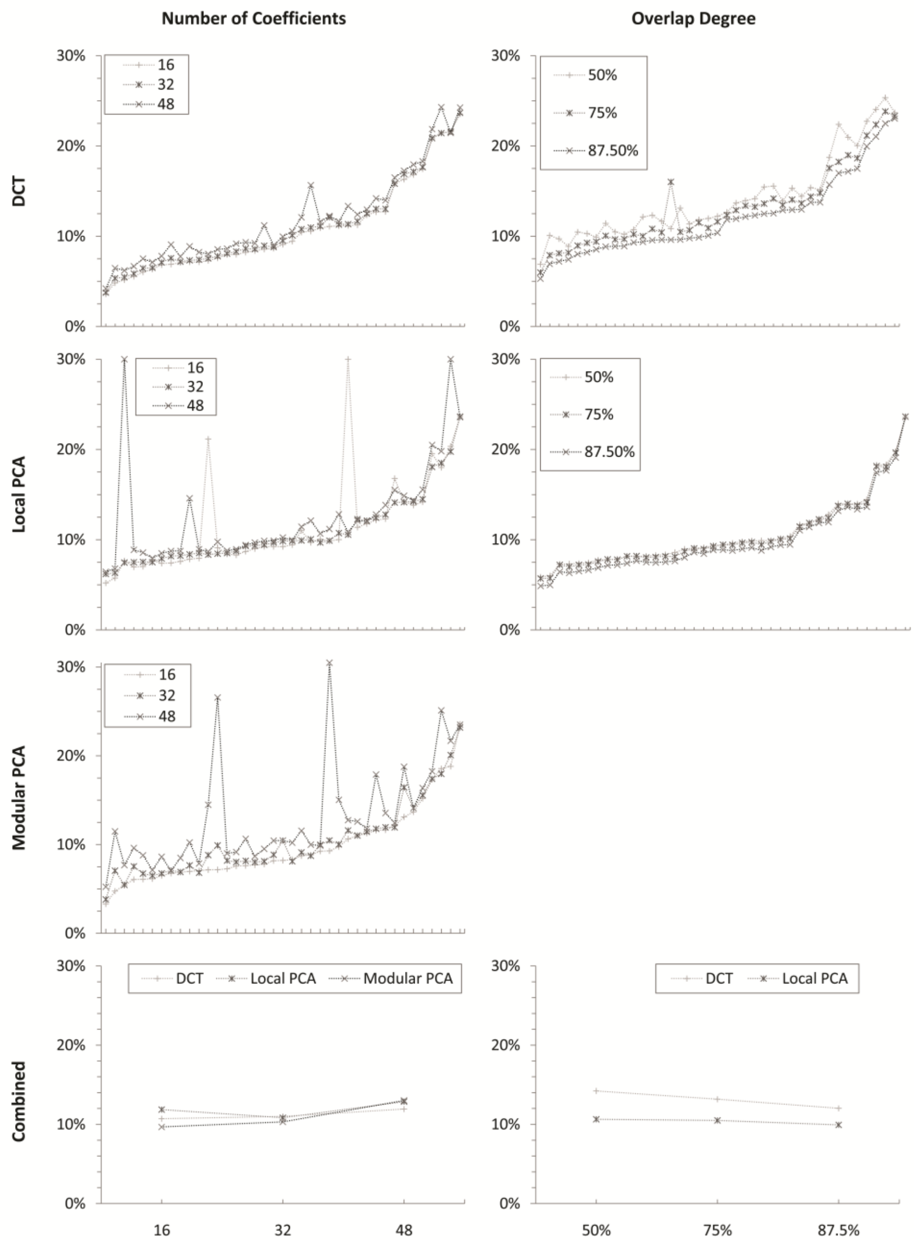

FIGURE 2: Fractional error by number of coefficients and overlap degree for DCT, local PCA and modular PCA conditions. The first three rows show the individual fractional errors. The last row shows combined marginal means for DCT, local PCA and modular PCA conditions.

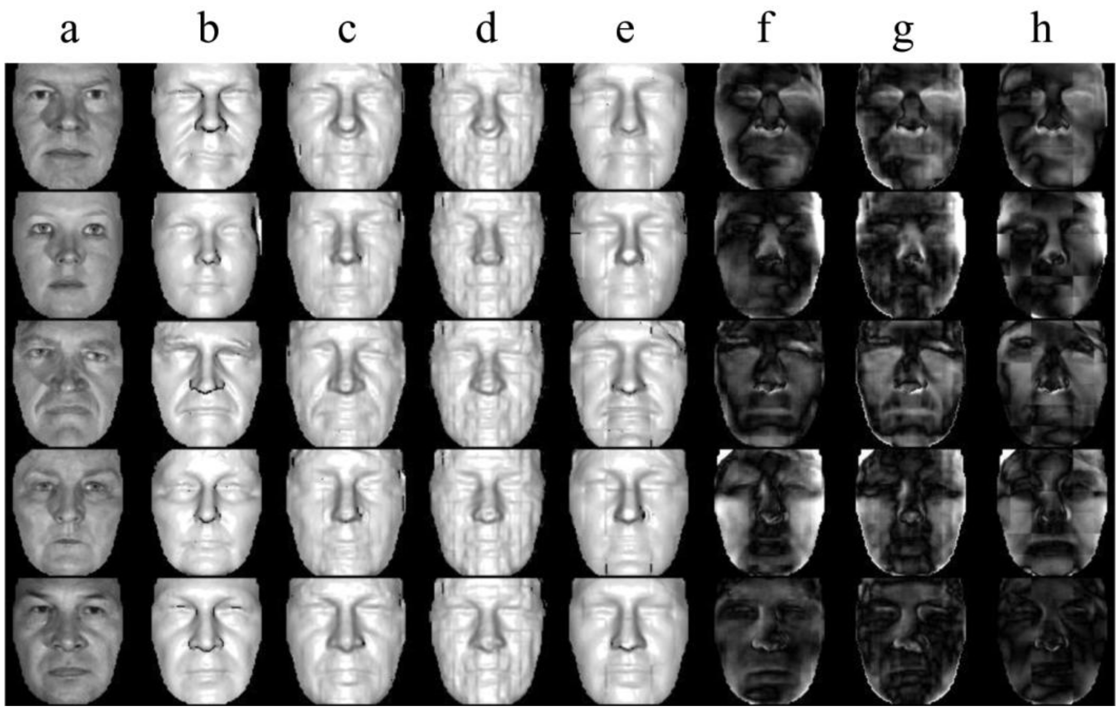

FIGURE 3: Predicted depths as frontal re-illumination for several out-of-training images using different local predictors. For every row: a) the input image, b) true depth, c) predicted depth for DCT, d) local PCA, e) modular PCA, f) error or difference between prediction and true depth for DCT, g) local PCA and h) modular PCA. Lighter values indicate greater error.

5. ILLUMINATION VARIATIONS

We conceptualized image illumination as a four spherical harmonic approximation [16] and rendered new intensity images by varying the light source angle over the horizontal plane from 0º (frontal position, lighting vector [0,0,1]) to 45º (to the left) at five levels at regular intervals and the proportion of ambient illumination over the total illumination from 0.2 to 0.9, at five levels at regular intervals, and a level 0 condition. The effect of illumination variations over the logarithm of coefficients variances and prediction fractional error was studied.

Figure 4 shows that most of between-illumination variance is concentrated in the first coefficient for DCT [8]. In the rest of the conditions the between-illumination variance is shared between the different coefficients. It's noted that between-illumination variance is almost negligible in the case of local PCA, while in the first block grid from modular PCA (a), which includes the blocks of the edge of the face, coefficients variance is greater than in the second grid (b). Between-subjects variances were greater than between-illumination for all conditions when performing a t-test. For DCT [$t(64) = 4.67$, $p < .001$, $d' = 0.82$, $r = .38$], local PCA [$t(64) = 4.68$, $p < .001$, $d' = 0.83$, $r = .38$], modular PCA grid a [$t(64) = 41.91$, $p < .001$, $d' = 6.26$, $r = .95$] and modular PCA grid b [$t(64) = 50.71$, $p < .001$, $d' = 7.98$, $r = .97$].

Figure 5 shows that the effect of horizontal light source angle and proportion of ambient illumination is small but consistent for the DCT condition ([$F(4,152) = 25.1$, $p < .001$, $\eta_p^2 = .49$], [$F(5,190) = 12.18$, $p < .001$, $\eta_p^2 = .24$]) and local PCA condition([$F(4,152) = 92.6$, $p < .001$, $\eta_p^2 = .71$], [$F(5,190) = 18.15$, $p < .001$, $\eta_p^2 = .32$]). But it is more noticeable in the case of modular PCA ([$F(4,152) = 15.97$, $p < .001$, $\eta_p^2 = .28$], [$F(5,190) = 10.33$, $p < .001$, $\eta_p^2 = .21$]). Specifically, increasing the horizontal light source angle is accompanied by an increase in the fractional error in modular PCA, while increasing the proportion of ambient illumination results in a decrease of fractional error as shown in the combined averages row of Figure 5.

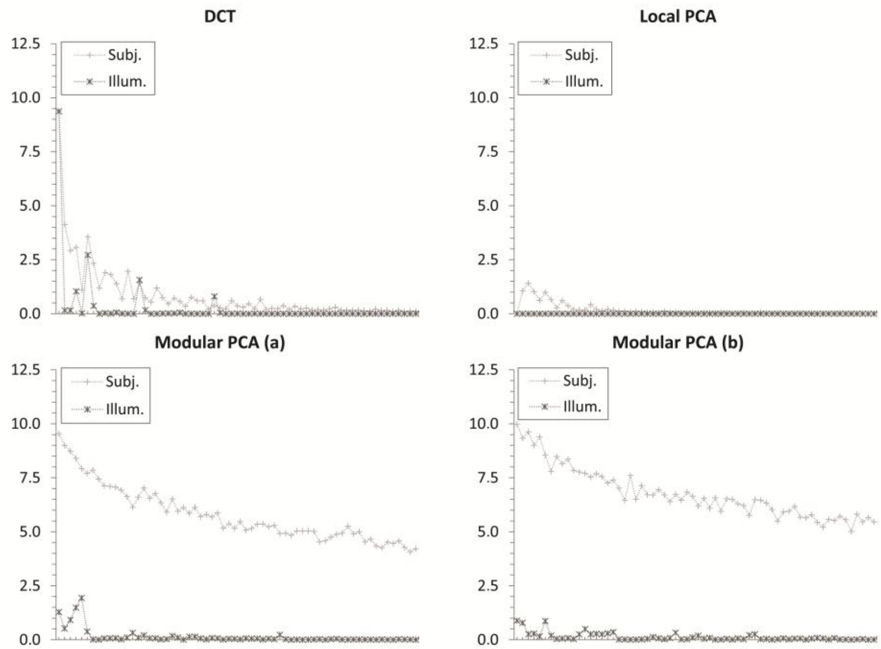

FIGURE 4: Logarithm of between-subjects and between-illumination variances for DCT, local PCA and modular PCA conditions. For modular PCA results for grid starting at (0,0) (a) and starting at (16,16) (b) are shown.

6. EXTERNAL IMAGE TEST

Finally, we performed an experiment with 10 real world images from the Yale B database. In Figure 6, we show the input image of each subject, with a frontal and side view of the modular PCA predicted depth map as frontal re-illumination. The results using modular PCA are visually acceptable and resemble some of the subjects' distinctive features in the frontal and also in the side view.

7. CONCLUSIONS AND FUTURE WORK

We applied a new method that allows reconstruction of 3D face shape from local features prediction using a MLR model with acceptable results for out-of-training and external images. The best results are obtained by using the eigenfeatures coefficients from modular PCA in images frontally illuminated in the same way as in the training phase, but it deteriorates as the illumination angle departs from the frontal position. The presence of ambient illumination compensates for this effect. Local PCA and DCT with the first coefficient removed produces poorer depth predictions that are robust against illumination variations. The proposed method is in contrast with state-of-the-art methods to the extent that it is local. It applies a local regression method to face features instead of doing it to the whole face globally. Global methods try to extract regularities between the whole face intensity and depth. We feel that a local approach may improve their results by taking into account variation modes between intensity and depth that are specific to certain facial features but not to the whole face as a unit.

A drawback of every method that uses aligned face images is that no alignment method produces a perfect match between facial landmarks. In future research, we'll try to improve the results by improving the images alignment by morphing every image to the average shape of a set of manually located features points [17] and by using more elaborate regression models, as partial least squares [18] and intensity-depth coupled models [3]. This will allow us to yield a quantitative comparison to state-of-the-art global methods.

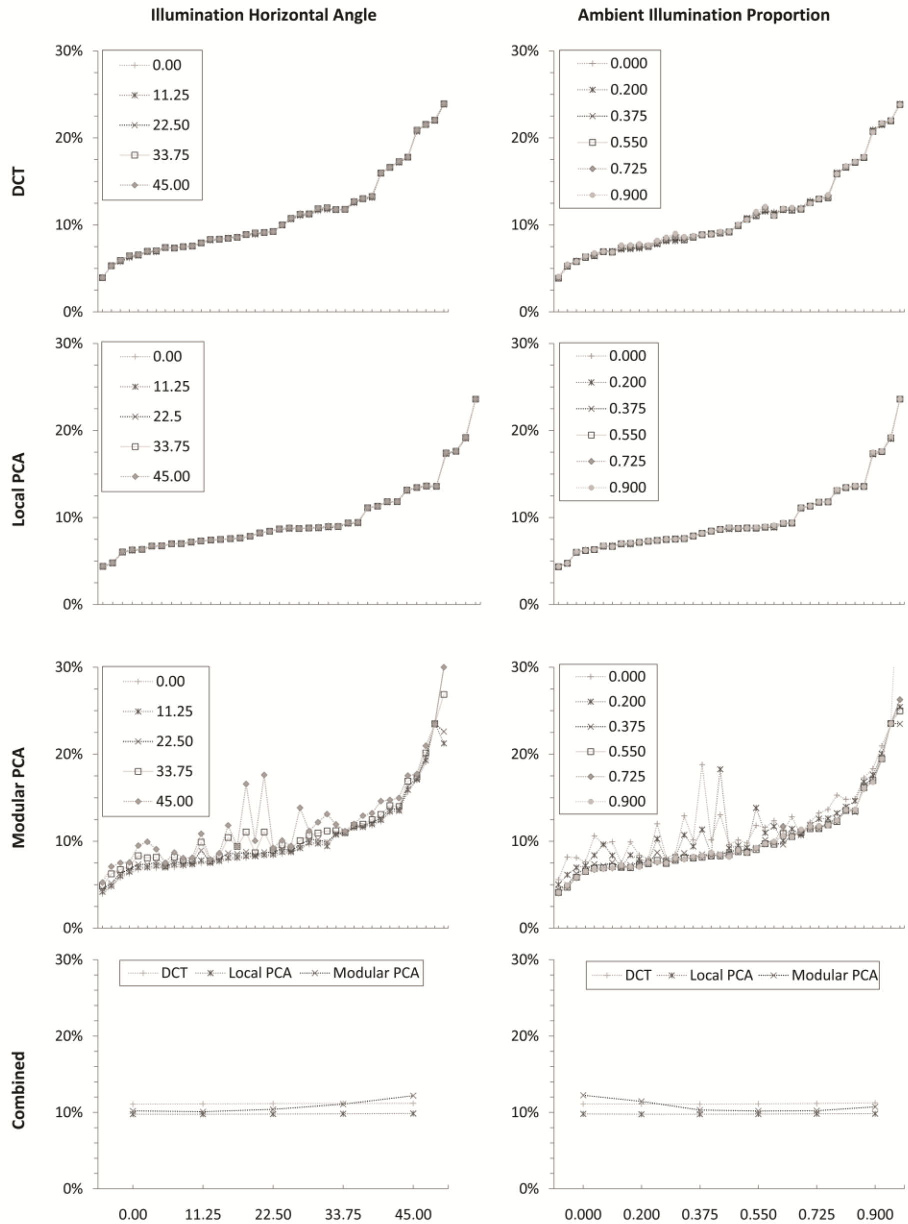

FIGURE 5: Fractional error by illumination horizontal angle and ambient illumination proportion for DCT, local PCA and modular PCA conditions. The first three rows show the individual fractional errors. The last row shows combined marginal means for DCT, local PCA and modular PCA conditions.

FIGURE 6: Predicted depths as frontal re-illumination for external images from the Yale B database using the modular PCA (eigenfeatures) predictor with 16 coefficients and a 87.5% overlap degree. Both grids starting at (0,0) and starting at (16,16) were used and their results were averaged. For every row in each panel the input image and two reconstructed views are shown.

8. REFERENCES

[1] J. J. Atick, P. A. Griffin and A. N Redlich. "Statistical approach to shape from shading: Reconstruction of three-dimensional face surfaces from single two-dimensional images." *Neural Computation*, vol. 8, no 6, pp. 1321-1340, 1996.

[2] V. Blanz and T. Vetter. "A morphable model for the synthesis of 3D faces." *Proceedings of the 26th annual conference on Computer graphics and interactive techniques*, 1999, pp. 187-194.

[3] M. Castelán, W. A. Smith and E. R. Hancock. "A coupled statistical model for face shape recovery from brightness images." *IEEE Transactions on Image Processing*, 2007, vol. 16, no 4, pp. 1139-1151.

[4] M. Turk and A. Pentland. "Eigenfaces for recognition." *Journal of cognitive neuroscience*, vol. 3, no 1, pp. 71-86, 1991.

[5] R. Gottumukkal, and V. K. Asari. "An improved face recognition technique based on modular PCA approach." *Pattern Recognition Letters*, vol. 25, no 4, pp. 429-436, 2004.

[6] P. Quintiliano, A. N. Santa-Rosa and R. Guadagnin. "Face recognition based on eigenfeatures." *Multispectral Image Processing and Pattern Recognition*, 2001, pp. 140-145.

[7] Z. M. Hafed and M.D. Levine. "Face recognition using the discrete cosine transform." *International Journal of Computer Vision*, vol. 43, no 3, pp. 167-188, 2001.

[8] W. Chen, M. J. Er, and S. Wu. "Illumination compensation and normalization for robust face recognition using discrete cosine transform in logarithm domain." *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 2006, vol. 36, no2, pp. 458-466.

[9] V. P. Vishwakarma, S. Pandey and M. N. Gupta. "Adaptive histogram equalization and logarithm transform with rescaled low frequency DCT coefficients for illumination normalization." *International Journal of Recent Trends in Engineering*, vol. 1, no 1, pp. 318-322, 2009.

[10] A. R. Chadha, P. P. Vaidya, and M. M. Roja. "Face recognition using discrete cosine transform for global and local features." *International Conference on Recent Advancements in Electrical, Electronics and Control Engineering*, 2011, pp. 502-505.

[11] C. Sanderson, M. Saban and Y. Gao. "On local features for GMM based face verification." *International Conference on Information Technology and Applications*, 2005, vol. 1, pp. 650-655.

[12] C. Sanderson, S. Bengio and Y. Gao. "On transforming statistical models for non-frontal face verification." *Pattern Recognition*, vol. 39, no 2, pp. 288-302, 2006.

[13] Y. Wong, C. Sanderson and B.C. Lovell. "Regression based non-frontal face synthesis for improved identity verification." *Computer Analysis of Images and Patterns*, Springer Berlin Heidelberg, 2009, pp. 116-124.

[14] Chai, X., Shan, S., Chen, X., & Gao, W. "Local linear regression for pose invariant face recognition." *International Conference on Automatic Face and Gesture Recognition*, 2006, pp. 631-636.

[15] S. Gupta, K. R. Castleman, M. K. Markey, and A. C. Bovik. "Texas 3D face recognition database." *IEEE Southwest Symposium on Image Analysis & Interpretation*, 2010, pp. 97-100.

[16] R. Basri, and D. W. Jacobs. "Lambertian reflectance and linear subspaces." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2003, vol. 25, no 2, pp. 218-233.

[17] I. Craw and P. Cameron. "Parameterising images for recognition and reconstruction." *British Machine Vision Conference*, Springer London, 1991, pp. 367-370.

[18] M. Castelán and J. Van Horebeek. "3D face shape approximation from intensities using Partial Least Squares." *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, 2008, pp. 1-8.