Iterative Hyperplane Merging: A Framework for Manifold Learning

Harry Strange
hgs08@aber.ac.uk
Reyer Zwiggelaar
rrz@aber.ac.uk

Department of Computer Science, Aberystwyth University, Aberystwyth, SY23 3DB, Wales, UK

Manifold learning algorithms have received much focus in the computer vision and pattern recognition communities over the past decade. Many problems in these fields require a low dimensional representation to be found as working in higher dimensions can often be problematic. Dimensionality reduction and manifold learning techniques have been used to reveal patterns in such high-dimensional data [6]. As an example, consider a data set consisting of a sequence of images showing a rotating 3-dimensional object. Across the dataset the object rotates around one of its axes. If each of these images were to be thought of as a point in high-dimensional space (the dimensionality being equal to the number of pixels in the image) then they would lie on a simple circular manifold that is parameterized by the degree of rotation of the object. This means that each image can be discriminated using only 1-dimension - the degree of rotation - as opposed to, in the case of a 128×128px image, $16,384$ dimensions. This reduction of dimensionality overcomes many computational and mathematical problems associated with high-dimensional learning.

Many techniques have been proposed to perform manifold learning, ranging from simple linear transforms, for example Principal Components Analysis (PCA) [3], to more advanced non-linear learning algorithms like ISOMAP [5] and Local Tangent Space Alignment (LTSA) [7]. We present a manfiold learning algorithm, Iterative Hyperplane Merging, that can be used to find non-linear manifolds in high-dimensional data. Iterative Hyperplane Merging (IHM) can be intuitively thought of as a form of local PCA where PCA is applied at a local scale to produce low-dimensional local hyperplanes. These local hyperplanes are then globally aligned to produce the final low-dimensional embedding. A clustering algorithm is employed to create the partitions needed to form the local hyperplanes, in this paper we use either a Gaussian Mixture Modelling scheme [2] or a constrained k-means clustering algorithm [1] to form the partitions. A Minimum Spanning Tree (MST) of the inter-hyperplane distance graph is used as the basis for forming the global alignment of the hyperplanes as it provides a skeleton of the manifold along which we can walk. To obtain a faithful global embedding we perform a pre-order traversal on this Minimum Spanning Tree. When walking from one node in the MST to another we merge the hyperplanes gradually building a global embedding of the data. The merging function is defined as

$$f(\Pi_a \rightarrow \Pi_b) = \mathbf{A}\mathbf{U}_b\mathbf{U}_b^T + (\bar{\Pi}_b - (\bar{\Pi}_b\mathbf{U}_b\mathbf{U}_b^T)) \qquad (1)$$

and is a simple projection from one hyperplane, Π_a, to another, Π_b. Here $\mathbf{A}$ is the matrix of samples in all the hyperplanes already visited and is increased and updated after each call to the above merging function. If we are re-visiting a hyperplane in the traversal then we need to adapt the above merging function to improve the robustness of the embedding. We do this by applying Procrustes analysis [4] to translate, scale and rotate the image of the hyperplane in the global embedding to its original representation. This backtracking function is similar to the above merging function but with an extra constraint:

$$f(\Pi_{a'} \rightarrow \Pi_a) = \mathbf{b}(\mathbf{A}\mathbf{U}_a\mathbf{U}_a^T + (\bar{\Pi}_a - (\bar{\Pi}_a\mathbf{U}_a\mathbf{U}_a^T)))\mathbf{T} + \mathbf{v} \qquad (2)$$

where $\mathbf{b}$ is the isomorphic scale value, $\mathbf{T}$ is the rotation matrix and $\mathbf{v}$ is the translation vector. Once the traversal along the MST has finished we can find the low-dimensional embedding by performing Principal Components Analysis on the obtained global embedding.

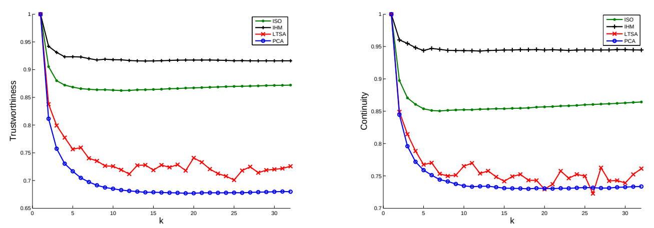

Figure 1: Graph of results for the trustworthiness (left) and continuity (right) of different algorithms when trying to unroll the Swiss Roll dataset with 2000 samples.. The neighborhoods for Isomap and LTSA were averaged over the range $k = [2,32]$. For IHM we use GMM to find optimal cluster sizes with an initial cluster size estimate of 32.

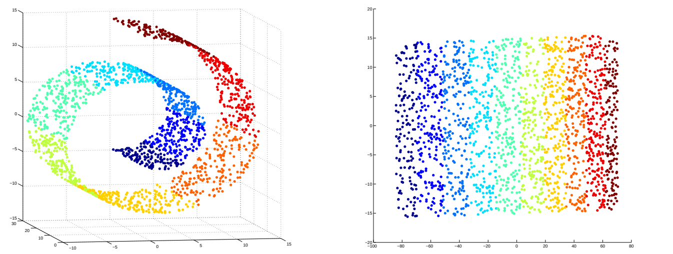

Figure 2: 3D Swiss Roll data set (left) and the IHM embedding using Gaussian Mixture Modelling for the clustering stage (right).

We have tested our algorithm on both toy data and real world image data. For the toy data we used the benchmark Swiss Roll data set and compared our results against three widely used algorithms - Principal Components Analysis [3], ISOMAP [5] and Local Tangent Space Alignment [7]. Our results show that over a range of local neighborhood sizes Iterative Hyperplane Merging is more stable than any of the other algorithms (Figure 1). It is also able to find the low-dimensional embedding without normalization or any global distortion (Figure 2). In this paper we also show how Iterative Hyperplane Merging can learn a high-dimensional image manifold by applying our algorithm to the Frey Faces dataset. Our results show that there is structure in the data with expressions distributed across the manifold and at a local scale images in the local input neighborhood match those in the local output neighborhood. When compared with other results using this dataset our embedding shows more meaningful structure. This suggests that Iterative Hyperplane Merging could be a useful tool for discovering structure and patterns in many real world image datasets.

[1] K. P. Bennett, P. S. Bradley, and A. Demiriz. Constrained k-means clustering. Technical Report MSR-TR-2000-65, Microsoft Research, May 2000.

[2] C. A. Bouman. Cluster: An unsupervised algorithm for modeling Gaussian mixtures. Available from http://www.ece.purdue.edu/~bouman, April 1997.

[3] H. Hotelling. Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24:417–441, 1933.

[4] D. G. Kendall. A survery of the statistical theory of shape. *Statistical Science*, 4(2):87–99, 1989.

[5] J. B. Tenenbaum, V. de Silva, and J. C. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290: 2319–2322, 2000.

[6] K. Q. Weinberger and L. K. Saul. Unsupervised learning of image manifolds by semidefinite programming. *International Journal of Computer Vision*, 70:77–90, 2006.

[7] Z. Zhang and H. Zha. Principal manifolds and nonlinear dimension reduction via local tangent space alignment. *SIAM Journal on Scientific Computing*, 26(1):313–338, 2004.