Technical Report

Roy Schestowitz

July 2, 2012

Abstract

Statistics of faces are intricate and delicate when posed as a task requiring the ability to discern identities. In 3-D, where the data is essentially range images void of any textural information, the task is further complicated, especially in the presence of facial expressions. This work explores our ability to carry out verification tasks, where various tools are tested and compared, ranging from intensity-based PCA (superimposed upon range images) to more sophisticated algorithms such as Generalised Multi-Dimensional Scaling (GMDS). We find that the performance reaches its peaks at around 97% for particular datasets and hovers at the range of 90-95% for FRGC data, depending on the method being assessed. The work is unique in the sense that it thoroughly but not exhaustively explores what can be attained *without* any textural information, using tools which are claimed to have already been outperformed, e.g. by Al-Osaimi *et al.* We are unable to replicate results of such groups due to the proprietary nature of their data and methods, but by bringing together under the same roof several families of methods we make reasonably valuable benchmarks possible.

1 Introduction

Face identification (or verification) helps associate a person with his/her real identify when it cannot be reliably inferred from a document in one's possession. Authentication commonly requires that a person's biometric characteristics closely resemble a model or reference digitally stored – one corresponding to the same person. In order to reliably carry out this task, one can measure different attributes that are perceived to be immutable, e.g. resistant to illumination changes, pose-agnostic, and independent of the acquisition method. If for each person a large (but finite) set of surfaces can be considered a proper "match", then we wish to identify the subspace in which those surfaces lie. Only through good separation in a (very high) parametric space can one person never be confused with another. The problem is harder than typically realised, especially when depth information is all we are being provided. Since people's faces are topologically similar, it is the fine differences that tell apart one person from another and since the face is very morphologically flexible, there is no guarantee of a neutral expressions being presented all the time.

In Section 2, a very concise summary of past work is presented. In Section 3, a discussion of key concepts (and hypothesis) is outlined. Section 4 deals with data and a tiny portion of our practical work is presented in Section 5, emphasising only the best of our results. Section 6 concludes.

2 Past Work

Face recognition using Multi-dimensional Scaling goes a reasonably long way back. The basic idea is that expressions can be treated by learning their effect on the surface of a flattened face. Each expression can then be treated using isometries, which are an area explored by others too [10]. The surface of the face is deformed to a Canonical Form using Multi-Dimensional Scaling (MDS) such that geodesic distances between the points are preserved. This helps remove the impact of expressions on the surface in a different way than the one adopted with PCA for example. There is an extension to this work, which is known as Generalised Multi-dimensional Scaling (GMDS). Bronstein *et al.* used variants of such a non-rigid method to tackle the face recognition problem, whereas many others stick to rigid methods which preserve the geometry of the faces as they approach the recognition problem. GMDS can also be used in a wide range of other problems, including deformation-invariant comparisons, similarity of deformable shapes with partial similarity, and correspondence of deformable shapes.

In contrast to the work of Bronstein *et al.,* Faisal R. Al-Osaimi *et al.* [1] extend work on ICP to account for expressions and annul them. They propose "An Expression Deformation Approach" where PCA learns (so as to embody) variation induced by expressions and then warps each 3-D face onto an expression-neutral equivalent. The FRGC (v2.0) dataset is used to demonstrate good results, but some of the data used in these experiments is proprietary and we are not allowed access to it.

Figure 1: Expression parameterisation in action (image from Al-Osaimi *et al.*)

Figure 1 depicts the sorts of 3-D deformations used to alter expressions in a consistent fashion. Using similar tools or methodologies, ear-based verification of identity is demonstrated in [8]. In our work we are trying to learn those methodologies and compare them to GMDS.

3 Theory and Replication of Work

As the first step of our work we wish to test the aforementioned approach for ourselves and replicate those results. Unfortunately, however, the training with PCA required a very large database of faces that are proprietary and cannot be accessed. Instead, we used a collection of many faces from hundreds of different individuals and used those for training. Given that prior groups working on this sort of challenge were unable to get high verification rates (e.g. Russ *at al.* [13], Mena-chalco *at al.* [11], and Gervei *at al.* [7] with rates of 83.3% for 540 3-D images from 60 individuals) we were hoping to reach success rates of at least 90% but were not overly optimistic. At our disposal, initially, we have PCA. Disney works on PCA for animation these days [14] and other groups working on verification in 3-D reported to have approached rates of about 85% (verification) several years ago, using piece-by-piece PCA alone.

PCA and GMDS have some clear similarities at some lower level. In MDS and GMDS we treat shapes in a metric space and assume that shape similarity can be reliably measured in terms of the distance between metric spaces. The duality of this problem can be outlined visually. To conceptualise this, it might look akin to Figure 2.

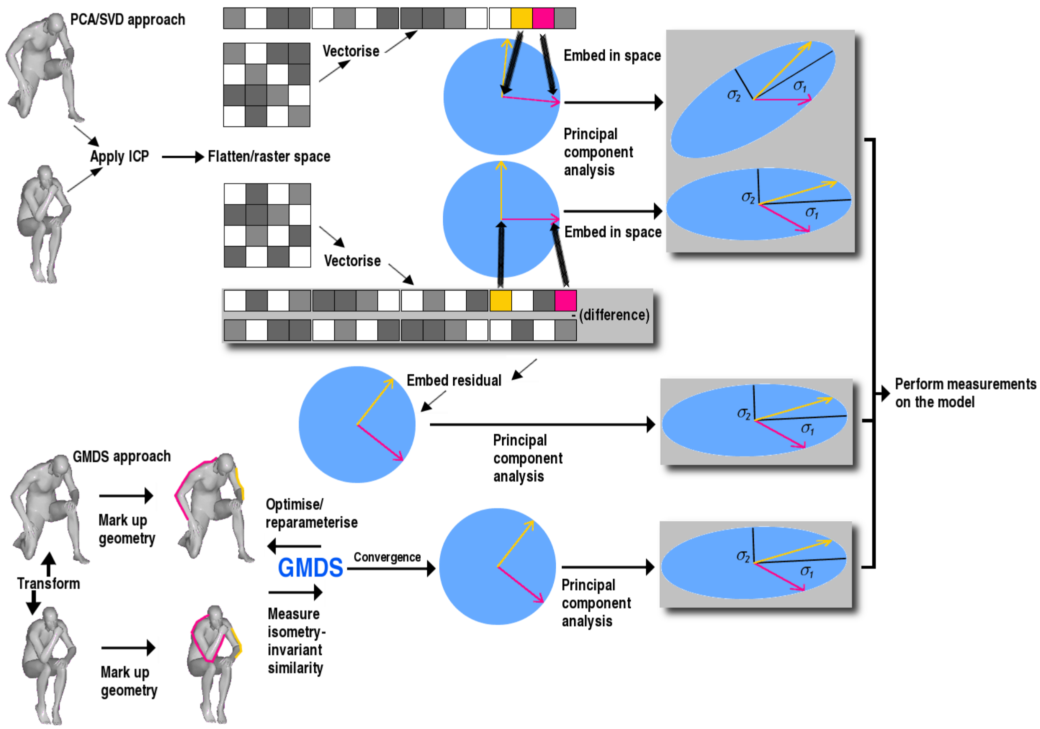

Figure 2: The proposed framework for GMDS improvement

It is possible to treat it somewhat differently, e.g. use ICP for rough alignment, GMDS for intrinsic fine alignment. If the alignment is onto a generalised face, then spectral decomposition takes place in the refined space – an Eigen functions of the generalised face. We need to address those issues, as shown in Figure 3.

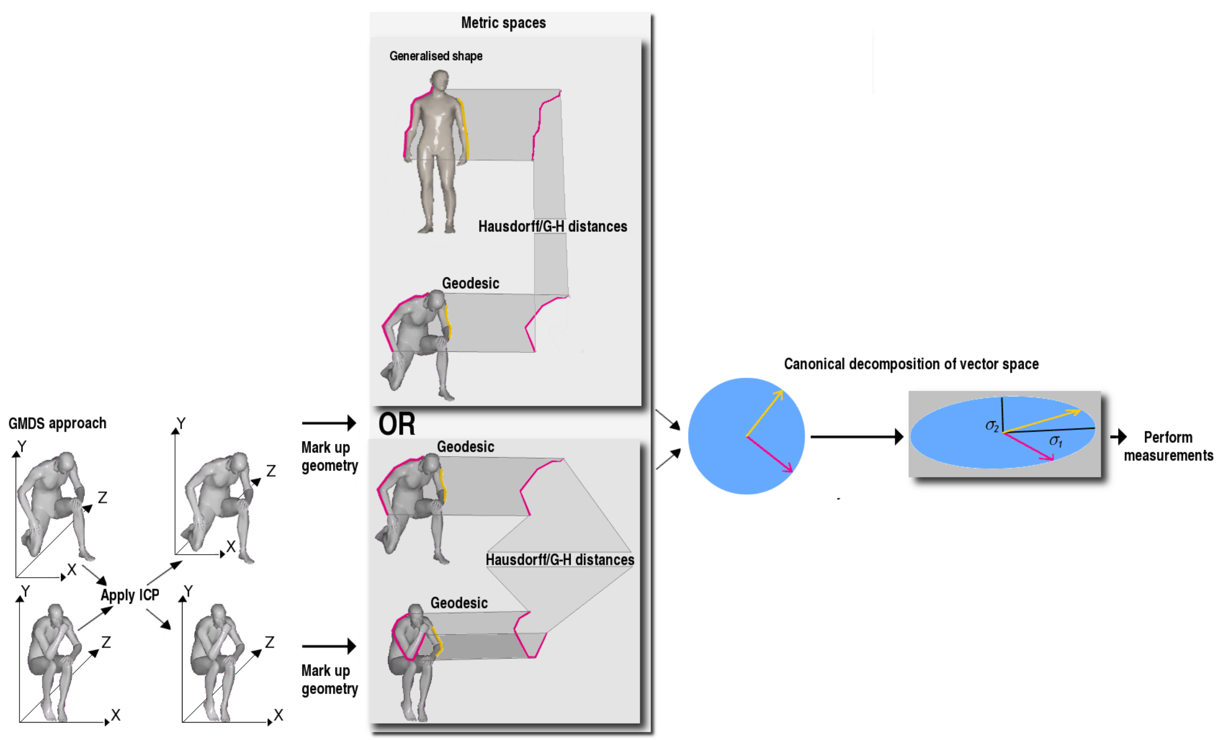

Figure 3: A closer look at the GMDS approach

ICP could be considered Gromov-Hausdorff (G-H) in Euclidean space. The G-H-inspired method strives to identify and then calculate minimal distances for a group of geometric points with commonality in a more rigid space, wherein harmonic variation occurs in inherently non-orthogonal spaces. One way to model this type of variation and then explain its nature would be high-dimensional decomposition, which evidently requires that data be represented in a high-dimensional form such as vector of coordinates, intensities, energy, or discrete/quantised G-H distances (geometric terms). Figure 4 provides an example of that subtle point.

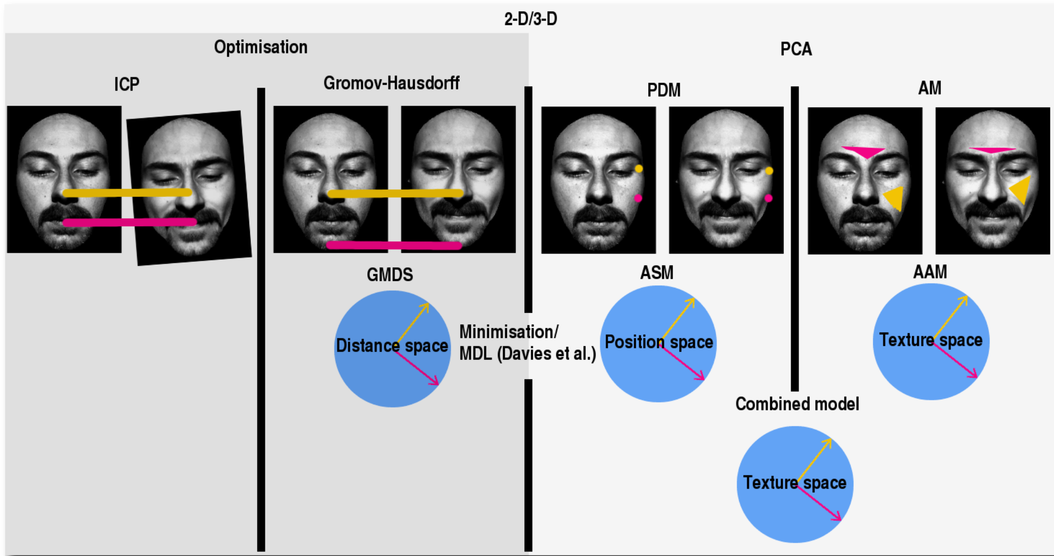

Figure 4: Crude visual example of how typical PCA and GMDS relate to one another, approach-wise

Depending on the circumstance, different measurable attributes can be added to the space, even a hybrid of them (e.g. shape and texture, so as to reconstruct/recover the relationship between image intensity and the image shape in 2- and 3-D). For synthesis of images belonging to a particular class/subspace, e.g. a canonical form (bar embedding error), one requires that the model should be specific and generic. Specific – for the fact that it need preferably not be confused with similar images belonging to another class, and generic – for the fact that it must span a sufficiently large cloud in hyperspace in order to capture the variation of all images of the same class.

When it comes to sampling geodesic distances for PCA, furthest point sampling would do the job of meaningful sampling[12]. It is 2-optimal in sense of sampling. $d_{GH(S,Q)}$ where distances are Euclidean is like ICP.

4 Data

In the remainder of the article and in all subsequent sections, results of experiments are presented which not only use FRGC data but also the T3FRD database. An example range image from FRGC is presented in Figure 5 or Figure 6 where it is shown in 3-D.

Figure 5: Example range image from the FRGC dataset

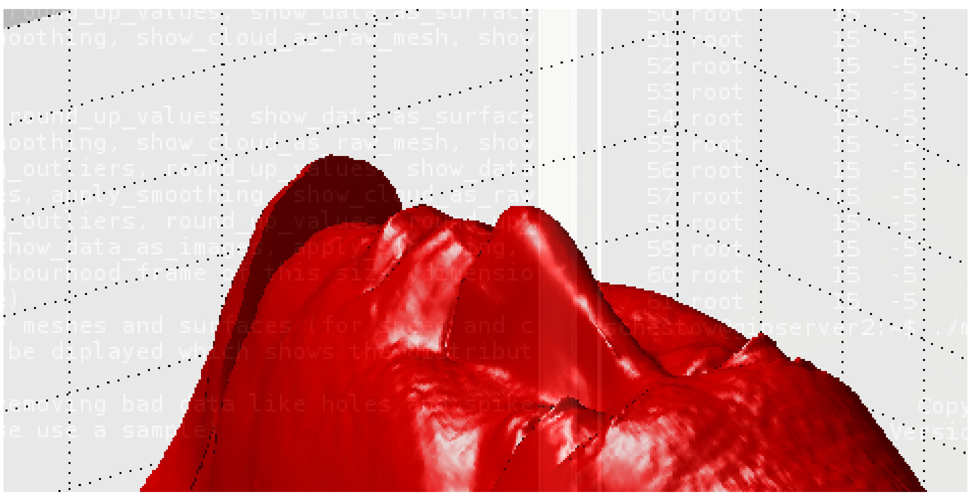

Figure 6: A 3-D representation of an example image from the FRGC dataset

Although we have used 4 types of datasets, including synthetic data at one point, the experiments that follow use just the standard sets. We begin by experimenting with PCA on faces whose 3-D difference image looks like in figures 7 and 8.

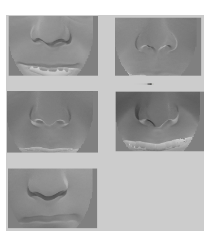

Figure 7: Difference images of the first 5 pairs taken from the same people

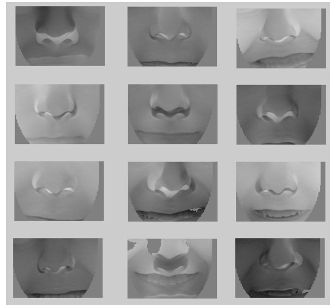

Figure 8: Difference images of the first 12 pairs taken from different people

For the sake of the experiments being consistent with past work, in most of the experiments we carve out only a subset of all faces, as shown in Figure 9, leaving only more rigid parts of the face.

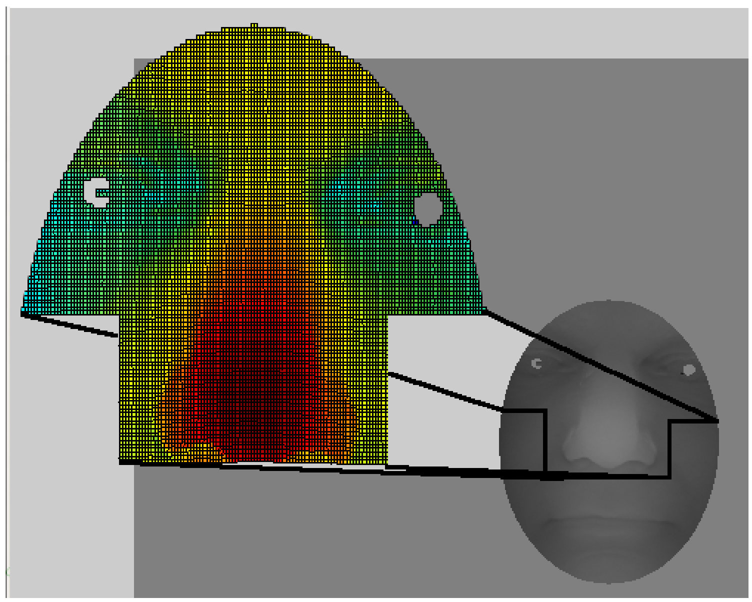

Figure 9: A slice or subset of the data being used for ICP (on the left) and the masked face from which it is extracted (right)

5 Experimental Framework

The framework constructed over the first couple of months can be seen in Figure 10.

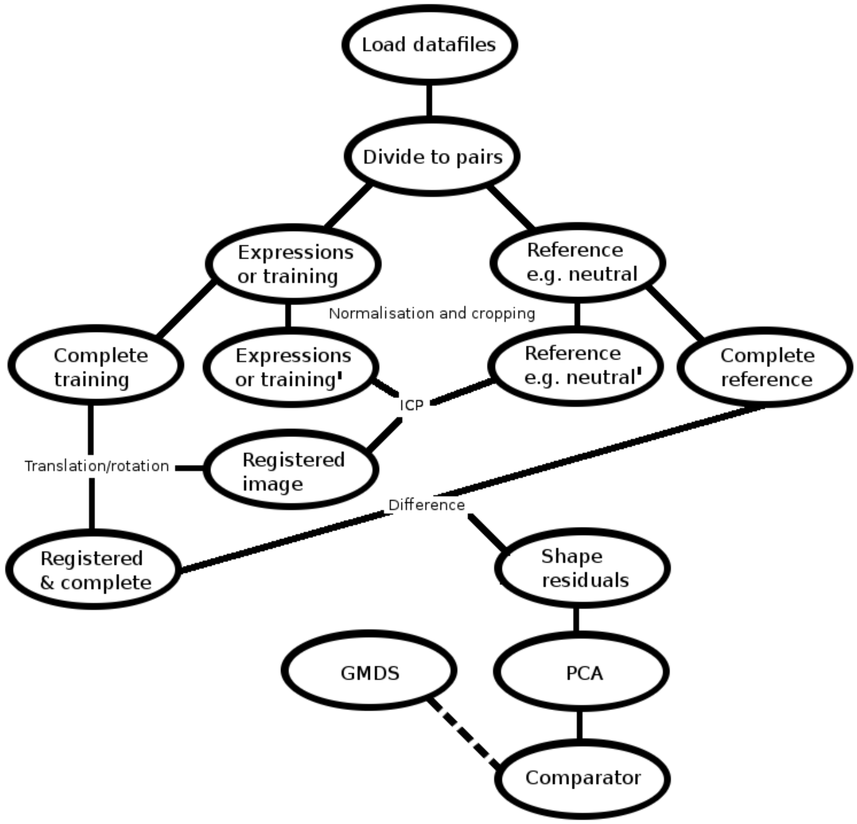

Figure 10: Program steps broken down into an overview-type flowchart

For the time being we concentrate on PCA and not GMDS. The key part is about PCA [9], which MATLAB implements with princomp; it is essential for constructing statistical models, via decomposition of face characteristics as derived automatically from the dataset.

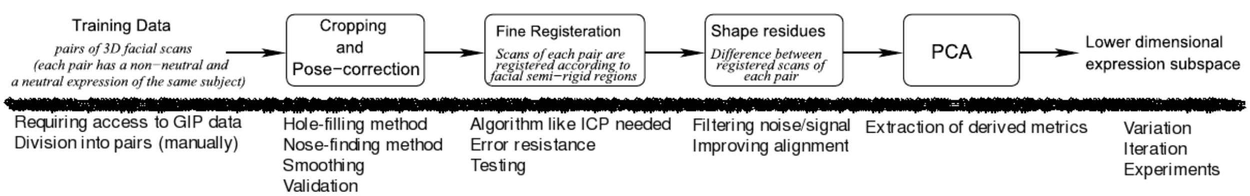

Figure 11: A replotted block diagram of the components in the IJCV paper [1] and our proposed extension/modifications (bottom), and the already-implemented procedures

Shown in Figure 11 is an annotated version of the original figure from the paper. The overview is simplistic in the interests of abstraction and elegance. The graphical user interface of the program we built looked like Figure 12 about a year ago (there have been many feature enhancements since).

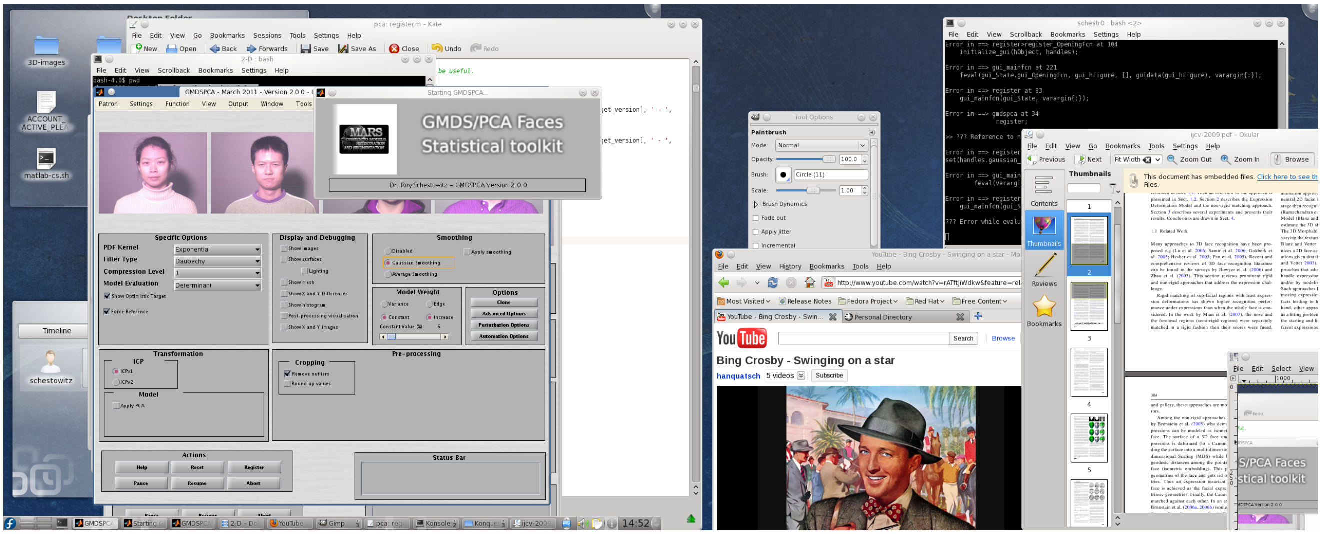

Figure 12: Early prototype of the GUI

Using the PCA-based approach, we were initially able to get results of reasonable standard. Figure 13 shows the ROC curve that accounts for hundreds of image pairs from the FRGC dataset.

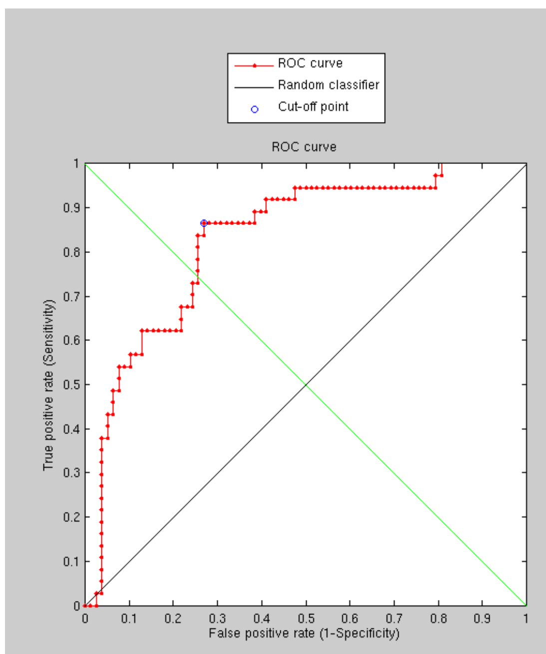

Figure 13: ROC curve showing the PCA-based approach with its performance on 120 image pairs

In a separate experiment, the images were downsampled by a factor of 10 along each dimension, lowering by two orders of magnitude the Z axis data that gets sampled by PCA based on a grid. This ought to have kept the models more manageable for the purpose of algorithm/performance testing. Interestingly enough, downsampling hardly affected the ability to recognise faces. As Figure 14 shows, the classification remained almost the same, even though the images were tiny (see Figure 14 at the top left).

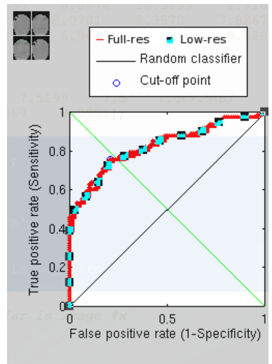

Figure 14: The results from full-resolution image sets and low-resolution equivalents (as seen at the top left)

Using the same data and preprocessing as before (for the sake of a sound comparison), we have applied a PCA-based approach to get the following results, which are clearly by far superior. The variation incurred by expression is detected by PCA in the sense that it is not seen as a new type of variation.

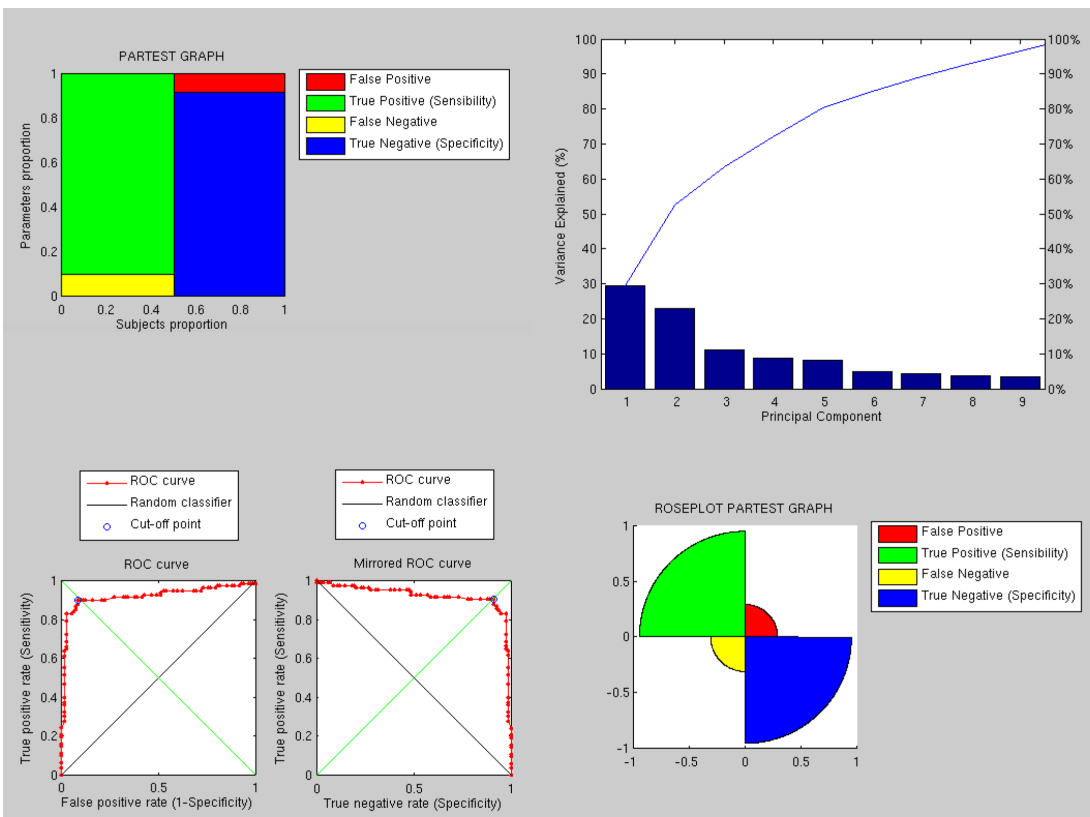

Figure 15: The FP analysis of the results of a model-based approach, with the breakdown of modes shown at the top right

The performance, as shown in Figure 15, is therefore greatly improved and there is room for further improvements as this implementation uses tiny images to save time and it does not use the sophisticated approaches partly implemented by now.

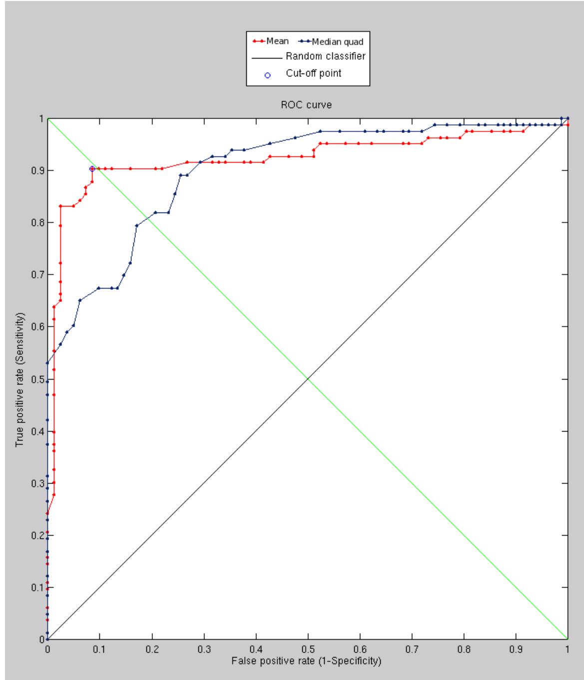

Figure 16: Performance comparison between an approach where the median of squared differences gets compared to mean of model changes

Using GMDS, the recognition performance reached at the early stages was around 90%. There were limitations related to image resolution and algorithms whose performance does not degrade linearly. These experiments took a long time to set up and run (manually) because of some freezing and stability issues at a resolution which translates to 4,600 vertices. The combination with C++ code for stress minimisation improved speed. The ROC curves in Figure 17 helps show the impact of the number of vertices on performance.

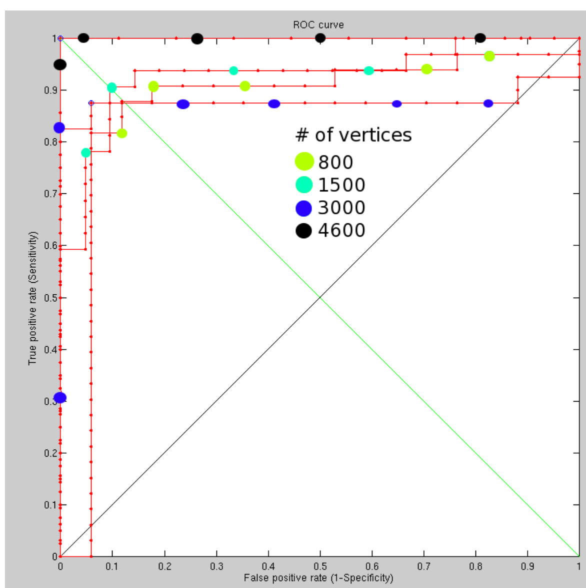

Figure 17: A set of verification results obtained by systematically varying the number of vertices used

Areas of mismatch have been studied more closely in order to understand what causes them. Several large images were looked at along with the GUI (with previews of images enabled), showing quite clearly that we must remove hair from the surfaces as many mis-detections are caused by this. The aim is to get close to 99% recognition rate. The sample size is not large enough for an sufficiently informative ROC curve, but there are only a few wrong classifications. One is a borderline case where pairs from different people almost seem like belonging to the same person. The other case is mostly a case of GMDS not working, not quite a wrong classification. At all resolutions attempted so far, one pair of faces (same person imaged) cannot be made correspondent. Other than that, there is almost an order of magnitude apart in terms of separation between correct pairs and incorrect pairs. One important issue to tackle is the rare case where GMDS hardly latches onto facial features at all, as shown in Figure 18.

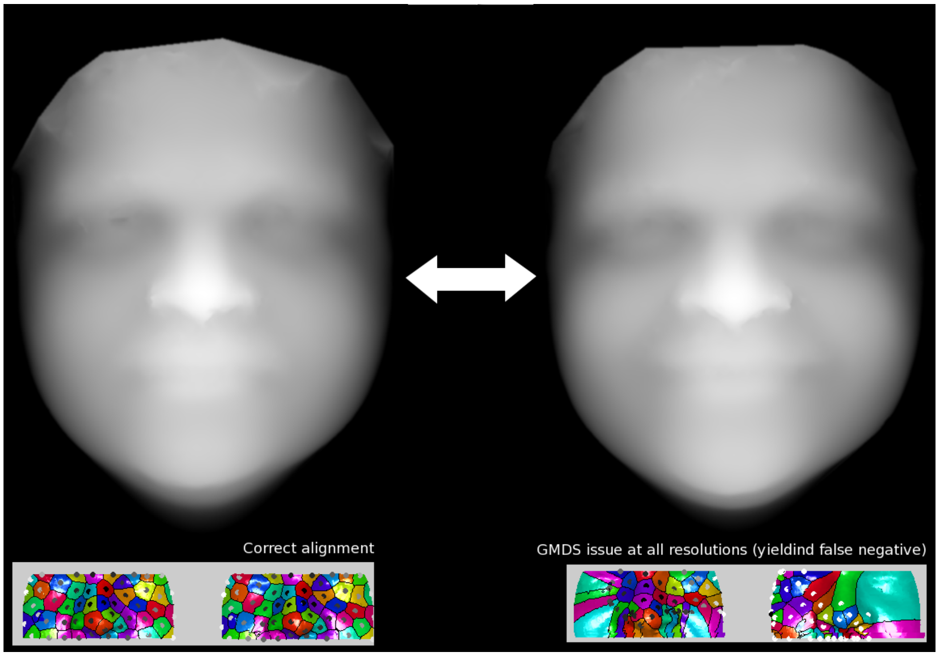

Figure 18: A pair that causes GMDS to fail

A GMDS-based identity verification task, with smoothed surfaces where the resolution is increased for accuracy and for improved performance, still works rather well (room remains for improvement). In the following late experiment only one image was problematic, only slightly bordering the threshold because of pose variation on the face of it (see Figure 19). There was only one case where GMDS failed. The ROC curve is shown in Figure 20 and it is the best performance level that we reached using this method.

Figure 19: A problematic pair which is seen as too different to quality as a match

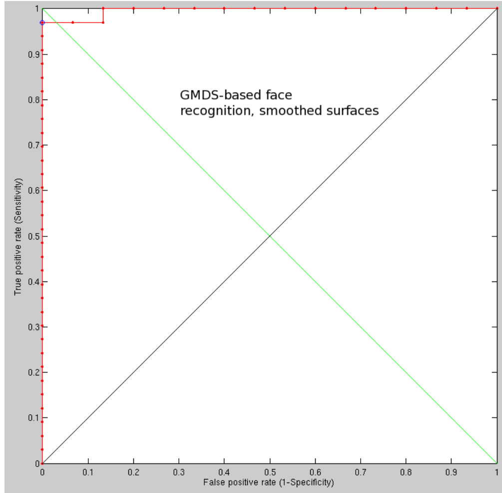

Figure 20: ROC curve based on the smoothed surfaces variant of the algorithm

That last curve was obtained by using a kernel/window 13 pixels across, moving average (horizontal and vertical). The 2-D Gaussian filter is another option.

A newly-worked-on approach strives to measure distances between images in hyperspace based on their parameterised version, where these parameters are basically a small set of distances, each (hopefully) encompassing a sort of concise digital signature corresponding to a person's facial surface alone. The sketch below shows the approach. It is a brute force implementation that measures many geodesic distances and then compares surfaces based on distance-to-distance subtractions. It is not particularly clever, but the results of recognition tests are not too bad, either. They help validate the premise that by measuring Euclidean distances in XY, YZ, and XZ (based upon geodesic operators like FMM) we are able to carve out the surfaces and extract meaningful measures from the sub-surfaces.

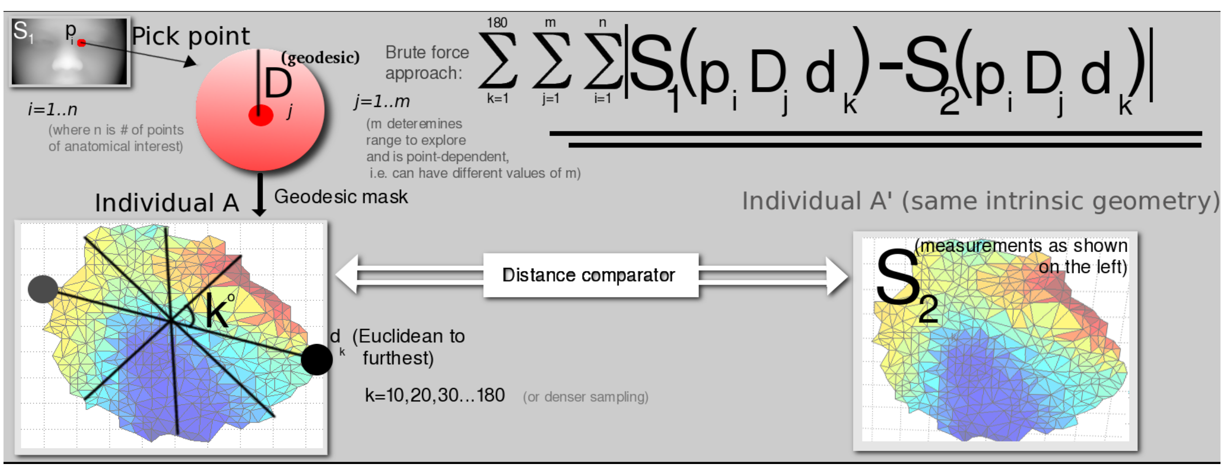

Figure 21: Brute force implementation that measures many geodesic distances

Another figure, Figure 22, shows the next step.

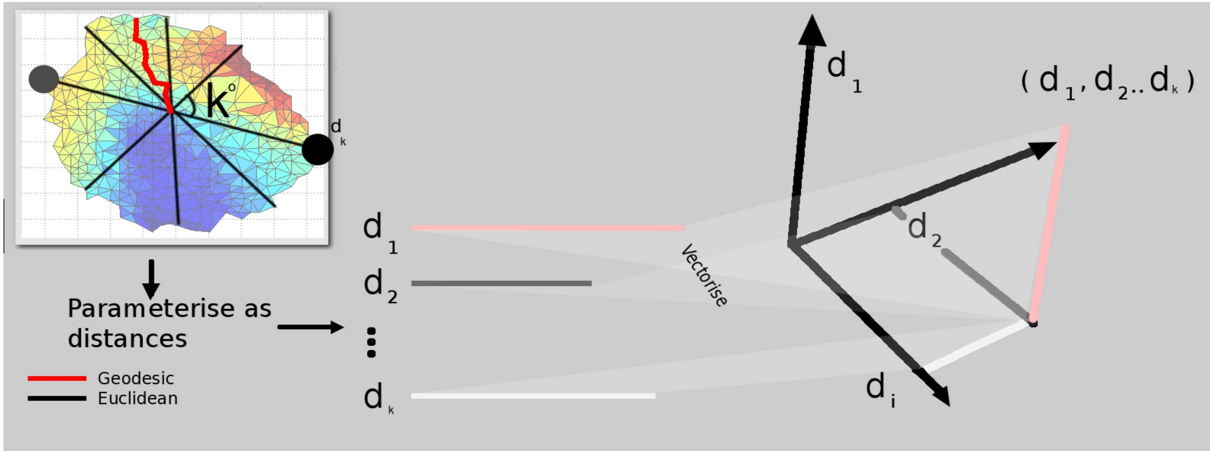

Figure 22: This figure visualises the idea of encoding surfaces as a vector not of surface vertices but an ordered list of Euclidean-upon-geodesic distances, which are fast to compute and sensitive to isometric/mildly detectable alterations

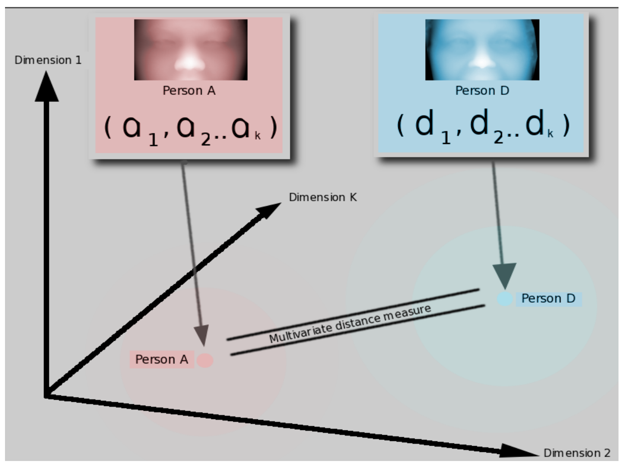

Figure 23: Separability testing in hyperspace

Taking the first imaged individual vs different imaged individuals (92 different individuals), the following results are obtained using the new method, which was refined and adjusted to the task at hand (Figure 24), unlike GMDS which is generic and adaptable.

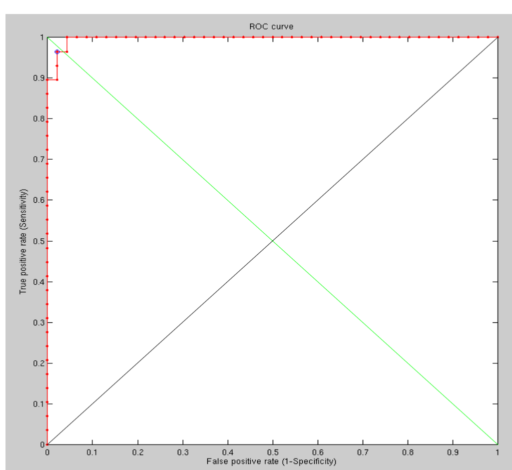

Figure 24: ROC curve obtained by measuring geodesic-Euclidean distances on the first imaged individual vs the same on different individuals

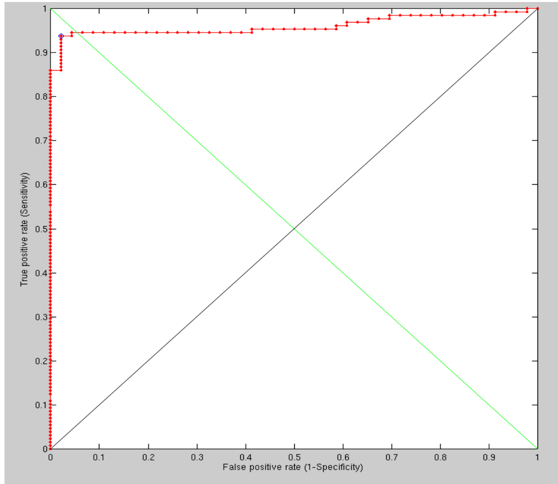

Figure 25: The result of running the test set further (not for comparative purposes)

We are able to see slight improvement incurred by the use of smoothing in the new FMM-based method. It makes sense to do this around the eyes, but currently the filter is applied uniformly to the entire image. At present, GMDS continues to show potential (more so than PCA), but its performance falls short of the modified algorithm detailed above.

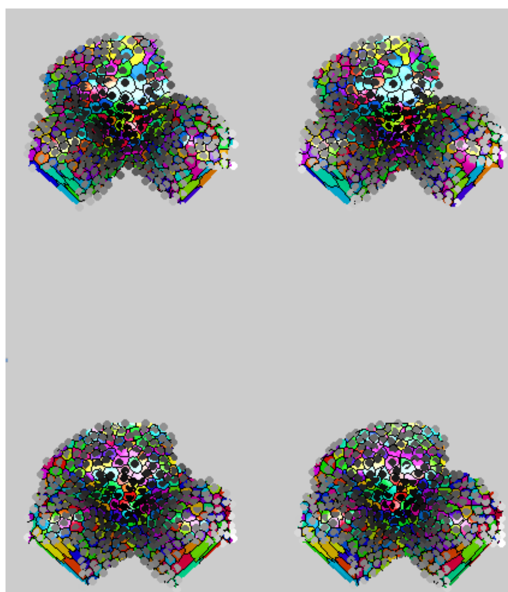

Figure 26: Example of correct matching between pairs of faces, where geodesic distances are being used to carve out a subsurface for each face

6 Summary and Conclusions

Verification tasks that are limited to geometric data (in 3-D) remain challenging. After more than a year exploring and studying this problem, it seems safe to say that diffusion-based methods are less suitable than geodesic distances for the task of identity validation and while PCA-based methods show promise, they are outperformed by GMDS and surrogate methods, at least within our experimental framework which primarily examines two publicly available datasets. Further work could explore the effect of measuring exact geodesics, combining several different classifiers, and maybe fusing photometric data.

Acknowledgements: the project was funded by the European Research Council.

References

[1] F. Al-Osaimi, M. Bennamoun, and Ajmal Mian, “An Expression Deformation Approach to Non-rigid 3D Face Recognition," Int’l Journal of Computer Vision (IJCV), vol. 81(3), pp. 302–316, 2009. 2, 7

[2] A. Bronstein, M. Bronstein, and R. Kimmel, “Three-dimensional face recognition. International Journal on Computer Vision,” 64(1), pp. 5–-30, 2005.

[3] A. Bronstein, M. Bronstein, and R. Kimmel, “Generalized multi-dimensional scaling: a framework for isometry-invariant partial surface matching. Proceedings National Academy of Sciences (PNAS),” 103(5), pp. 1168-–1172, 2006.

[4] A. Bronstein, M. Bronstein, and R. Kimmel, "Calculus of non-rigid surfaces for geometry and texture manipulation," IEEE Trans. Visualization and Computer Graphics, vol. 13, pp. 902–913, 2007.

[5] A. Bronstein, M. Bronstein, and R. Kimmel, “Expression-invariant representations of faces,” IEEE Trans. Image Processing, vol. 16(1), pp. 188–197, 2007.

[6] A. Bronstein, M. Bronstein, and R. Kimmel, “Generalized multidimensional scaling: a framework for isometry-invariant partial surface matching,” Proc. National Academy of Sciences (PNAS), vol. 103(5), pp. 1168–1172, 2006.

[7] O. Gervei, A. Ayatollahi, N. Gervei, “3D Face Recognition Using Modified PCA Methods,” World Academy of Science, Engineering and Technology, 63 2010. 2

[8] S. M. S. Islam, R. Davies, M. Bennamoun, and A. S. Mian, “Efficient Detection and Recognition of 3D Ears," Int. Journal of Computer Vision, 2011. 2

[9] I.T. Joliffe, “Principal Component Analysis,” Springer Series in Statistics, Springer, New York, 1986.

7

[10] F. Mémoli and G. Sapiro, “A theoretical and computational framework for isometry invariant recognition of point cloud data. Foundations of Computational Mathematics,” 5(3), pp. 313–-347, 2005. 2

[11] J. Mena-chalco, I. Macedo, L. Velho, and R. Cesar-Jr,“PCA-Based 3D Face Photography,” Brazilian Symposium on Computer Graphics and Image Processing, pp. 313—320, 2008. 2

[12] C. Moenning and N. Dodgson, “Fast Marching farthest point sampling,” 2003. 4

[13] T. Russ, C. Boehnen, and T. Peters, “3D Face Recognition Using 3D Alignment for PCA,” Computer Vision and Pattern Recognition (CVPR), vol. 2, pp. 1391–1398, 2006. 2

[14] J. R. Tena 1 F. De la Torre, and I. Matthews, “Interactive Region-Based Linear 3D Face Models,” ACM Transactions on Graphics (Proc. ACM SIGGRAPH), Aug. 2011 (to appear).

2