

Learning a Manifold as an Atlas*

Supplementary Material

Nikolaos Pitelis

Chris Russell

Lourdes Agapito

School of EECS, Queen Mary, University of London

[nikolaos.pitelis, chrissr, lourdes]@eeecs.qmul.ac.uk

1. Introduction

In the supplementary material we provide the exact form of the objective function used in the unwrapping of manifolds for visualisation. We also include additional examples of unwrapped manifolds, high-dimensional data visualisation, and more detailed results of our classification experiments. More specifically, Section 2 explains the mapping from the original high-dimensional space into a low-dimensional one, suitable for data visualisation, and shows further examples of manifold unwrapping for the visualisation of synthetic datasets and face datasets (ISOfaces and Frey faces); Section 3 shows more detailed classification results on the Semeion dataset and additional results on the Yale Faces dataset on the data provided by [4].

2. Manifold unwrapping for visualisation

Unlike existing approaches to manifold learning, our method has no requirement to unwrap the manifold, and after characterising it as an atlas, we can immediately perform classification or 3D reconstruction (see main paper). However, one of the principal uses of manifold learning [2, 3, 5] is in creating a mapping from a high dimensional space \mathbb{R}^n into a low-dimensional one \mathbb{R}^d , suitable for visualising data. As a way of illustrating our method's robustness to sparse data, the presence of noise and to systematic holes we illustrate 2D unwrapping on standard datasets. The focus on $d = 2$ in the mathematical formulation is for the sake of exposition, but the approach generalises to any choice of target dimensionality d , with Fig. 1 and 5 showing unwrapping of 3D manifolds.

Given a partitioning of the data into overlapping charts isomorphic to \mathbb{R}^2 , it is possible to write down an objective function. For each affine chart c_i , we wish to find an affine mapping \mathbf{E}_i from $\mathbb{R}^2 \rightarrow \mathbb{R}^n$ into the original space in such a way that a point is projected into the same location by each chart that it belongs to (2); the centres of each chart are far from one another (3); the affine projection is not degenerate

and distinct points in the chart do not collapse on top of one another (4); and the reconstruction is as flat as possible, with most of the coefficients that do not correspond to a projection into the plane being small (5). As the objective we are minimising is non-convex, we unwrap the manifold in the original embedding space, where the extra degrees of freedom make it less likely that we will get stuck in a poor quality local minimum.

More formally, given a point p we use \mathbf{p}_i to refer to its projection into chart c_i ,

$$\mathbf{p}_i = \left[\mathbf{C}_i(\mathbf{p} - \mu_i) \right], \quad (1)$$

where μ_i , \mathbf{C}_i are the mean and orthonormal vectors of projection from the embedding space into chart c_i . Writing \mathbf{E}_i for the $n \times (2+1)$ affine embedding of chart c_i into \mathbb{R}^n , and \mathbf{E} for the set of all \mathbf{E}_i we minimise the following objective:

$$\arg \min_{\mathbf{E}} \sum_{p \in \mathcal{P}} \sum_{i,j \in \mathcal{X}_p} w_{\mathcal{O}_{i,j}} \|\mathbf{p}_i^T \mathbf{E}_i - \mathbf{p}_j^T \mathbf{E}_j\|_2^2 \quad (2)$$

$$+ w_{\mathcal{A}} \sum_{i,j \in \mathcal{A}} \|\mathbf{E}_i^{[h,1]} - \mathbf{E}_j^{[h,1]}\|_2^{-2} \quad (3)$$

$$+ w_d \sum_{i \in \mathcal{A}} (\mathbf{E}_i^{[h,2]} \mathbf{E}_i^{[h,3]T})^2 \quad (4)$$

$$+ \sum_{i \in \mathcal{A}} \|\mathbf{E}_i^{[v,3]}\|_F^2, \quad (5)$$

where $\mathbf{E}_i^{[h,k]}$ refers to the k^{th} horizontal row of matrix \mathbf{E}_i and $\mathbf{E}_i^{[v,k]}$ refers to the k^{th} vertical column. The normalising constants $w_{\mathcal{O}_{i,j}}$, $w_{\mathcal{A}}$, and w_d guarantee that each term contributes about the same to the cost function regardless of the size of the overlap, the number of charts, or their local dimensionality respectively. Specifically,

$$w_{\mathcal{O}_{i,j}} = \frac{1}{\mathcal{O}_{i,j}}, \quad (6)$$

where $\mathcal{O}_{i,j}$ is the number of points that belong to both charts c_i and c_j ,

$$w_{\mathcal{A}} = \frac{2}{|\mathcal{A}| - 1}, \quad (7)$$

*This research was funded by the European Research Council under the ERC Starting Grant agreement 204871-HUMANIS

where $|\mathcal{A}|$ is the total number of charts in the atlas, and

$$w_d = \frac{2}{d(d-1)}. \quad (8)$$

We initialise \mathbf{E} as the embedding learnt in section 2.2 in the main paper with

$$\mathbf{E}_i = \begin{bmatrix} \mu_i^T \\ \mathbf{C}_i \end{bmatrix}. \quad (9)$$

As the final reconstruction is not exactly 2-dimensional, we apply PCA to the unwrapped data and extract the two main components. Our approach readily generalises to a mapping from an embedding space of any dimensionality and to any choice of target dimensionality, by adjusting terms (4) and (5) to account for the additional rows and columns.

Toy datasets In the main paper we have illustrated the failure of existing manifold learning methods (with optimally tuned parameters), versus the success of our method in unwrapping two manifolds: the Swiss roll dataset with noise and the Gaussian without noise. Figure 2 shows additional unwrapping results on the Swiss roll dataset without noise and the Gaussian with noise. Even in the absence of noise most methods fail to deal with the geometry of Swiss roll. While Hessian LLE and LTSA succeed at unfolding the manifold they do not recover its aspect ratio. Figure 3 shows the effect of creating a hole in the Swiss roll (with and without noise). Notice how only Atlas succeeds in the presence of noise in the data. The unwrapping of the Twin-peaks manifold is shown in Figure 4. Although the peaks are not as sharp as in the case of the Gaussian, no method does better than ours in dealing with the non-uniformity of the sampling and preserving the rectangular shape of the plane.

ISOMAP faces This dataset, from the original ISOMAP paper [3], consists of 698 synthetically created 64×64 greyscale face images under different pose and illumination conditions. Specifically, there are three degrees of freedom: horizontal and vertical orientation, and lighting direction, with the vertical orientation having a smaller scale than the other two. Figure 5 shows the visualisation results for all methods that produced an informative result. The parameters for our method were $\lambda = 10^3$ and $k = 5$. Notice the success of our method at unfolding the manifold and preserving the correct scaling.

Frey faces Figure 1 shows the 3D visualisation of the Frey faces dataset, which consists of 1965 20×28 greyscale face images taken from a video. The variations in the data include changes in facial expressions and head rotations along the horizontal and vertical axes. These changes of orientation are captured along the horizontal and vertical axes

Table 1. Yale Faces dataset classification error (%) of NN classifier on the uncorrected data provided by [4]. The classification error of NN in the original space is 5.05, and the lowest classification error of PCA NN is 5.04 for $d = 299$. For Atlas $\lambda = 10^5$ and $k = 4$, while for Atlas+ $\lambda = 800$, $k = 3$, and $\theta = 1100$. #c denotes the number of subspaces found.

	PCA	LLE C&E [4]	LTSA C&E [4]	wLTSA C&E [4]	Atlas	Atlas+
d	$\epsilon(\%)$	$\epsilon(\%)$	$\epsilon(\%)$	$\epsilon(\%)$	$\epsilon(\%)$ #c	$\epsilon(\%)$ #c
8	47.12	5.93	3.88	3.01	3.04 64	1.88 53
9	43.30	6.85	4.40	3.22	2.98 65	1.94 55
10	39.54	6.59	4.18	3.13	2.92 65	1.93 55

of the figure while similar expressions are captured by each chart, showing spatial correlation in the low-dimensional space. The parameters for our method were $\lambda = 10^2$ and $k = 5$.

3. Classification

Extended Yale Face Database B As discussed in the main paper, careful examination of the data provided by [4] revealed an “off by one” labelling error in 37 out of the 38 classes (subjects). Nevertheless, for fair and complete comparison, we ran experiments on the uncorrected data and verified that the labelling error results in approximately 1% higher error for NN in the original space and NN preceded by PCA, Atlas, or Atlas+. Table 3 (which corresponds to Table 1(a) in the main paper) shows the scores for the Extended Yale Face Database B on the uncorrected data as provided by [4]. Atlas+ significantly outperforms LLE, LTSA, and wLTSA, while Atlas outperforms LLE, LTSA, and wLTSA, in all but one cases ($d = 8$), when it achieves similar performance to wLTSA. Again, for this Table, we kept the parameters of Atlas and Atlas+ fixed.

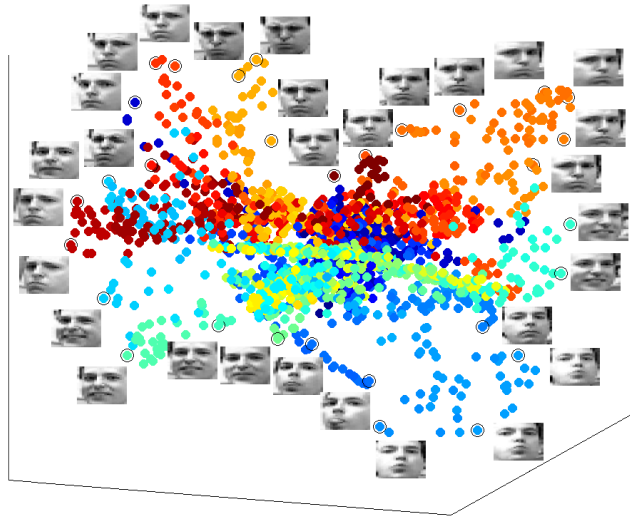


Figure 1. Frey faces Atlas 3D subspace unwrapping and visualisation. Different colours correspond to different subspaces.

Semeion In the main paper we showed classification results for $d = \{12, 14, 16, 18, 20\}$ (Table 2(a)) comparing with LLE, LTSA, and wLTSA from [4] and for $d = 21, 23, \dots, 33$ (Table 2(b)) comparing with PCA, LLE, and LTSA from our experiments. In this section we include Table 3, which contains the numerical values used to create Figure 3(a) in the main paper. Therefore this table shows extended values to those shown in the paper. SMCE [1] results are also included in the table.

References

- [1] E. Elhamifar and R. Vidal. Sparse manifold clustering and embedding. In *Advances in Neural Information Processing Systems*, pages 55–63, 2011. 3
- [2] L. Saul and S. Roweis. Think globally, fit locally: unsupervised learning of low dimensional manifolds. *The Journal of Machine Learning Research*, 4:119–155, 2003. 1
- [3] J. B. Tenenbaum, V. de Silva, and J. C. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 2000. 1, 2
- [4] Z. Zhang, J. Wang, and H. Zha. Adaptive manifold learning. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 34(2):253–265, feb. 2012. 1, 2, 3
- [5] Z. Zhang and H. Zha. Principal manifolds and nonlinear dimension reduction via local tangent space alignment. *SIAM Journal of Scientific Computing*, 26:313–338, 2002. 1

Table 2. Semeion Classification error for reduced dimensionality 1 to 50. The lowest error is chosen for each row for each method from our experiments. The classification error of Nearest Neighbour classifier in the original space is 10.92. The lowest classification error row-wise is highlighted in bold while the lowest column-wise error is underlined.

	PCA	LLE		LTSA		SMCE clust		SMCE NN		Atlas		
d	$\epsilon(\%)$	$\epsilon(\%)$	k	$\epsilon(\%)$	k	$\epsilon(\%)$	λ	$\epsilon(\%)$	λ	$\epsilon(\%)$	λ	k #c
1	74.44	35.45	2	53.03	11	-	-	-	-	18.52	10^1	2 93
2	61.57	28.58	2	39.52	5	28.56	210	75.80	210	15.16	10^{-1}	2 107
3	46.53	20.71	8	32.11	19	27.78	120	31.61	120	13.03	10^1	2 110
4	33.35	18.24	5	28.80	9	27.80	89	22.37	100	11.82	10^{-1}	2 122
5	25.90	15.84	5	23.94	42	27.57	77	18.53	61	10.83	10^{-1}	2 128
6	19.53	15.21	10	19.82	30	28.91	51	15.76	59	11.11	10^{-1}	2 134
7	17.07	13.75	9	17.28	36	29.87	46	12.44	43	11.59	10^{-1}	2 140
8	13.86	13.79	9	13.85	39	29.33	32	11.33	35	11.39	10^2	2 77
9	13.13	13.34	16	12.75	39	28.04	29	10.10	26	11.22	10^2	6 4
10	12.15	12.89	10	11.78	36	28.04	21	9.58	24	11.37	10^3	2 12
11	11.47	12.48	14	11.51	41	27.97	17	9.87	20	10.85	10^2	6 4
12	11.37	12.56	10	10.64	44	27.58	14	<u>9.26</u>	16	10.31	10^3	2 12
13	10.94	12.38	5	10.52	41	<u>27.24</u>	9	10.26	12	10.15	10^3	2 12
14	10.28	12.02	5	10.13	41	31.44	8	10.77	8	10.14	10^3	2 12
15	10.31	11.70	5	10.25	39	28.95	4	11.24	4	9.96	10^2	6 4
16	10.41	11.35	5	10.25	43	28.05	0.89	11.72	0.97	9.88	10^2	6 4
17	10.08	10.76	5	9.98	36	27.98	0.82	11.33	0.71	9.49	10^2	5 5
18	9.63	10.51	9	10.16	39	28.61	18	11.40	0.59	9.51	10^2	6 4
19	9.42	10.14	9	10.01	39	28.70	0.36	11.45	0.45	9.37	10^2	5 5
20	9.41	10.20	9	9.87	37	28.64	0.27	11.59	0.32	9.26	10^1	5 12
21	9.34	9.80	8	<u>9.56</u>	34	29.17	0.22	11.67	0.2	8.99	10^2	6 4
22	9.17	9.71	8	9.89	41	29.51	0.11	11.85	0.13	9.17	10^{-1}	15 1
23	9.23	9.61	8	9.89	33	35.55	0.05	12.02	0.05	8.88	10^1	5 12
24	9.11	9.56	8	9.80	39	-	-	-	-	8.50	10^2	5 5
25	8.97	9.73	9	9.75	39	-	-	-	-	8.62	10^2	5 5
26	8.96	9.68	16	9.74	43	-	-	-	-	8.70	10^2	6 4
27	8.77	9.57	8	9.86	40	-	-	-	-	8.61	10^2	6 4
28	8.82	9.49	8	9.79	40	-	-	-	-	8.58	10^2	6 4
29	8.74	9.41	16	9.75	42	-	-	-	-	8.48	10^2	6 4
30	8.74	9.49	8	9.95	43	-	-	-	-	8.44	10^2	6 4
31	8.66	9.54	8	10.16	43	-	-	-	-	<u>8.27</u>	10^2	6 4
32	<u>8.58</u>	9.60	8	10.20	40	-	-	-	-	8.45	10^2	6 4
33	8.67	9.69	16	10.29	44	-	-	-	-	8.50	10^2	6 4
34	8.60	9.54	15	10.25	42	-	-	-	-	8.41	10^2	6 4
35	8.67	9.43	15	10.14	47	-	-	-	-	8.57	10^2	6 4
36	8.62	9.43	15	10.30	49	-	-	-	-	8.62	10^2	6 4
37	8.65	9.36	8	10.23	49	-	-	-	-	8.62	10^2	6 4
38	8.85	9.28	7	10.37	47	-	-	-	-	8.78	10^1	5 12
39	8.81	<u>9.26</u>	7	10.28	48	-	-	-	-	8.73	10^2	6 4
40	8.85	9.40	16	10.54	46	-	-	-	-	8.76	10^2	6 4
41	8.85	9.40	16	10.66	48	-	-	-	-	8.64	10^2	6 4
42	8.96	9.47	16	10.37	50	-	-	-	-	8.74	10^1	5 12
43	8.84	9.50	16	10.52	50	-	-	-	-	8.74	10^1	5 12
44	8.94	9.54	15	11.02	50	-	-	-	-	8.73	10^2	6 4
45	9.00	9.60	15	10.98	50	-	-	-	-	8.83	10^2	6 4
46	9.08	9.63	15	11.69	50	-	-	-	-	8.87	10^2	6 4
47	9.13	9.56	14	13.30	50	-	-	-	-	8.84	10^1	5 12
48	9.06	9.65	14	34.78	50	-	-	-	-	8.91	10^1	5 12
49	9.06	9.67	7	81.44	49	-	-	-	-	8.95	10^2	6 4
50	8.97	9.68	15	81.29	50	-	-	-	-	8.97	10^{-1}	15 1

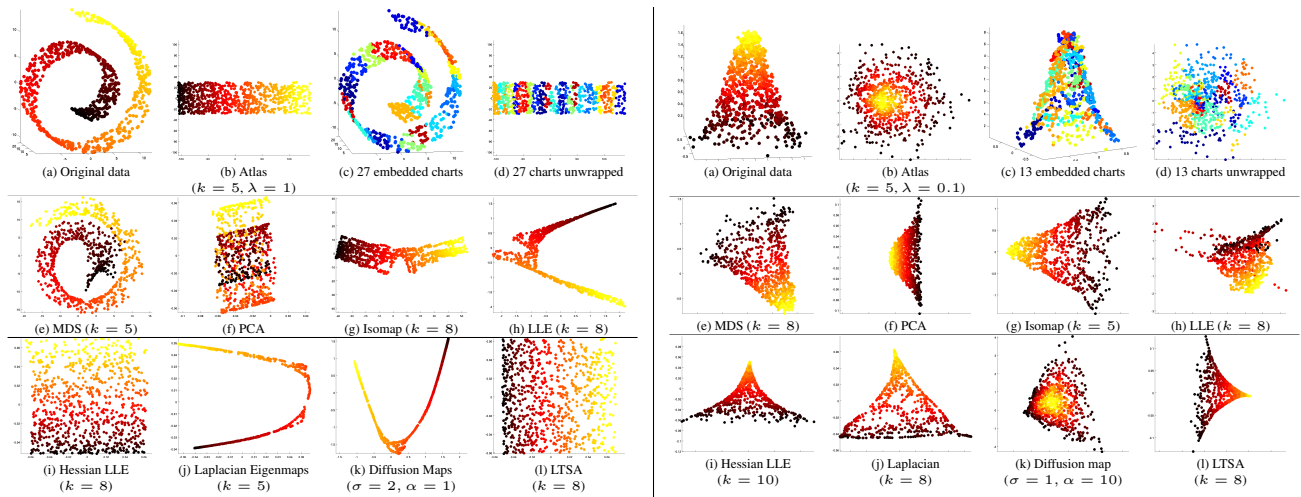


Figure 2. **Left:** Swiss Roll with no noise **Right:** Sharply peaked Gaussian manifold with Gaussian noise.

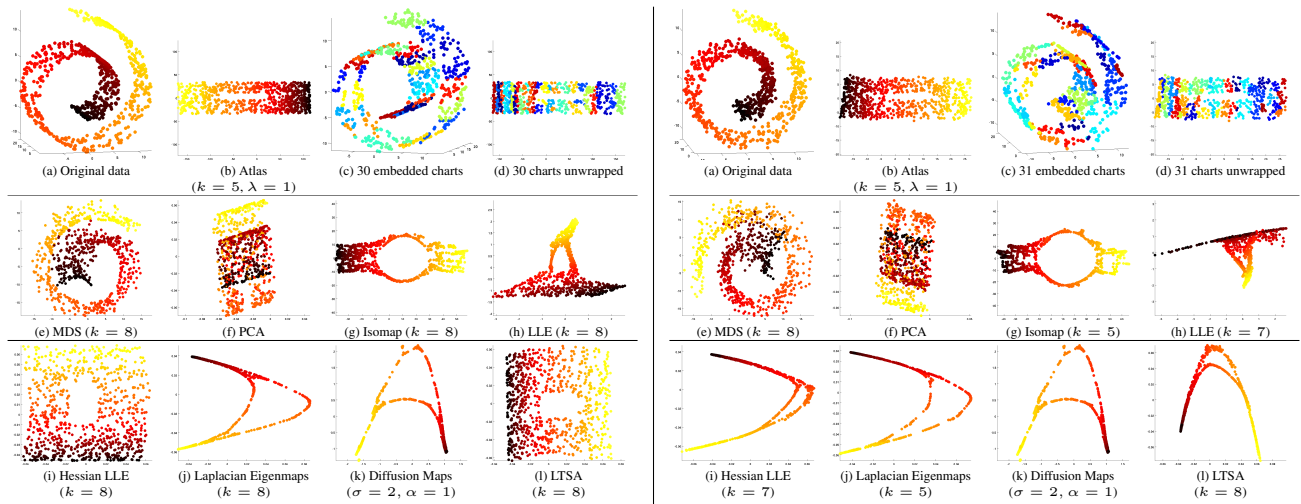


Figure 3. Swiss Roll with a hole **Left:** with no noise **Right:** with Gaussian noise.

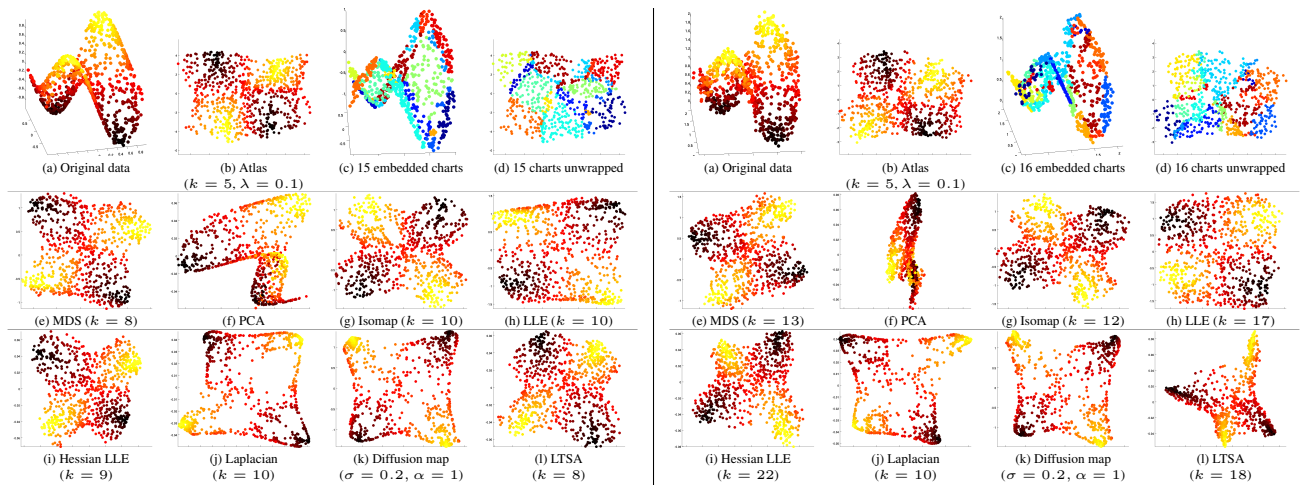


Figure 4. Twin Peaks **Left:** with no noise **Right:** with Gaussian noise.

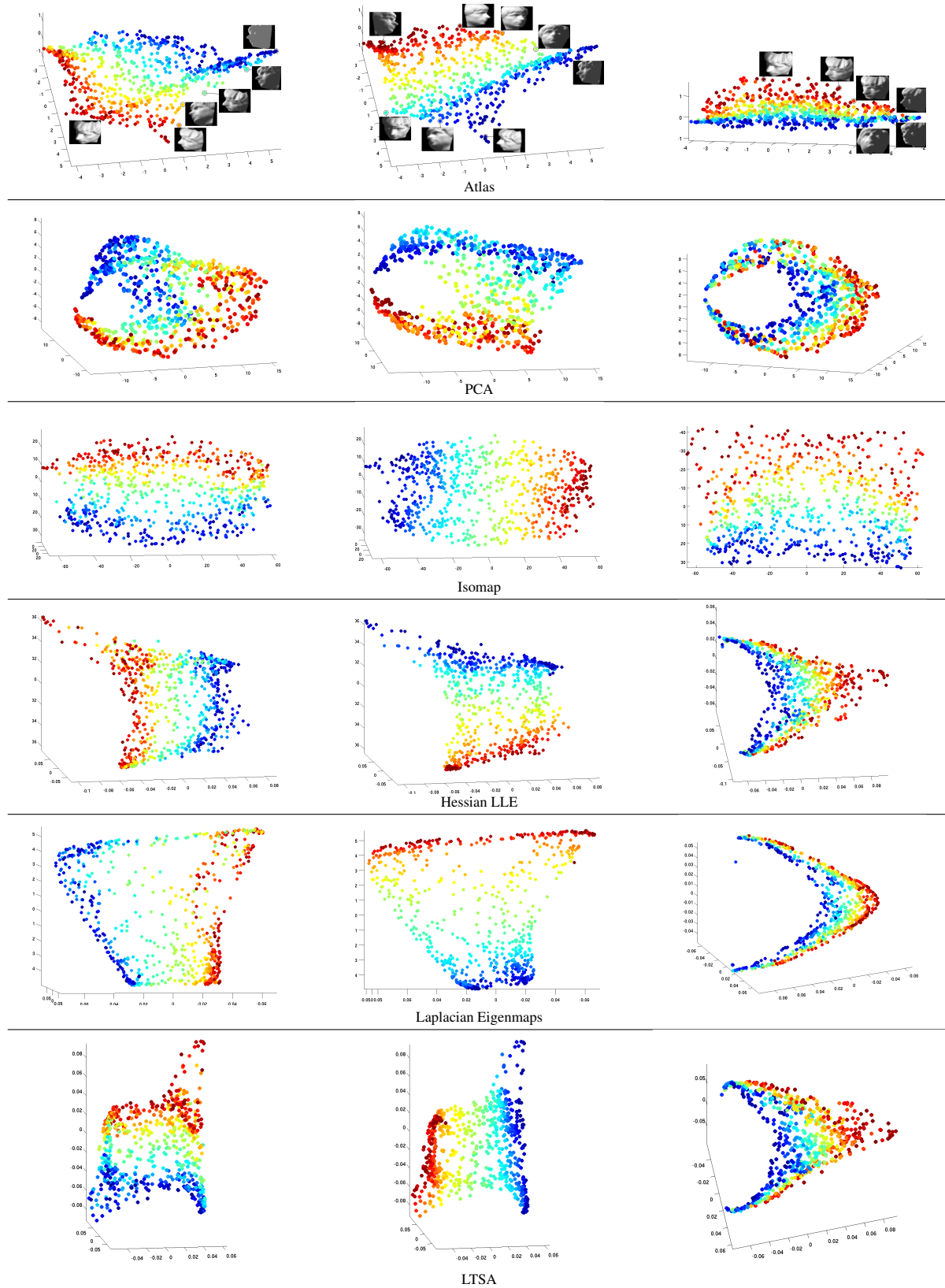


Figure 5. Best viewed in colour. Intrinsic modes of variation of the ISOMAP faces manifold. **Left column** Lighting direction left (red) to right (blue). **Middle column** Horizontal orientation left (blue) to right (red) pose. **Right column** Vertical orientation bottom (blue) to top (red) pose.