

Non-Linear Dimensionality Reduction With Isomaps

Galina Paskaleva*

Supervised by: Stefan Jeschke

Institute of Computer Graphics
Vienna University of Technology
Vienna / Austria

Abstract

This paper discusses the Isomap method for dimensionality reduction and studies its performance on both artificial and natural datasets. While linear methods for dimensionality reduction such as Principle Component Analysis (PCA) detect a linear subspace of the original domain that represents the data with maximal accuracy, the Isomap method detects the tangent space of a manifold embedded in the original domain. PCA remains globally linear as it simply transforms the data from one vector space into another. Isomap is only locally linear; globally it maps a low-dimensional manifold embedded in a high-dimensional domain into a lower-dimensional vector space by globally aligning the local manifold tangent spaces. The critical precondition for its success is that the sampling frequency of the data is sufficient to avoid 'short-circuits' in the spatial dissimilarity matrix. This is demonstrated by means of artificial datasets with dimensionality ranging from two (e.g. planar geometry) to several thousand (e.g. images of geometrical models) since they allow absolute control over the sampling frequency. The results of Isomap applied to a natural dataset (pressure images from a foot scan) are much more ambiguous. Since we have no prior knowledge of the 'hidden' dimensionality of the data we do not know if the sampling frequency is sufficient. The difficulty here lies in differentiating between classification outliers, i.e. atypical samples, and 'unexpected' degrees of freedom, that may initially seem 'wrong' if they contradict our expectations.

Keywords: Isomap, Principle Component Analysis, dimensionality reduction, non-linearity, sampling frequency

1 Introduction

Dimensionality reduction can be viewed as the process of finding the domain best suited for embedding a set of high-dimensional data samples [2]. Ideally, the dimensionality of this domain is considerably lower (e.g. an order of magnitude or more) than the dimensionality of the data samples, yet sufficient to reproduce them within a given margin of error. A highly useful byproduct of dimensionality

reduction is an interpretation of the given data set within the new domain [2] that is not possible in the original domain due to its high complexity. The purpose of this work is to discuss the Isomap method for dimensionality reduction and to compare it to the classical method of Principal Component Analysis (PCA) in terms of effectiveness with particular emphasis on the quality of insight provided into both geometrical and image data.

Human visual perception is capable of reducing complex data to its underlying dimensionality in less than a second (i. e. the lighting in a photograph), particularly in the presence of prior knowledge or expectations [9, 3]. While during the first phases of object recognition linear methods such as frequency decomposition play an important role [9], in the latter stages of conscious search for yet undiscovered interdependencies they might prove too restrictive. In this work the ability of the most commonly used linear method (PCA) and one non-linear method (Isomaps) to help a human observer gain new insight into complex high-dimensional (visual) data is examined in detail.

The next section provides a brief overview over the existing methods for dimensionality reduction. Section 3 reviews Isomaps and PCA in detail. Section 4 compares the results of the application of both Isomaps and PCA to artificially produced data sets. Section 5 performs the same comparison on the basis of a natural data set. Section 6 discusses the results.

2 Methods for Dimensionality Reduction

All methods for dimensionality reduction can be divided into two large classes - the classical linear techniques and the more recently developed non-linear techniques.

As the term *linear* will be used extensively we provide a definition. A linear map ($A : X \rightarrow Y$ with X and Y vector spaces) or a linear operator ($A : X \rightarrow X$ on the vector space X) is a function with the following two properties:

$$\begin{aligned} \text{additivity :} & & A(\vec{x} + \vec{y}) &= A(\vec{x}) + A(\vec{y}) \\ \text{homogeneity of degree 1:} & & A(\vec{x}\alpha) &= (A\vec{x})\alpha \text{ with } \alpha \in \mathbb{R} \end{aligned}$$

Whenever a reduction step is described as linear it is to

*g.paskaleva@gmx.at

be understood that it defines a linear map from the vector space of the original data points X into a new vector space Y .

2.1 Linear Methods for Dimensionality Reduction

The classical methods for dimensionality reduction include PCA, classical Multidimensional Scaling (MDS), Factor Analysis (FA), and Independent Component Analysis (ICA). What is common to all of them is that they look for a linear subspace in the sampled data [2, 5]. A common advantage of all linear methods is that the embedding of new data samples in the already calculated subspace is reduced to applying the linear mapping to the new samples. For most non-linear methods, however, only estimation techniques exist [5].

2.2 Non-Linear Methods for Dimensionality Reduction

Methods that can deal with non-linear cases can be subdivided in techniques that preserve global data properties in the low-dimensional representation - e.g. Kernel PCA, Semidefinite Embedding (SDE), MDS with a non-Euclidean distance function, Isomaps, and Maximum Variance Unfolding (MVU), techniques that preserve local properties in the low-dimensional representation - e.g. Locally Linear Embedding (LLE), Laplacian Eigenmaps (LEM), Hessian LLE, and techniques that calculate local linear models and subsequently align them globally. Markov processes and neural networks have also been successfully used for non-linear dimensionality reduction. Detailed discussion of the above methods can be found in [2, 5, 8].

3 Mathematical Background of PCA and Isomaps

In this section the mathematical theory behind PCA and Isomaps is discussed in detail. The fundamental difference between the two methods is presented in Figure 1.

3.1 PCA

Let us assume that there are N data samples of dimensionality D and let the i -th data point be represented as a D -dimensional vector $\vec{x}_i = (x_{i1}, x_{i2}, \dots, x_{iD})$ with $i = 1, \dots, N$. The result of PCA is a linear map that projects the data points from the original vector space into another while retaining as much of the variability of the data as possible and orienting the coordinate axes (or principal components) spanning the new vector space according to it. The principal components are linear combinations of the existing data points. As they are orthogonal to each other there can be at most D of them.

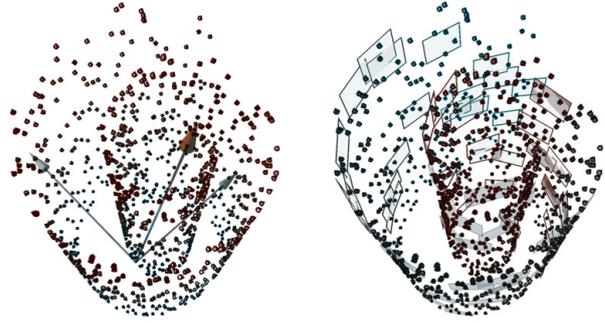


Figure 1: On the left PCA determines an optimal embedding of the 'Swiss roll' dataset in a 3-dimensional vector space. On the right Isomap performs local linear fitting before mapping the detected local vector spaces into a global vector space.

Let us assume that the axis aligned to the maximum variability of the data is $P_1 = \sum_{i=1}^n (w_i \vec{x}_i)$.

If we combine the coefficients w_i in a vector $\vec{w} = (w_1, w_2, \dots, w_N)$ and the data points \vec{x}_i - column-wise in a $N \times D$ matrix X we can write the previous equation as $P_1 = \vec{w}^T \times X$.

The variance along P_1 can be calculated as $VarP_1 = E[(P_1 - E[P_1])^2]$ where $E[P_1]$ is the expected value (mean) of P_1 , which exists whenever P_1 has a variance [7]. We can assume $E[P_1] = 0$ since this would only remove one translation from the linear map that can be re-applied in the very last step. Thus

$$\begin{aligned} VarP_1 &= E[P_1^2] \text{ or} \\ &= (\vec{w}^T \times X) \times (\vec{w}^T \times X)^T \\ &= \vec{w}^T \times (X \times X^T) \times \vec{w} \\ &= \vec{w}^T \times S \times \vec{w} \end{aligned}$$

with S the (scaled) covariance matrix of the original data points [7].

As $VarP_1$ is not bounded above for an arbitrary \vec{w} we impose the restriction $\|\vec{w}\| = 1$. We want to derive P_1 from the covariance of the data samples, not scale it to infinity. In order to maximize $\vec{w}^T \times S \times \vec{w}$ with the constraint $\vec{w}^T \times \vec{w} = 1$ we apply a Lagrange multiplier α_1 :

$$L(\vec{w}, \alpha_1) = \vec{w}^T \times S \times \vec{w} - \alpha_1 \times (\vec{w}^T \times \vec{w} - 1).$$

After differentiating with respect to \vec{w} we have $S \times \vec{w} = \alpha_1 \times \vec{w}$. Obviously, $VarP_1$ is maximal for the biggest α_1 , i.e. the largest eigenvalue of S corresponding to the eigenvector \vec{w} .

Essentially, in its first step PCA calculates the $D \times D$ covariance matrix of the original data samples. Then the R eigenvectors ($R \leq D$) corresponding to the R largest non-zero eigenvalues of the covariance matrix build the base B_R of the R -dimensional linear domain in which the data samples can be projected by means of the map $B_R B_R^T$ with minimal loss of variability [2]. A significant difference between the R th and the $(R + 1)$ th largest eigenvalue indicates the existence of an R -dimensional linear subspace 'hidden' in the original domain. In that case the remaining

$D - R$ Eigenvectors can be interpreted as noise obscuring the true dimensionality of the data.

An obvious drawback of PCA is that the size of the covariance matrix depends on the dimensionality of the data samples and becomes impractical for high-dimensional domains such as image data. One way to avoid this, especially if $N < D$, is to apply dual PCA [5].

The projection calculated by PCA has the effect of minimizing the squared reconstruction error. Using this property as a starting point, Ghodsi shows in [2] that the linear map of PCA can be obtained as the matrix U in the singular value decomposition of $X = U \times E \times V^T$ as it contains the eigenvectors of XX^T (V contains the eigenvectors of $X^T X$ and E - the square roots of the non-zero eigenvalues of both XX^T and $X^T X$ in its diagonal). Once U , V and E can be reduced in size according to the rank of E it follows that E is invertible and $U = XVE^{-1}$. This produces the linear map $U^T U$ of dual PCA (by solving the dual problem to that of classical PCA).

3.2 Isomaps

In its first step Isomap uses a neighborhood criterion (k nearest neighbors or an ϵ -neighborhood [4]) to build a neighborhood graph of the data samples. Then it builds a $N \times N$ spatial dissimilarity matrix by calculating the shortest paths between each pair of data samples, which allows non-linear subspaces to come into consideration [6]. The goal is to use the shortest paths between data samples to detect a manifold embedded in the vector space of the original data points.

In its second step Isomap applies classical MDS. This method takes the $N \times N$ dissimilarity matrix of N D -dimensional data points produced in the previous step as its input and calculates a linear transformation that preserves the pair-wise shortest path distances between the points while disregarding their coordinates in the original domain entirely. In effect, instead of minimizing the reconstruction error relative to the the original coordinates (which is what PCA does) MDS minimizes the reconstruction error relative to the dissimilarity matrix. Since it contains only scalar values it does not depend on the original dimensionality of the data points; in fact, it is the function of MDS to find the minimal suitable dimensionality for data points with (nearly) the same dissimilarity matrix (see Figure 2).

MDS uses the fact that the dissimilarity matrix, when employing the Euclidean distance as a dissimilarity measure, is expressed by means of the inner product of the $N \times D$ data point matrix X (the Gram matrix [8]): $G = X^T X = -1/2 \times H \times D^2 \times H$ with $H = I - 1/D \times \vec{e} \times \vec{e}^T$ where $\vec{e} = (1, 1, \dots, 1)$ is a N - dimensional vector [5]. There exist multiple methods for recovering a data matrix X when only the Gram matrix G is known [5, 8]. The result corresponds to a 'flattening' or linearization of the tangent space of any manifold concealed in the original domain.

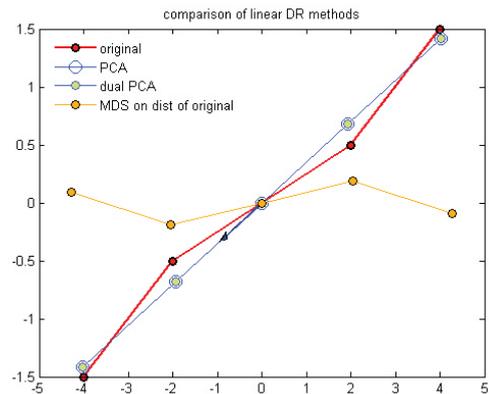


Figure 2: A comparison of PCA, dual PCA and MDS. Both PCA methods perform reconstruction with one component. MDS detects two degrees of freedom.

Tennenbaum *et. al.* provide a proof in [1] that as the number of data samples increases so does also the accuracy of the approximation of the underlying manifold performed by Isomap. Asymptotic convergence to the 'true' underlying structure is guaranteed for smooth manifolds isometric to a convex domain of Euclidean space [4]. These dependencies will be demonstrated in the next section.

4 PCA and Isomap Applied to Artificial Datasets

The experiments on artificial datasets documented in [5] show that Isomap is significantly better than PCA in detecting non-linear structures on the 'Swiss roll', the 'Helix' (i.e. on smooth highly non-linear manifolds, the first one Euclidean, the second one - non-Euclidean) as well as on a discontinuous 'broken Swiss roll' dataset. However, Isomap performs only slightly better than PCA on a dataset with high intrinsic dimensionality.

4.1 Low-dimensional Data Sets

We start our experiments with a one-dimensional dataset folded in two dimensions - a spiral. Applying Isomap with 4 nearest neighbors produces the mapping (represented by the straight lines) shown in Figure 3.

The next experiment applies Isomap with 4 nearest neighbors again to a one-dimensional set, but this time it

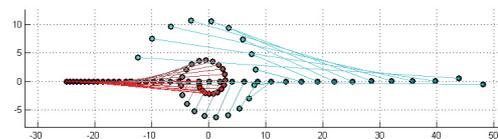


Figure 3: Isomap performed on a one-dimensional manifold folded in two dimensions.

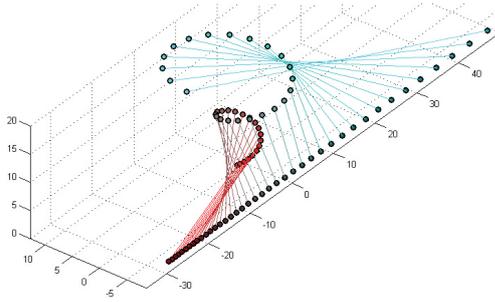


Figure 4: Isomap performed on a one-dimensional manifold folded in three dimensions.

is folded in three dimensions. As Figure 4 shows the result is the same as in the previous case.

Figure 5 shows a region in the data samples that causes the error in the mapping of the manifold into the one-dimensional domain (the loop in Figure 6 - the color gradient corresponds to the ordering of the data samples according to the first detected degree of freedom). Since the underlying manifold is actually one-dimensional, the loop represents the failure of Isomap to recognize this in the vicinity of the 'short-circuit'.

These three examples demonstrate a key requirement for the convergence of Isomaps [1]: *the sampling condition*. It states that arbitrarily close approximation of the underlying manifold M is possible if the neighborhood graph contains all edges connecting data samples at a distance less than a positive parameter ϵ , and for each point m on M there exists a data sample that is at a distance less than $\epsilon/4$ from m . In other words, the sampling density must be sufficient for every point on the manifold. In particular, it must be at most half of the shortest distance between manifold 'folds' (see Figure 5).

The next two experiments utilize the 'Swiss roll' dataset

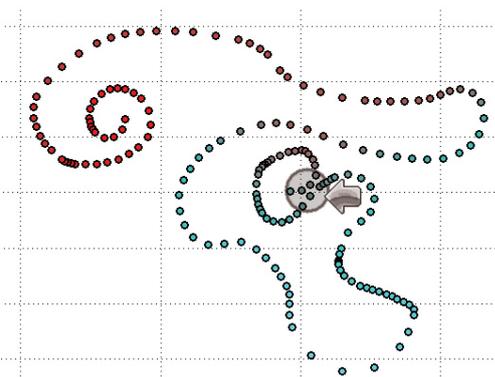


Figure 5: An example of insufficient sampling density.

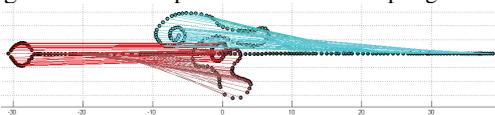


Figure 6: The loop on the far left showcases the error resulting from insufficient sampling density.

used in [4] and [5]. Figures 7 and 9 visualize the difference resulting from the choice of the number of nearest neighbors k . For $k = 7$ Isomap manages to recognize the underlying two-dimensionality of the dataset in spite of the numerous 'short-circuits' resulting from insufficient data sampling (see Figure 7 - the change in color corresponds to the ordering along one detected degree of freedom, the edges are those of the neighborhood graph). The 'unwrapped' manifold is shown in Figure 8. For $k = 20$, however, we obtain the same result as with PCA - there is no meaningful dimensionality reduction, only a change in the coordinate system. Since for all examples above PCA cannot reduce the dimensionality of the data (as there is no significant difference in the detected eigenvalues) its results are not shown.

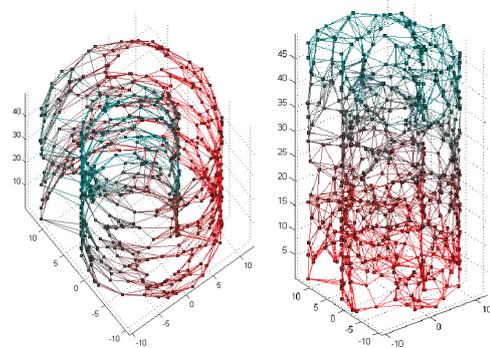


Figure 7: The 'Swiss roll' dataset colored according to the first and second degrees of freedom detected by Isomaps with $k = 7$.

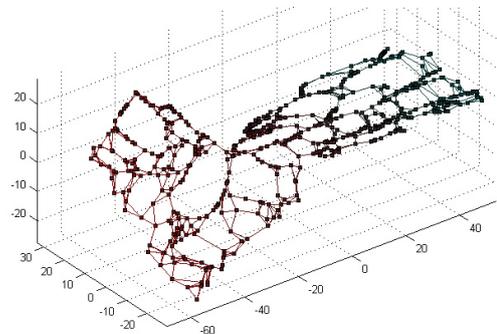


Figure 8: The unwrapped 'Swiss roll' dataset from Figure 7 shows a slight 'thickening' in the third dimension due to 'short-circuiting' errors.

The 'Swiss roll' examples showcase another important condition for successful approximation: *the neighborhood condition*. It requires that the neighborhood graph does not contain edges connecting data samples whose Euclidean distance is larger than certain parameters derived from the minimum radius of curvature r_0 and the minimum branch separation s_0 of M . The minimum branch separation is the largest Euclidean distance between points on M that still guarantees a geodesic distance of less than πr_0 [1]. In the case of 20 nearest neighbors this condition is clearly violated and thus Isomap fails.

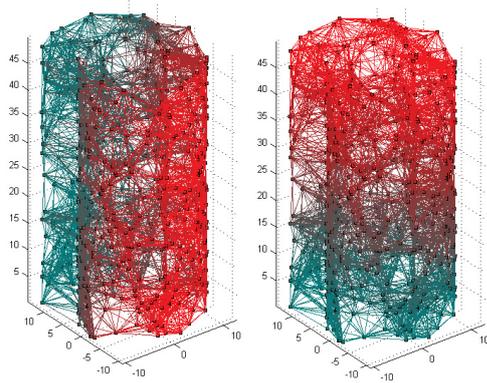


Figure 9: The 'Swiss roll' dataset colored according to the first and second degrees of freedom detected by Isomaps with $k = 20$.

In [1] Bernstein *et al.* provide a proof that a random sampling of sufficient density can satisfy the two conditions above with probability depending on the volume of M and on the volume of the smallest metric ball in M . However, while these conditions are easy to check in an artificially generated data set with known degrees of freedom it is quite difficult to do so for natural datasets where the prior knowledge is insufficient, as will be demonstrated in Section 5.

4.2 High-dimensional Data Sets

We now move to a very high-dimensional domain - that of image data. The generated images satisfy both the sampling and the neighborhood conditions from Section 4.1.

We generate 300 images with resolution 64×86 pixels of a glass ball illuminated by several light sources (as shown in Figure 11), only one of which changes position - it traverses one half of a circular trajectory from left to right. The resulting data samples are 5504-dimensional vectors ($64 \times 86 = 5504$) containing 8-bit integral intensity values.

Figure 10 shows the residual variances after consecutive data reconstruction with 1 to 6 components detected by PCA and by Isomap ($k = 2$ nearest neighbors considered in the neighborhood graph) respectively. Where PCA performs a frequency decomposition of the data (first image group) and the residuals diminish rather gradually, Isomap (second image group) shows a sharper drop in the residuals after the first detected degree of freedom and correctly identifies the main degree of freedom in the data as movement.

Figure 11 shows in the first row the ordering of the images along the first principal component calculated by PCA. From the viewpoint of a human observer this ordering does not carry any meaningful insight. The ordering in the second row on the other hand corresponds to a step-by-step rotation of the mobile light source around the ball.

The next experiment introduces two degrees of freedom. We generate 5000 images with resolution 100×50 pixels of a statue (as shown in Figure 12). It is illuminated

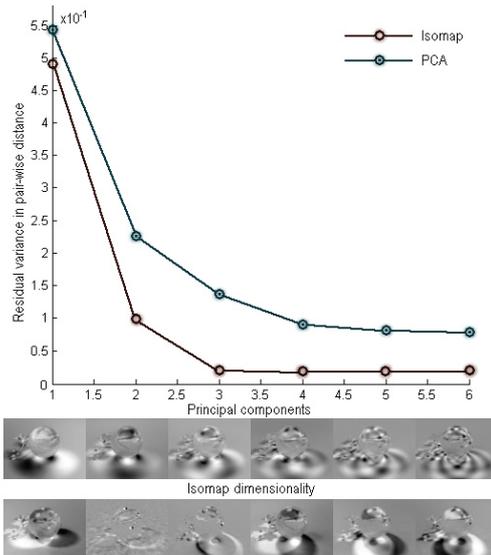


Figure 10: The residual (unaccounted for by the currently used model) variance along the first 6 degrees of freedom detected by PCA and Isomap respectively. The sharp drop in the graph indicates that PCA as well as Isomap has detected the intrinsic dimensionality of the data.

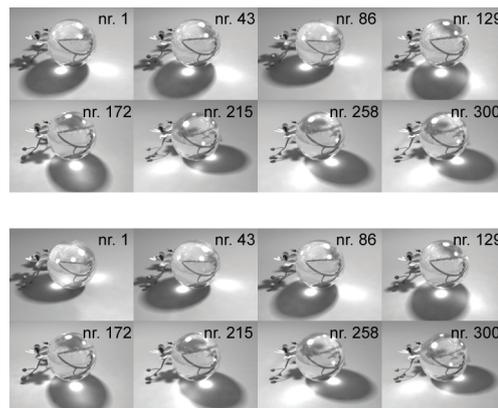


Figure 11: The first group shows the ordering according to the first principal component. The second group shows the ordering according to the first degree of freedom detected by Isomap.

by two light sources, one of which traverses a similar trajectory as in the previous experiment. What also changes in these images is the camera position, which traverses a one-dimensional path folded in three dimensions around the statue.

Similar to the previous experiment, Figure 12 shows the first 6 principal components detected by PCA and the resulting ordering of the data; Figure 13 shows the variations along the degrees of freedom with the 6 highest residuals resulting from Isomap with $k = 10$ nearest neighbors considered in the neighborhood graph, again followed by the resulting ordering of the data. The difference is less obvious this time. For PCA the following observation can be made: the smaller the corresponding eigenvalue, the higher the frequency of the information carried by the principal component. Again there is a significant drop after the

largest residual to indicate a sub-space of the original domain well suited for representing the data, in this case - the overall lightness of the scene. The behavior of Isomap is also similar to that in the previous case with the difference that the sharp drop in the residuals occurs after the second detected degree of freedom and the result gives a human observer information about the parameters responsible for the variation in the image data.

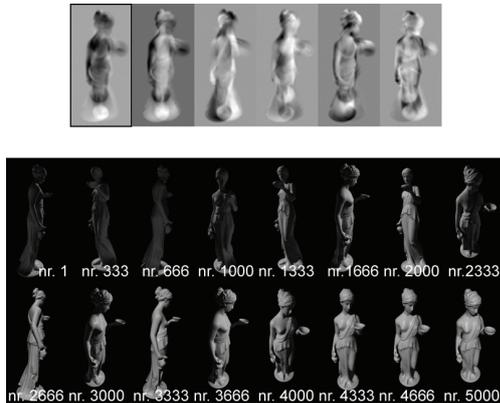


Figure 12: The ordering of the statue images according to the first principal component (highlighted in the top row) detected by PCA.

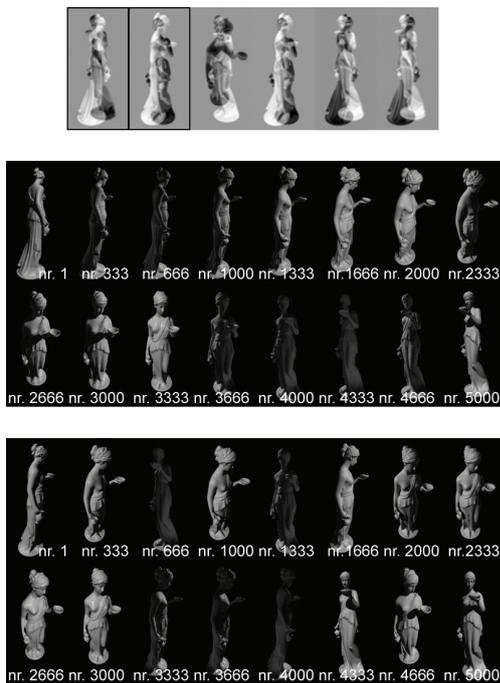


Figure 13: The ordering of the statue images according to the first and second degrees of freedom detected by Isomap: the camera path and the light source path. The variation (gradient) along the first 6 degrees of freedom is in the top row.

The orderings in Figure 13 show a step-by-step traversal of the camera path, and a step-by-step change of the position of the mobile light source. The second ordering is remarkable in its independence from the camera position

- the ordering is correct regardless of the overall lightness of the scene.

5 PCA and Isomap Applied to a Natural Dataset

The experiments on natural datasets documented in [5] show that Isomap performs poorer than PCA in four out of five cases. The first reason for that according to Maaten *et. al.* is the 'curse of dimensionality' - the fact that with the rising of the dimensionality of the underlying manifold the number of data samples needed for its proper detection rises exponentially. The second reason is the presence of noise in the natural datasets that results in local overfitting. On the other hand, outliers result in very similar performances. While even a single outlier can cause a 'short-circuit' in the neighborhood graph of Isomaps, the more 'short-circuits' Isomap produces the more the spatial dissimilarity matrix approaches the squared Euclidean distances matrix for the data samples and consequently PCA (see Section 3). Consequently, in the presence of enough outliers Isomap fails to detect non-linear subspaces to the same degree PCA does.

The natural dataset used in this last experiment consists of 1298 archive images with resolution of 410×230 pixels of the maximal walking pressure distribution over a human foot of volunteers aged 12 to 85, courtesy of Sandrina Illes, MSc, TU Chemnitz, Department Forschungsmethoden und Analyseverfahren. The images were obtained by means of a RSScan 0.5 Gait Scientific foot scanner with spatial resolution of 4 sensors per cm^2 and temporal resolution of 300 Hz.

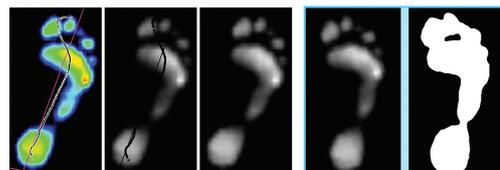


Figure 14: Dataset preparation. The second image from the right represents dataset GS, the rightmost image - the dataset BW.

In the initial stage we prepared the images as illustrated in Figure 14. First the color coded pressure distribution was converted into intensity values. Subsequently all discontinuities were eliminated through linear interpolation in order to improve the conditions for the application of Isomaps (see Section 3). This was followed by the global alignment of all images. The thus prepared grayscale images built dataset GS. The binary dataset BW was derived from GS by employing thresholding followed by a low pass filter to remove noise. In order to reduce dimensionality we scaled the images to a resolution of 64×36 pixels (2304 dimensions). Tests with different did not produce significant changes in the final result.

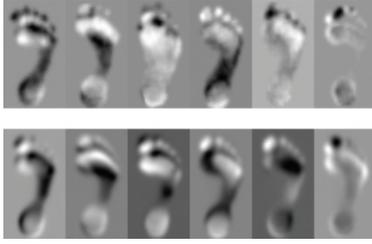


Figure 15: The first row shows the variance along the first 6 degrees of freedom detected by Isomap. The second row shows the first 6 principal components calculated by PCA.

Figure 15 shows the comparison between the first 6 principal components detected by PCA and the variations along the degrees of freedom with the 6 highest residuals resulting from Isomap with $k = 20$ nearest neighbors applied to dataset GS. We performed the Isomap algorithm with values for k ranging from 2 to 50 and chose 20 as it produced the least number of apparent ordering errors. PCA failed to detect any sharp drop in the Eigenvalues corresponding to the first 20 principal components. Isomap exhibited the same behavior in respect to the residuals corresponding to the first 20 detected degrees of freedom.

Figure 16 shows the ordering along the first, second, and third degrees of freedom detected by Isomap respectively. As opposed to the results in the previous section, the interpretation here is difficult due to lack of prior knowledge of the intrinsic dimensionality of the original domain. A tentative interpretation of the results was attempted by an expert in the field of gait analysis, Sandrina Illes, MSc.

The first degree of freedom seems to indicate the amount of pressure on the joint of the big toe. The second degree of freedom seems to correspond to the pressure distribution on the front of the foot, the foot arches, and the heel. The third degree of freedom seems to relate to the pressure on the base of the second and fifth metatarsals. The fourth degree of freedom seems very similar to the first one. The fifth degree of freedom seems to indicate the amount of pressure on the toes compared to the pressure on the front of the foot, and the sixth degree of freedom seem to give an ordering according to the presence of Hallux Valgus (bunion).

Figure 17 shows the comparison between the first 6 principal components detected by PCA and the variations along the degrees of freedom with the 6 highest residuals resulting from Isomap with $k = 20$ nearest neighbors applied to dataset BW. We performed the Isomap algorithm with values for k ranging from 2 to 50 and again chose 20 as it delivered the most consistent results. PCA failed to detect any sharp drop in the Eigenvalues corresponding to the first 20 principal components. Isomap exhibited the same behavior with respect to the residuals corresponding to the first 20 detected degrees of freedom.

Figure 18 shows the ordering along the first, second, and third degrees of freedom detected by Isomap respectively. A tentative interpretation of the results was again attempted by Sandrina Illes, MSc.

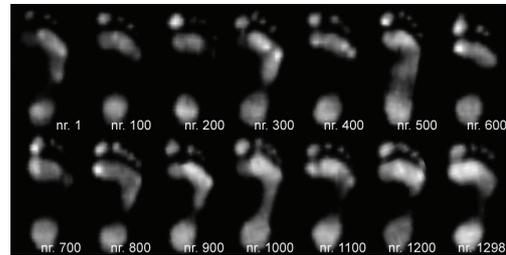
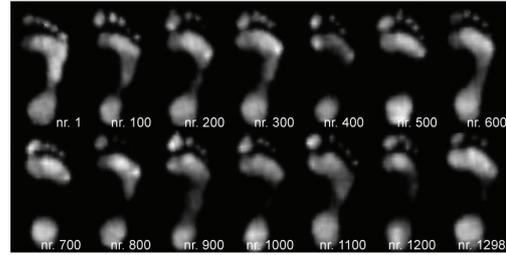
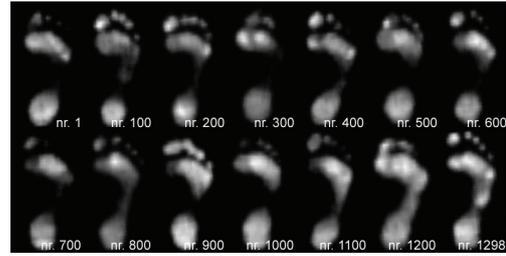


Figure 16: The ordering of the GS foot images according to the first three degrees of freedom detected by Isomap.

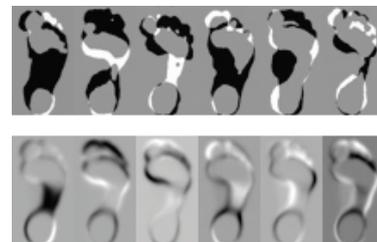


Figure 17: The first row shows the variance along the first 6 degrees of freedom detected by Isomap. The second row shows the first 6 principal components calculated by PCA.

The first degree of freedom seems to produce an ordering from very high to completely collapsed foot arches. The second degree of freedom seems to correspond to the overall shape of the foot with an emphasis on the relation between the length of the longitudinal arch and the foot length. The third degree of freedom seems to relate to the form of the front of the foot and the degree to which the small toes are used. The fourth degree of freedom seems somewhat similar to the first one. The fifth degree of freedom seems to indicate the amount to which the transversal arch touches down, and the sixth degree of freedom seems to be in relation to the amount of pressure on the joint of the big toe as well as the function of the other four toes.

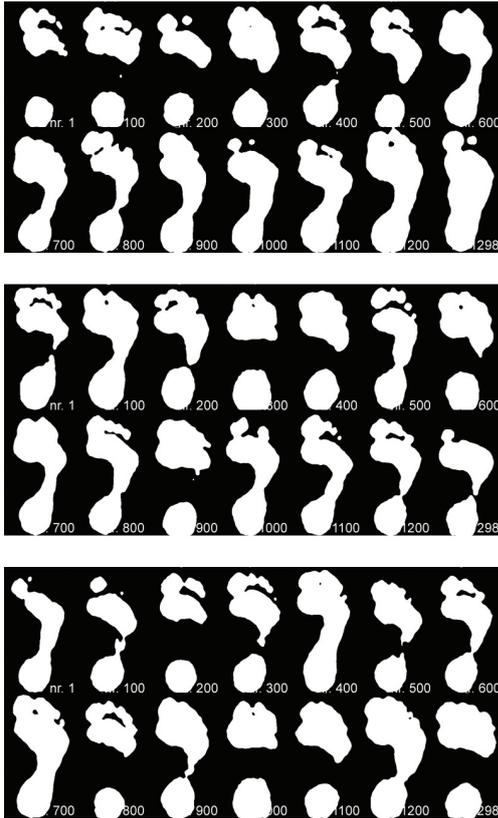


Figure 18: The ordering of the BW foot images according to the first three degrees of freedom detected by Isomap.

6 Results

The results of our experiments in Sections 4 and 5 confirm the performance predictions in [1] and the findings in [5]. Isomap delivers the best results in recognising the intrinsic dimensionality of artificial datasets since they can be sampled adequately due to our a priori knowledge of their structure. The dimensionality of the original data samples does not influence the results. The failure to satisfy the sampling condition (see Section 4.1) results in 'thickened' manifolds [5] due to noise from 'short-circuiting'. Nevertheless Isomap still delivers good results. The failure to satisfy the neighborhood condition however causes convergence of the results from PCA and Isomap - i.e. inability to detect non-linear subspaces in the sampled data.

The main difficulty with natural datasets lies in our lack of prior knowledge. Since we cannot determine if the data samples satisfy the two conditions in Section 4.1 it is highly likely that one or both of the problems described in the paragraph above will occur, which in turn makes interpretation of the result even more challenging.

For the low-dimensional examples in Section 4.1 both PCA and Isomap required less than 1 min. to calculate in *MatLabTM* on a dual core 1.8 GHz, 2 GB RAM 32-bit Windows platform. For the image examples in Section 4.2 Isomap needed 4 min. for the ball, 10 min. for the foot BW, 12 min. for the foot GS and 20 min. for the statue.

Classical PCA needed 6 min. for the ball, 5 min. for the foot BW, 7 min. for the foot GS and 20 min. for the statue datasets. Dual PCA needed less than 2 min. per example.

7 Conclusions

Isomaps is a non-linear method for dimensionality reduction that preserves the geodesic distances between the data samples in their lower-dimensional representation, if such representation can be found. When applied to sufficiently densely sampled artificial datasets it outperforms PCA significantly both in detecting underlying lower-dimensional highly folded subspaces as well as in achieving readability for human observers. On natural datasets PCA and Isomaps deliver comparable results due to our inability to ensure a satisfactory sampling density (at least half of the smallest distance between manifold 'folds') and an adequate neighborhood definition. In order for a human observer to glean new insight in a heretofore unknown domain by means of Isomap one needs to proceed iteratively, adjusting the number of nearest neighbors to be considered in the neighborhood graph and, even more importantly, the data sampling density in each iteration.

References

- [1] Mira Bernstein, Vin de Silva, John C. Langford, and Joshua Tennenbaum. *Graph Approximation to Geodesics On Embedded Manifolds*. Stanford University, Palo Alto, CA, USA, December 2000.
- [2] Ali Ghodsi. *Dimensionality Reduction a Short Tutorial*. Department of Statistics and Actuarial Science University of Waterloo, Waterloo, Ontario, Canada, 2006.
- [3] Christopher Summerfield and Tobias Egner. *Expectation (and attention) in visual cognition*. Trends in Cognitive Sciences, vol. 13, no. 9, pp. 403-409, September 2009.
- [4] Joshua Tennenbaum, Vin de Silva, and John C. Langford. *A Global Geometric Framework for Nonlinear Dimensionality Reduction*. Science, vol. 290, pp. 2319-2323, December 2000.
- [5] L.J.P. van der Maaten, E.O. Postma, and H.J. van den Herik. *Dimensionality Reduction: A Comparative Review*. MICC, Maastricht University, Maastricht, The Netherlands, January 2008.
- [6] Dan Ventura. *Manifold Learning Examples - PCA, LLE and ISOMAP*. Department of Computer Science, Brigham Young University, Provo, UT, USA, October 2008.
- [7] R. Viertl. *Einfuehrung in die Stochastik*. Springer-Verlag, Wien, Germany, 2003.
- [8] Tynia Yang, Jinze Liu, Leonard McMillan, and Wie Wang. *A Fast Approximation to Multidimensional Scaling*. IEEE workshop on Computation, 2006.
- [9] Xu Yaoda and Marvin M. Chun. *Selecting and perceiving multiple visual objects*. Trends in Cognitive Sciences, vol. 13, no. 4, pp. 167-174, April 2009.