Estimation of face depths by conditional densities

John A Robinson, Justen R Hyde
Department of Electronics, University of York, YO10 5DD
jar11@ohm.york.ac.uk

Abstract

The expected value of missing data in a sample taken from a multivariate normal probability distribution is the mean of the conditional distribution of the missing dimensions given the known dimensions. We explain the derivation of this result, demonstrate its application to face image processing, then use it in a new method for recovering shape from image data. The context of our work is the use of 3D facial models to aid in recognition of human faces by humans. We explain the requirement for such models and review the practical possibilities for encoding depth information alongside photographs in identity documents like passports. The best alternative is to derive depths automatically from the photos, as this requires no side information. We show experimentally that conditional density estimation provides accurate face depth recovery, without recourse to explicit modelling of surface shape.

1 Introduction

In this paper we are principally concerned with the analysis of images of human faces. These may either be 2D images, where an array of pixels represents greyscale, or 2.5D (or, informally, 3D) images, where a second array represents depths. In common with other appearance-based approaches (e.g. [1]-[4]), we stack all the measurements (greyscales and depths) for an image into an *n*-dimensional column vector, which we consider as a sample point in a multidimensional space. Unlike most other appearance-based approaches, we do not assume that "face space" is a low-dimensional subspace of "image space". Rather, we model faces as an *n*-dimensional normal distribution, with the parameters derived from training data. With this representation, conditional distributions can be used to estimate any number of missing measurements in a sample. Section 2 explains the theory, highlights some properties of the estimation, then shows how it can be applied to image reconstruction from partial data. All examples are faces, but the generalization to other domains is briefly considered.

In section 3 we introduce the application with which we are concerned: the recovery of depth information from face photos.

Because the number of samples available for training is less than *n* and therefore far too low for the scatter matrix of training samples to characterize the covariance matrix, we use a regularized covariance estimation method [5]. This mixes the scatter matrix and the identity matrix according to a reconstruction criterion, as described in section 4, which also contains the results of experiments to find the appropriate mixing parameter.

Section 5 reports experiments on the recovery of depth data from photos of faces.

2 Conditional distributions for data estimation

2.1 Theory

Consider the random vector $\boldsymbol{X}$, distributed as $N_n(\mu,\Sigma)$. Suppose we have a particular sample from the distribution $\boldsymbol{P}$, which we call the "probe", in which some of the measurements are

BMVC 2005 doi:10.5244/C.19.77

known and some are not. A binary vector ***M*** (or "mask") may be used to indicate which dimensions in ***P*** are known values, and a permutation matrix ***R*** can then be defined which will reorder ***M*** into a column of q zeroes followed by p ($= n - q$) ones.

When ***R*** is applied to ***P***, it moves all the unknown values to the top:

$$P_{perm} = RP = \begin{bmatrix} P_1 \\ \hline P_2 \end{bmatrix} \tag{1}$$

where P_1 and P_2 are column vectors of dimensionality q and p respectively (P_1's values are undefined). Similarly,

$$X_{perm} = RX = \begin{bmatrix} X_1 \\ \hline X_2 \end{bmatrix} \tag{2}$$

We also define $\mu_{perm} = R\mu = \begin{bmatrix} \mu_1 \\ \hline \mu_2 \end{bmatrix}$ and $\Sigma_{perm} = R\Sigma R = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \hline \Sigma_{21} & \Sigma_{22} \end{bmatrix}$ (3)

I.e. μ_{perm} is the mean vector reordered to match ***RP***, the reordered probe vector. Similarly Σ_{perm} is the covariance matrix with rows and columns appropriately reordered.

We now follow [6] (page 170) and define a matrix A with submatrices and dimensions as shown:

$$\underset{(n\times n)}{A} = \begin{bmatrix} \underset{(q\times q)}{I} & \underset{(q\times p)}{-\Sigma_{12}\Sigma_{22}^{-1}} \\ \hline \underset{(p\times q)}{0} & \underset{(p\times p)}{I} \end{bmatrix} \tag{4}$$

Now the random vector

$$A(X_{perm} - \mu_{perm}) = A\begin{bmatrix} X_1 - \mu_1 \\ \hline X_2 - \mu_2 \end{bmatrix} = \begin{bmatrix} X_1 - \mu_1 - \Sigma_{12}\Sigma_{22}^{-1}(X_2 - \mu_2) \\ \hline X_2 - \mu_2 \end{bmatrix} \tag{5}$$

is a linear transformation of the normally-distributed random vector X_{perm} and so itself is normally distributed with mean $E\left[A(X_{perm} - \mu_{perm})\right] = AE\left[(X_{perm} - \mu_{perm})\right] = 0$ and covariance matrix $A\Sigma_{perm}A^T$ where the T denotes transpose.

Using the identities $\Sigma_{22}^T = \Sigma_{22}, \Sigma_{12}^T = \Sigma_{21}$ (because of the symmetry of covariance matrices), we can calculate equation (6):

$$A\Sigma_{perm}A^T = \begin{bmatrix} I & -\Sigma_{12}\Sigma_{22}^{-1} \\ \hline 0 & I \end{bmatrix}\begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \hline \Sigma_{21} & \Sigma_{22} \end{bmatrix}\begin{bmatrix} I & 0 \\ \hline (-\Sigma_{12}\Sigma_{22}^{-1})^T & I \end{bmatrix} = \begin{bmatrix} \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} & 0 \\ \hline 0 & \Sigma_{22} \end{bmatrix}$$

Equation (6) above is the motivation for choosing A as we did: when the covariance matrix is calculated, the top right and bottom left corners turn out to be 0 submatrices, meaning that $X_1 - \mu_1 - \Sigma_{12}\Sigma_{22}^{-1}(X_2 - \mu_2)$ and $X_2 - \mu_2$ have zero covariance and are therefore independent. We can therefore consider the quantity $X_1 - \mu_1 - \Sigma_{12}\Sigma_{22}^{-1}(X_2 - \mu_2)$ as a distinct q x q multivariate normal distribution. When X_2 takes the value P_2, the random variable becomes $X_1 - \mu_1 - \Sigma_{12}\Sigma_{22}^{-1}(P_2 - \mu_2)$. As shown above, $E\left[A(X_{perm} - \mu_{perm})\right] = AE\left[(X_{perm} - \mu_{perm})\right] = 0$, so $E[X_1 - \mu_1 - \Sigma_{12}\Sigma_{22}^{-1}(P_2 - \mu_2)] = 0$. But $\mu_1 - \Sigma_{12}\Sigma_{22}^{-1}(P_2 - \mu_2)$ is a constant, so the mean of X_1, i.e. the expected value of the missing data we want to fill in, is given by:

$$E[X_1] = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(P_2 - \mu_2) \quad (7)$$

We therefore have a direct method for estimating P_1 from P_2. All that then remains is to apply $\boldsymbol{R^{-1}}$ to P_{perm} to recover the full image corresponding to the probe.

The above derivation is indirect. It is also possible to construct a proof that uses the densities directly ([6] pp 217-218). Also note that, as well as the mean, the covariance matrix of the estimated data has been derived as $\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$. This could be used to explore the principal components of variation, i.e. the modes in which the real values are likely to differ from the estimate. We do not pursue use of the conditional density covariance further here.

2.2 Geometric interpretation of the estimation.

Figure 1 illustrates the estimation of a missing value in two dimensions. The geometric interpretations it suggests apply in *n*-dimensional space, namely: (a) The use of the expectation of the conditional density is equivalent to finding the point where the known subspace is tangent to an equivariance contour, or, equivalently, finding the minimum Mahalanobis distance from the known subspace to the class mean. (b) Given a particular known subspace (i.e. a particular distribution of pixels in the probe), the derivation of the missing values is a linear transformation. (c) It is important to have a full *n*-dimensional covariance matrix for the distribution. Were the method to be applied in a principal component subspace, for example, the missing values would be estimated by reflecting in the principal components. This will lead to large errors when the retained principal components are near parallel to missing value axes.

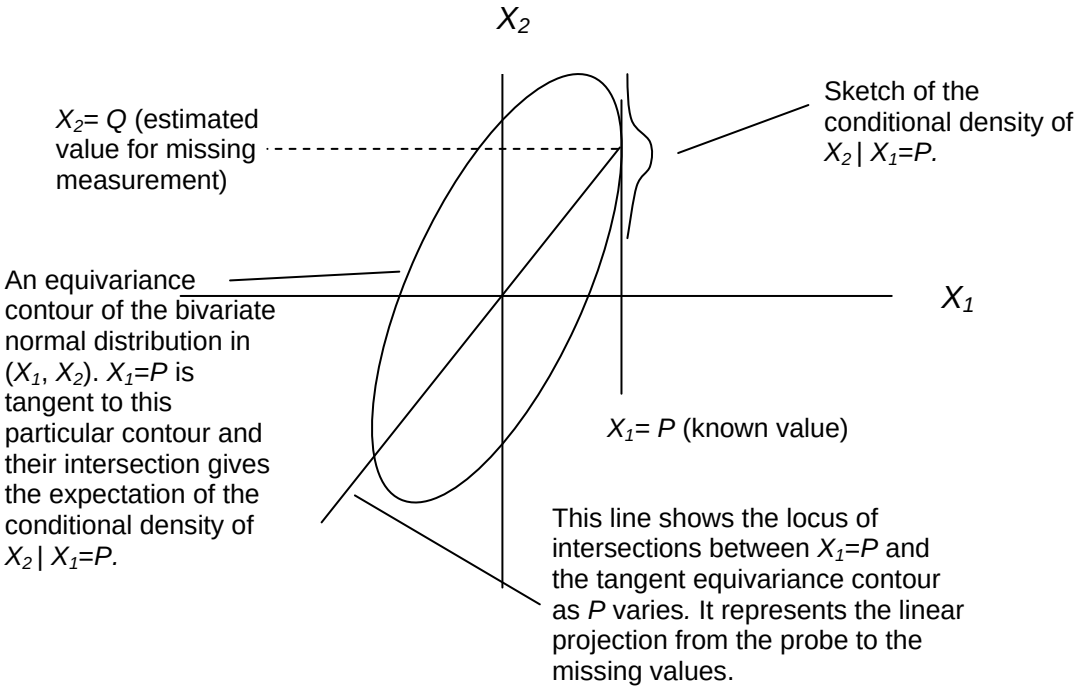

Figure 1: Illustration of estimation for a 2D distribution

2.3 Applying conditional estimation to the recovery of missing data in face images

Although the focus of this paper is the recovery of depths from greylevels, we briefly illustrate the use of conditional density estimation for filling in missing or damaged data in images.

Figure 2 shows the mean of a training set of 2325 38x38 face images together with the individual pixel variances (i.e. the diagonal of the scatter matrix). The faces are modelled as a multivariate normal distribution using the sample mean as maximum likelihood estimate of the distribution mean, and with the covariance matrix estimated as described in Section 4 below (and, more fully, in [5]).

Figure 2: Sample mean and pixel variances for a 38 x 38 face-image training set

Figure 3: Example recovered pictures by estimation from pixels extracted by probes

Figure 3 shows a table of pictures recovered by application of equation (7) to sets of probe pixels from face images. Five original test images (none of which were in the training set) are shown along the top of the figure, while six different probe patterns are shown down the left. In the probe images, white pixels represent points used in the probe. The dimensionality n is 38x38 = 1444, while the number of points p in the probe is as follows: (a) the probe subsamples the image at regular intervals, $p = 100 = n/14$, (b) the probe subsamples the image at coarse regular intervals, $p = 49 = n/29$, (c) the probe is a fovea-like logarithmic spiral, $p = 104 = n/14$, (d) the probe consists of all the face except the right-hand triangle, $p = 1083 = 3n/4$, (e), the probe consists of the whole face except the eyes, $p = 1015 = 7n/10$, (f) the probe consists of the whole face except the mouth region, $p = 1015 = 7n/10$. Note that the known data in every original probe is taken directly from a test image and is not filtered.

Figure 3 shows that recovery is possible no matter what the probe. Each additional pixel in the probe provides additional data about how the recovered picture should diverge from the

mean, but the divergence is conservative. For example, in row (b), the recovered faces are all plausible though blurred, and reflect the lighting and pose of the original. So far as identity and expression are concerned, however, they tend towards the mean, as would be expected given the dimensionality requirements for identification and expression analysis [7]. Reconstructions improve with increasing probe size, though wherever there are large areas missing, the fill-in again has the property of being appropriate in both geometry and shading and therefore plausible, but neutral so far as expression is concerned.

3 3D photos for human recognition

The accuracy of human face recognition from single photographic images as on passports, driving licenses, etc., is surprisingly low, as demonstrated strikingly by Kemp *et al* [8]. Here, six subjects in the environment of a supermarket were asked to be alert for fraudulent use of credit cards. A test set of volunteers used credit cards with photographic ID. 35 percent of cards with significant variation between image and bearer were accepted, as were 64 percent of cards with similar but different faces to those of the bearer. 14 percent of non-fraudulent cards were rejected.

Part of the reason for this poor performance is believed to be that in many practical cases the faces being recognised are not well known to the observer [9]. This is significant as with only one viewpoint to compare with a testing view, the observer's perception of the ground truth (the person in the photo) is heavily reliant upon the environmental conditions when the 2D image was taken – lighting, pose, and the presence of any transient features (spectacles, hats, etc) at the time. This is not such a significant problem when the face is already known by the observer, as familiar faces are recognised more accurately than unfamiliar faces under most circumstances. Familiarity appears to compensate for distracting influences that could be considered external. Two of these are significant for our purposes:

1. As a face becomes more familiar, the observer becomes more adept at remembering, and therefore using as recognition cues, the "internal" features of the face – that is, features such as eye separation, mouth width, etc., which are not necessarily immediately apparent, but which are highly distinctive between individuals. Less familiar faces are recognised by using more obvious, large scale features, such as hairstyle, facial hair, spectacles, etc, which are readily apparent, highly distinctive when compared to the average face, and also potentially highly misleading as they are both easily changed and do not distinguish between – for example – all people with the same hair colour and style very effectively.
2. As a face becomes familiar, more views of the face are stored and associated with the identity. In this model, transient features (such as hairstyle and facial ornaments) also become less significant, though for different reasons. Here, the presence or absence of, say, spectacles, becomes irrelevant as the observer associates both states to the person in question. The more significant element of this recognition process is the ability of the observer to recognise the subject under varying lighting conditions and in different poses as images of the subject in increasing numbers of different poses and conditions are stored in the associative memory of the observer. Troje and Kersten [10] show the significance of this process effectively. Here, subjects are presented with a selection of familiar faces (close colleagues) from a variety of angles along with an image of their own face (which they are familiar with in the frontal view) in profile. The speed at which they recognised and correctly named the faces was recorded, and showed while other familiar faces could be recognised equally quickly independent of view, the subject's own face was consistently recognised more slowly in profile than in frontal view. Troje concludes that this is because although we are familiar with our own faces in the frontal view (as from our reflection), we rarely see our own face in profile, and so are less primed to recognise it.

It has long been understood that familiar and unfamiliar faces, as subsets of familiar and unfamiliar objects, are recognised so differently that completely different memory processes are used – familiar using semantic memory, and unfamiliar using episodic memory. This makes it difficult to draw conclusions about one type of recognition from the other, as both

have distinct and very different properties [11]. However, we can approach the problem from a different angle: we now have a list of external distractors which make unfamiliar face recognition rates poorer than familiar recognition rates. If these distractors could be in some way removed, then recognition should be less difficult.

A further, very significant, interpretation of the Troje study [10] is that recognition is based on a collection of views, rather than a mental 3D model of a face. If a 3D mental model were used, recognising a familiar face from different angles should have no impact upon performance, whereas recognition based on a collection of views (akin conceptually to searching through a photo album of many views of a candidate before deciding upon recognition) will degrade when a familiar face is presented in a novel orientation. Therefore, the use of a virtual 3D model should provide functionality not available implicitly to the human recognition process.

A natural extension of the study by Troje, and indeed the concept of pose as a problem for unfamiliar face recognition, is to posit that if a face is unfamiliar then pose, lighting and external feature changes will have a disproportionate impact on the observer's identification of the face. Several studies [12,13] have suggested that changes in pose have a detrimental effect upon recognition though Troje and Bulthoff [14] suggest that the training view provided is significant, and the testing view is unimportant – i.e., with a good training view, any change in pose becomes unimportant. [15] demonstrates that changes in illumination direction can hinder identification, while Liu and Chaudhuri [16] support the view that facial structure is determined principally from shading – not stereopsis – which suggests that lighting has a strong effect upon our perception of face shape. Also of note is the finding [17] that perspective distortion caused by camera distance can have a severe effect upon recognition.

In subjective tests we have shown that both lighting and pose changes do indeed have an impact on fully textured 3D facial masks, lighting being greater but pose being far from insignificant. To summarise briefly, test subjects were exposed to a set of twenty training faces, in a random order, posed either facing the camera or looking off at approximately twenty degrees, for five seconds each. Lighting for each image varied between sources to the lower left or upper right of the face. The same twenty faces, all shown with either the alternate pose, lighting or both relative to the training image were then shown to the subjects mixed with twenty novel faces, rendered under the same range of conditions. After testing, a distinctiveness survey was used to remove unusually distinctive faces from the results. As can be seen in table 1, both pose and lighting changes pose a significant problem for recognition. Further, with a compound change in both lighting and pose correct identification drops to less than 60% after normalising for the positive or negative bias of each test subject.

	Correct	False Negative	Normalised Correct	Normalised False Negative
Same Lighting, Different Pose	74.50%	25.50%	77.38%	23.44%
Different Lighting, Same Pose	60.00%	40.00%	72.07%	33.41%
Different Lighting, Different Pose	52.88%	47.13%	59.20%	43.52%

Table 1: Recognition of faces in different poses and lighting conditions

We conclude, therefore, that adjusting pose and/or lighting can significantly degrade the ability of an observer to correctly identify an unfamiliar face. If a full, accurate 3D model of the face in question were provided, this degradation may be ameliorated by mapping the initial 2D image onto the model, re-positioning and re-lighting the model to conform to any given environmental test conditions and re-rendering a new 2D image. For example, a face caught on camera in profile may be more accurately compared with a passport photo by applying the full frontal passport image to the 3D model of the passport owner's head and rotating to a profile view, re-illuminating to match the conditions in the test scene. Further manipulation, such as

the addition or removal of spectacles, beard, hats, etc, are also possible as is the adjustment of perspective distortion upon the model. While this does not confer any form of familiarity with a novel face, which would be the ideal method of improving recognition, we suggest that the use of a well-fitted 3D model would allow us to remove many of the distracting factors that cause significant degradation to unfamiliar face recognition, in essence trying to boost the performance of unfamiliar recognition by providing some of the empirically apparent functionality of the familiar face recognition process.

People do not carry laser-scanned 3D models of their own heads but only 2D photographs. Hyde and Robinson [18] suggest that there is a useful correlation between the greyscale information in a standard passport style photo and a depth map of the face which can be exploited in an application-specific coding scheme. We propose that the depth information for a face could be estimated from the greyscale information present in a 2D image with sufficient accuracy to provide a better 3D model than the mean face, and allow the re-posing and re-illumination of a 2D image. In the conditional density estimation process described in section 2, we have a mechanism for doing this. Training images consisting of greyscales and depths will be used to estimate a regularized normal model for faces. Greyscale faces will then be used as probes to recover all the missing depth dimensions.

4 Covariance matrix estimation

Robinson [5] reviews methods for covariance matrix estimation in the context of face classification and detection, then proposes an estimator of the form:

$$\mathbf{\Sigma}_i(\alpha,\beta) = \alpha\mathbf{S}_i + (1-\alpha)\mathbf{S}_{total} + \beta\mathbf{I} \tag{8}$$

where $\mathbf{\Sigma}_i(\alpha,\beta)$ is the estimated coefficient matrix for class *i*, $\mathbf{S}_i$ is the scatter matrix derived from a subset of the training samples of class *i*, $\mathbf{S}_{total}$ is the scatter matrix of all available training samples over all classes and α,β are regularization parameters estimated by classifying the remainder of the training samples for class *i* and choosing the α,β that give best performance. The estimator in [5] includes further regularization parameters γ_i which adjust the volumes of each class, again by maximizing classification performance. In our context, we have a single class so $\mathbf{S}_i = \mathbf{S}_{total}$, and α and γ_i have no effect. The estimator therefore reduces to

$$\mathbf{\Sigma}_i(\beta) = \mathbf{S}_i + \beta\mathbf{I} \tag{9}$$

In contrast to earlier regularization methods (e.g. [19,20]), the estimator of [5] optimizes its parameters according to the application. Therefore during training we consider the estimation of missing data as a recovery problem and search for the β that produces the average best estimate during a training phase using sample images with known depths.

The data available to us for the 3D face estimation comprises some 740 laser scanned images from the University of Notre Dame biometrics database. This database contains multiple images of some subjects in slightly different poses. In order to provide fair test images, 40 were removed from the set. These 40 were selected such that no other images of any of the subjects in the control set were present in the remaining 700 training images. The 3D models were automatically converted into depth map images by filtering the images to remove noise caused by surface reflectance peaks which can cause highly inaccurate readings from a laser scanner, and then centring the point closest to the camera. It can be assumed that all images are facing forwards, so this nearest point is always the nose. This is verified by checking that the image has approximately the same amount of non-zero data to both the left and right of the peak. A set-size sample window is set about this central peak, cropping the model and texture information to the face portion of the images only. A backplane is then set by scanning through all models to determine the nearest backplane which does not result in data loss. The texture and depth maps are then generated using a rendering library (OpenGL) to return depth and colour information for specified sample points.

As so few training samples are available relative to the size of the space of the problem, we have resized the images to 38 by 50 pixel greyscale texture and depth maps, limiting the dimensionality of the space to 3800 dimensions. All training images are mirrored to double the number of training samples available to 1400.

Optimization was performed by sweeping through values for β from 1 to 17500. The covariance matrix was computed using a "leave-one-out" method – that is, omitting one of the training images, computing the covariance with the remainder, regularising using the current control value, calculating the projected depth model and hence a mean square error. As several subjects in the set are represented in multiple images, all instances of a given test subject were omitted from the training set when testing using an image of that subject. As shown in figure 4, a clear minimum is evident in the results, at approximately 4500.

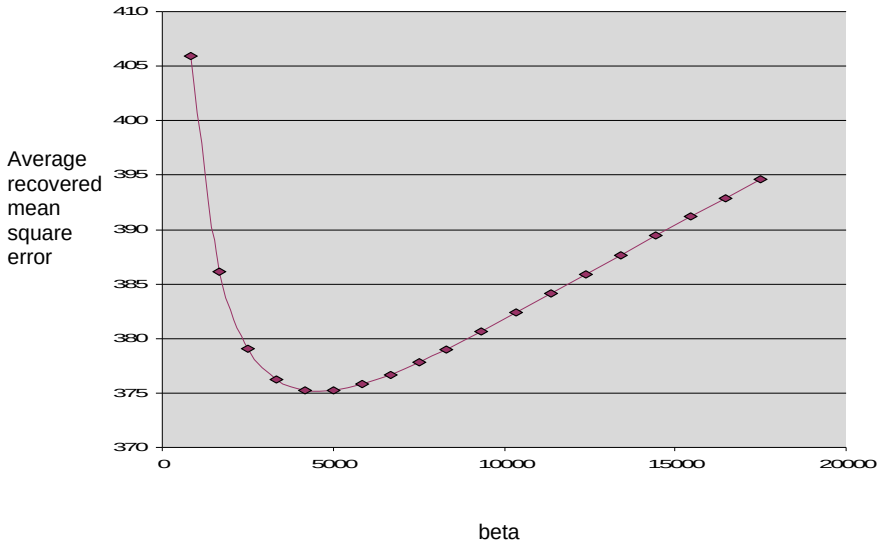

Figure 4. Average reconstruction error vs parameter β during training

In order to better understand the impact of varying the size of the training data, we have applied a simplified method of parameter estimation to cases where the scatter matrix was estimated from 175, 350, 525 and 700 images. In this experiment, we simply evaluated the optimal β for the forty test images. Table 2 shows the results:

Number of training images	175	350	525	700
Optimal β for the test set	20000	28889	11329	12210
Average MSE over the test set at this β	549	546	479	495

Table 2. Optimal parameter values for restricted training sets

In each case we also constructed the equivalent of figure 4. The graph's shape is always the same: there is an initial rapid improvement in performance as the parameter value is increased. This then reaches a minimum with a very shallow curve, and then degrades towards the standard deviation of the training set (i.e. the average distance between the training samples and their mean). It will asymptotically approach this value, which corresponds to the use of the mean depth for recovery, as β tends to infinity. While the estimation improves over the mean in all cases, improvement is clearly greater with more training samples. The β yielding the optimal performance decreases as the number of samples increases.

5 Recovery of depths from test images

Application of an estimation transform using the optimal β value of 4500 determined in section 4 to the forty test images not used in training yields results which consistently outperform the use of the mean depth image. Figure 5 shows that recovery via the conditional density provides on average an estimation with less than half the MSE of the mean depth. As shown in figure 6, the estimation provides a satisfactory model for use in a 3D render of the face. The top four rows of figure 6 are representative examples, while the bottom row shows the worst case (image 9 in figure 5). The original depths in this image were distorted and that has contributed to the poor MSE, but the reconstruction is also subjectively poor, probably because of the lack of images of similar framing and pose in the training set.

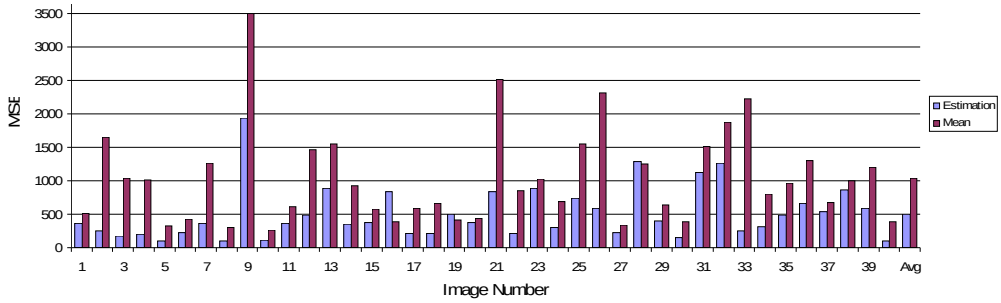

Figure 5. Depth recovery accuracy as Mean Square Error for 40 test images

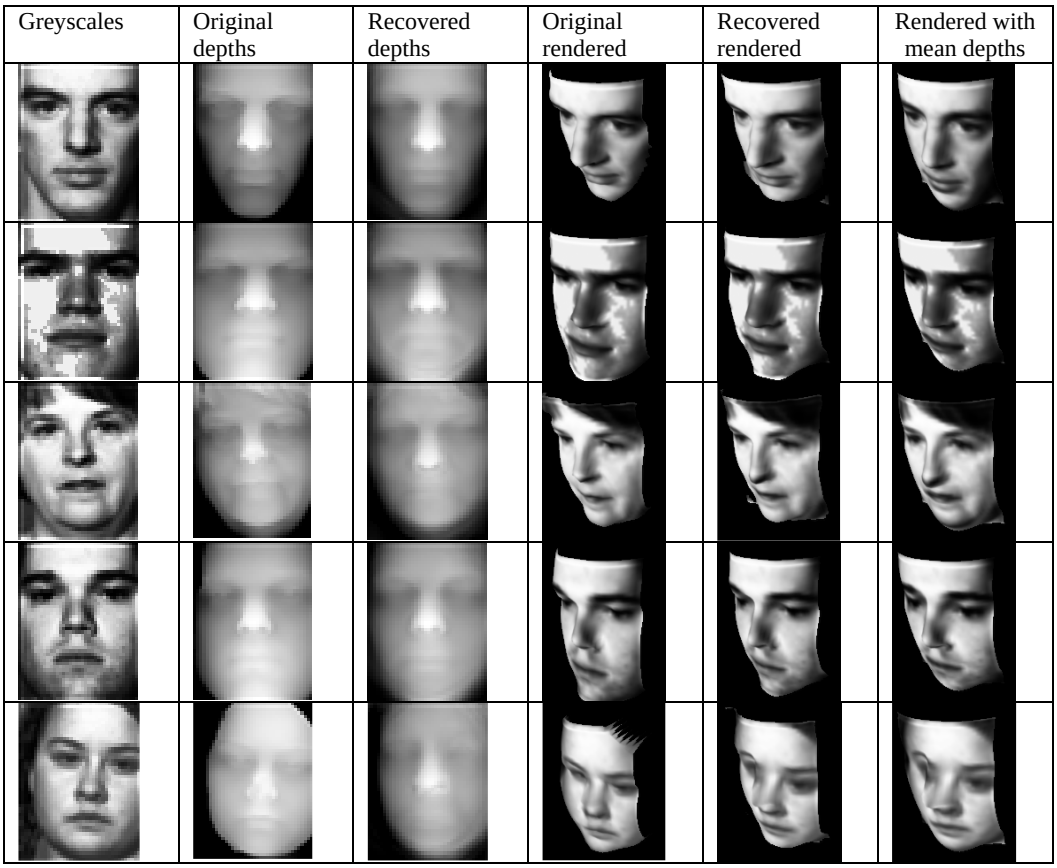

Figure 6. Example depth reconstructions

6 Conclusions

We have applied conditional distribution estimation to the recovery of missing data in images. In particular, we have shown how accurate estimates of face depths are recoverable from face photos. In future work we will use the depth data to relight the photographic information, and test the extent to which this aids human recognition.

References

[1] M Kirby, L Sirovich, "Application of the Karhunen-Loeve Procedure for the Characterization of Human Faces", *Transactions on PAMI*, Vol 12, No 1, January 1990, pp 103-108.

[2] V Belhumeur, J Hespanha, D Kriefman, "Eigenfaces vs. Fisherfaces: Recognition using Class Specific Linear Projection", *IEEE Transactions on PAMI*, Vol 19 No 7, July 1997, pp 711-720.

[3] B Moghaddam, A Pentland, "Probabilistic Visual Learning for Object Representation", *IEEE Transactions on PAMI*, Vol 19 No 7, July 1997, pp 696-710.

[4] B Schoelkopt, A Smola, K-R Muller, "Nonlinear Component Analysis as a Kernal Eigenvalue Problem", *Neural Computation*, Vol 10 No 5, 1998, pp1299-1319.

[5] J A Robinson "Covariance estimation for appearance-based image processing", *British Machine Vision Conference 2005*, Oxford, UK, September 2005.

[6] R A Johnson, D W Wichern, *Applied Multivariate Statistical Analysis*, Prentice-Hall, Upper Saddle River, NJ, 1998.

[7] P S Penev, L Sirovich, "The Global Dimensionality of Face Space", *Fourth IEEE Internat Conf Automatic Face and Gesture Recognition*, Grenoble, France, March 26-30 2000.

[8] R Kemp, N Towell, G Pike, "When Seeing should not be Believing: Photographs, Credit Cards and Fraud", *Applied Cognitive Psychology*, Vol 11, 1997, pp 211-222

[9] P J B Hancock, V Bruce, A M Burton, "Recognition of unfamiliar faces", *Trends in Cognitive Sciences*, Vol 4, 2000, pp 330-337

[10] N F Troje, D Kersten, "Viewpoint dependent recognition of familiar faces", *Perception*, Vol 28 February 1999, pp483-487

[11] E Tulving, *Elements of episodic memory*, Clarendon Press, 1985

[12] F L Krouse, "Effects of pose, pose change and delay on face recognition performance", *Journal of Applied Psychology*, Vol 66, pp 651 - 654

[13] V Bruce, T Valentine, A D Baddeley, "The basis for the ¾ view advantage in face recognition", *Applied Cognitive Psychology*, Vol 1, 1987, pp 109-120

[14] N F Troje, H H Bülthoff, "Face Recognition Under Varying Poses: The Role of Texture and Shape", *Vision Research*, Vol 36, No 12, 1996, pp 1761-1771

[15] H Hill, V Bruce, "The effects of lighting on the perception of facial surfaces", *Journal of Experimental Psychology: Human Perception and Performance*, Vol 22, 1996, pp 986 - 1004

[16] C H Liu, C A Collin, A Chaudhuri, "Does face recognition rely on encoding of 3-D surface? Examining the role of shape-from-shading and shape-from-stereo", *Perception*, Vol 29, 2000, pp 729-743

[17] C H Liu, A Chaudhuri, "Face recognition with perspective transformation", *Vision Research*, Vol 43, 2003, pp 2393-2402

[18] J R Hyde, J A Robinson, "Coding 3D facial models for mugshot applications", *Vision, Video and Graphics*, 2003, pp 127-133

[19] J H Friedman, "Regularized Discriminant Analysis", *Journal of the American Statistical Association*, Vol 84 No 405, March 1989, pp 165-175.

[20] B-C Kuo, D A Landgrebe, "A Covariance Estimator For Small Sample Size Classification Problems and its Application to Feature Extraction", *IEEE Transactions on Geoscience and Remote Sensing*, Vol 40, No 4, April 2002, pp 814-819.