Localised manifold learning for cardiac image analysis

Kanwal K. Bhatia\textsuperscript{a}, Anthony N. Price\textsuperscript{b}, Jo V. Hajnal\textsuperscript{b} and Daniel Rueckert\textsuperscript{a}

\textsuperscript{a} Department of Computing, Imperial College London, UK;\textsuperscript{b} Imaging Sciences Department, MRC Clinical Science Centre, Imperial College London, UK

ABSTRACT

Manifold learning is increasingly being used to discover the underlying structure of medical image data. Traditional approaches operate on whole images with a single measure of similarity used to compare entire images. In this way, information on the locality of differences is lost and smaller trends may be masked by dominant global differences. In this paper, we propose the use of multiple local manifolds to analyse regions of images without any prior knowledge of which regions are important.

Localised manifolds are created by partitioning images into regular subsections with a manifold constructed for each patch. We propose a framework for incorporating information from the neighbours of each patch to calculate a coherent embedding. This generates a simultaneous dimensionality reduction of all patches and results in the creation of embeddings which are spatially-varying. Additionally, a hierarchical method is presented to enable a multi-scale embedding solution.

We use this to extract spatially-varying respiratory and cardiac motions from cardiac MRI. Although there is a complex interplay between these motions, we show how they can be separated on a regional basis. We demonstrate the utility of the localised joint embedding over a global embedding of whole images and over embedding individual patches independently.

Keywords: Manifold learning, multiple manifold alignment, Laplacian eigenmaps, cardiac image analysis

1. INTRODUCTION

Manifold learning has been shown in recent years to be a useful tool for dimensionality reduction in medical imaging, with applications including ultrasound gating\textsuperscript{11}, biomarker discovery\textsuperscript{15} and image registration\textsuperscript{5}. It requires the construction of a graph of similarities (or distances) between pairs of data points (images) in the high-dimensional space. Manifold learning algorithms aim to reduce the data dimension while preserving some function of local or global distances. Traditionally, it has often been used on whole images\textsuperscript{15}, but, by employing a single measure of similarity (or distance) between each image in this manner, any information on the locality of differences is lost and smaller trends may be masked by the dominant global differences. Other work has used specific, delineated structures such as the hippocampus\textsuperscript{16}, but it is not always possible to know in advance which regions of interest to consider. In this paper, we propose the use of multiple local manifolds to analyse local regions of images without any a-priori knowledge of which such regions are important.

Localised manifolds are created by partitioning images into regular subsections with a manifold constructed for each patch. Since it is probable that anatomical structures will span or move across regular boundaries, adjacent patches are unlikely to be independent of each other. We therefore develop a framework for incorporating the information from the neighbours of each patch in calculating a coherent embedding. This generates a simultaneous dimensionality reduction of all patches and results in the creation of embeddings which vary based on location in the image.

We apply this to the detection of regional variations in time-resolved cardiac Magnetic Resonance Imaging (MRI) sequences. Here, the presence of respiratory and cardiac motions complicate acquisition and analysis, and so it is desirable to be able to assign a position in each cycle to a given image frame. Recent related work has included using manifolds to learn the respiratory cycle for ultrasound gating\textsuperscript{11} and for reconstruction of lung CT volumes\textsuperscript{3}. Learning the manifold structure of cardiac sequences has also been used to aid segmentation of the
Respiration can be measured during acquisition using a navigator or respiratory bellows. Free-breathing navigator-gated strategies typically monitor displacement of the diaphragm, with a patient-specific factor used to correlate this motion with that of the heart. Where to best place such a navigator for different breathing patterns has been the subject of work in Savill et al. Respiratory bellows instead monitor the movement of the chest wall during breathing. Again, this has to be correlated with the motion of the heart to be useful. This motivates the investigation of how the physical movements at various locations within the abdominal cavity are related, and how this information may be learned directly from time-resolved cardiac MRI.

In addition to obtaining models of respiratory motion, we also extract information on the cardiac cycle. The cardiac phase is commonly determined using electrocardiography (ECG), which relies on an interpretation of the electrical activity in the heart as it beats. However, with certain pathologies, for example, cardiac dyssynchrony, the physical contractions of the heart may not be fully concordant with the ECG. In cardiac resynchronisation therapy, it has been shown that mechanical dyssynchrony is a better indicator of likely response than electrical dyssynchrony. Image-driven models of cardiac phase, based on the physical motion of the heart, are therefore desirable.

In this paper, we investigate the use of localised manifold learning for cardiac images acquired during free breathing, and show that although there is a complex, spatially-dependent interplay between cardiac and respiratory motion, these can be separated, and their contributions analysed, without any prior knowledge of the motions involved.

2. METHODS

Manifold learning is an approach for non-linear dimensionality reduction, which has gained increasing popularity in recent years. The main concept is that the dimensionality of many datasets is artificially high: the same data can instead be well represented by an underlying low-dimensional manifold embedded in the higher dimensional space. The task of manifold learning is to extract this intrinsic dimensionality while preserving the structure of the dataset. This makes it particularly appealing for use in medical imaging, to transform high-dimensional images to a more interpretable representation. For example, each 2D frame of size 100 × 100 pixels within a cardiac sequence can be viewed as a point in a 10000-dimensional vector space. However, frames from the same sequence exhibit a large degree of similarity indicating that not all of these dimensions are necessary for analysis. Reducing the dimensionality of such data is therefore desirable.

Various manifold learning algorithms have been developed which differ primarily in terms of how they preserve structure within the data. Local approaches such as Locally linear embedding (LLE) and Laplacian Eigenmaps (LE) aim to maintain local relationships between data points by preserving the distances between each point and its neighbours. Global methods such as Isomap and Semi-definite embedding (SDE) instead preserve both local and global relationships between all data points. In this paper we base our work on Laplacian Eigenmaps, which has shown good performance in learning manifolds of medical images. In the following section, we adapt LE to analyse a time series of 2D images of fixed spatial location on a regional basis.

2.1 Laplacian Eigenmaps

Laplacian Eigenmaps preserves structure in the data by ensuring that data points which are “close” in the high-dimensional space remain “close” in the low-dimensional embedding. This is done by minimising the following cost function:

\[ C(x) = \sum_{ij} (x_i - x_j)^2 W_{ij} \]

which minimises the weighted Euclidean distance between the embedding coordinates \( x_i \) and \( x_j \) of points \( i \) and \( j \), respectively, in the low-dimensional embedding. The measure of closeness between points \( i \) and \( j \) is defined by the weight, \( W_{ij} \), which indicates their similarity. For medical imaging applications, these points can represent whole
images or image patches. Some common similarity metrics include functions of the Euclidean norm distance (equivalent to sums-of-squared differences between the images) and the correlation coefficient\(^{11}\) as well as those based on Gabor filter responses.\(^{8}\) One advantage of LE over other manifold learning techniques is its capacity to additionally handle non-metric similarity measures such as Normalised Mutual Information.\(^{15}\)

### 2.2 Simultaneous consistent embedding of multiple patch manifolds

In order to learn about the local structure of images, we can subdivide images into smaller patches and construct a manifold for each of these. It is possible to embed each patch separately and compare the resulting embeddings after alignment with Procrustes analysis.\(^{12}\) However, while this is feasible for a very low number of dimensions, it becomes more difficult as the dimensionality increases. This reduces its utility in, for example, brain image analysis, for which up to 20 dimensions have been shown to be useful.\(^{15}\) More than this, we expect that anatomical structures would not be well delineated by regular patch boundaries, and structures and movements may span across more than one patch. This motivates using information from neighbouring patches in the embedding. To do this, we simultaneously embed adjacent patch manifolds\(^{4,10}\) by amending the cost function in (1) to keep adjacent patches close together. For a set of \(t\) time series images for each patch \(k, k = \{1...m\}\), the joint embedding at each time-point is given by \(x_k = \{x_{k1}, x_{k2}, ..., x_{kt}\}\), which is obtained by minimising:

\[
C(x) = \sum_{kl} \mu_{kl} \sum_i (x_{ki} - x_{li})^2 + \sum_k \sum_{ij} (x_{ki} - x_{kj})^2 W_{ij}^{X_k}
\]  

subject to \(\sum_k x_k^T D x_k = 1\).

Here \(\mu_{kl}\) is a weighting factor which is non-zero only if \(k\) and \(l\) are adjacent patches, and \(W_{ij}^{X_k}\) is the weight representing the similarity between time frames \(i\) and \(j\) over patch \(k\). \(D^{X_k}\) is a diagonal weight matrix of the patch \(k\), with non-zero entries given by \(D_k^{X_k} = \sum_j W_{ji}^{X_k}\), as in the standard LE method.

Extending the ideas of Ham et al.\(^4\) to multiple manifolds, the solution to (2) can be shown to be given by the solution to the generalised eigenvalue problem:

\[
L x = \lambda D x
\]

where

\[
L = \begin{pmatrix}
\sum_{n(X_1)} \mu_n + L^{X_1} & -2\mu_{12} I & \cdots & -2\mu_{1m} I \\
-2\mu_{12} I & \sum_{n(X_2)} \mu_n + L^{X_2} & \cdots & -2\mu_{2m} I \\
\vdots & \ddots & \ddots & \vdots \\
-2\mu_{1m} I & \cdots & -2\mu_{2m} I & \sum_{n(X_m)} \mu_n + L^{X_m}
\end{pmatrix}
\]

and

\[
D = \begin{pmatrix}
D^{X_1} & 0 & \cdots & 0 \\
0 & D^{X_2} & \cdots & 0 \\
\vdots & \ddots & \ddots & \vdots \\
0 & 0 & \cdots & D^{X_m}
\end{pmatrix}
\]

\(\sum_{n(X_k)} \mu_n\) is the sum of the weights between patch \(k\) and each of its \(n\) neighbours. \(L^{X_k} = D^{X_k} - W^{X_k}\) is the Laplacian matrix of an individual patch \(k\) and the eigenvectors \(x\) of \(L\) provide the final embedding map. Note that although this requires the solution of a larger (sparse) eigenvalue problem, the main computational cost of computing all pairwise distances between images is no different to that of a single embedding of the whole image.

### 2.3 Hierarchical embedding of multiple manifolds

An alternative strategy is to embed in a hierarchical fashion leading to a multi-scale analysis of images. Given a first level embedding \(\tilde{x}\) of the full image, we can then subdivide the image into smaller parts. As each sub-part is contained within the current level, we would expect embeddings at successive levels to be similar in some way.
We obtain the new embedding $\mathbf{x}$ of one of these sub-parts by amending the LE cost function such that the new embedding is also close to the embedding obtained at the previous level, in a manner comparable to manifold alignment:

$$C(x) = \sum_i \mu(x_i - \bar{x}_i)^2 + \sum_{ij} (x_i - x_j)^2 W_{ij}$$

(6)

It can be shown that the analytical solution to this is given by:

$$\mathbf{x} = (\mu \mathbf{I} + \mathbf{L})^{-1}\mu \mathbf{x}$$

(7)

By repeatedly dividing the image into smaller parts, we can use this to obtain consistent embeddings at different scales.

Figure 1 compares how the same patch can be embedded under both simultaneous and hierarchical schemes. In the simultaneous method, all patches are embedded at the same time and neighbouring patches are constrained to have similar embeddings. Under the hierarchical scheme, the image is subdivided into successively smaller patches, with each new patch being constrained to have a similar embedding to the larger patch within which it is contained.

Figure 1. Comparison of simultaneous patch embedding (L) and hierarchical embedding (R)

2.4 Cardiac images

In this paper we apply LE, on a global and local basis, to a time-varying sequence of 2D cardiac images. Using the conventional method of manifold learning, we construct a graph where each node represents a 2D frame at one point in time. Since we expect similar contrasts for all frames, we weight each pair of frames using their $\ell_2$ (Euclidean) norm distance $d = \sum_{ij} \|u_i - u_j\|^2$ between the intensities $u$ of each image. The final weight between the frames $i$ and $j$ is obtained by using a heat kernel: $W_{ij} = e^{-\sum_{ij} \|u_i - u_j\|^2/2}$.

For the local alignment, we divide each frame into a grid of regular patches. All patches are embedded simultaneously, maintaining intra-frame weights only between a patch and its four immediately-adjacent neighbours. It is reasonable to assume that cardiac and respiratory motions are the two main causes of variation between the frames, and so we select a two-dimensional embedding space.
3. EXPERIMENTS

Real time MR was performed using a balanced steady state free precession (SSFP) sequence with spatial resolution of 2x2x10mm and temporal resolution of 117ms per frame (FA/TE/TR = 20/1.2/2.4ms). Short axis, 2-chamber and 4-chamber cardiac views were acquired for 200 dynamics (the last 180 being used for analysis). ECG and respiratory traces were recorded alongside sequence event markers (including ECG R-wave peak detection and image acquisition points) in order to align images to physiological motion. Healthy volunteers were imaged and so we expect heart contraction to follow the ECG R-wave peaks.

We compare the results of the embeddings to bellows measurements of respiration and ECG R-wave peaks. Note that although we do not regard bellows motion as ground truth, it is the best independent estimate we have for respiration. For cardiac motion, we only know the temporal locations of the R-wave peak; it is known that physical contraction of the heart muscle typically occurs soon after this. To compare with the ECG, we regress the location in time of the R-wave peak and the location of the subsequent maximum embedding coordinate. The sum of the residuals of the regression determines the correspondence.

3.1 Embedding of whole image

In the first experiment we embed the whole image in the conventional way. Figure 2 shows how each dimension compares to the respiratory and cardiac cycles for a particular subject. Here the "signal" given by the embedding has been rescaled to the same range as the respiratory trace (as absolute embedding scales are not meaningful). As can be seen, dimensionality reduction has managed to separate cardiac and respiratory motions into each embedding dimension. The correlation coefficient between the respiratory trace and the first embedding dimension was found to be 0.81, while the sum of residual errors between the locations of the R-wave peaks and the maximum value of the second dimension was 51.

3.2 Independent embedding of individual patches

To investigate local variation, we divide each frame into 20mm square patches. Examples of a patch which incorporates the heart and one which does not are shown in Figure 3.2. The first practical issue when embedding each patch separately is how to align the output. Reducing even the same dataset repeatedly may result in equivalent embeddings which differ by rotations or flipping in the low-dimensional space. This makes comparing and interpreting results from independently-generated embeddings difficult. Figure 4 shows a comparison of each embedding dimension with the bellows trace and ECG for the two patches embedded independently. The correlation coefficient between the non-cardiac patch and the respiratory trace was found to be 0.49. For the cardiac patch, the cardiac motion is well-recovered in the embedding, with a sum of absolute regression residuals of 28, but the respiratory embedding appears to be corrupted by the strong influence of the cardiac motion with a correlation coefficient of 0.36.
3.3 Simultaneous embedding of multiple patches

In the next experiment we embed all patches in the image simultaneously, using the proposed method. The results of embedding all patches simultaneously are shown in Figure 5 for the same two patches as in the individual case. The correlation coefficient between the bellows trace and the non-cardiac and cardiac patches were found to be 0.82 and 0.60, respectively. The residual error of the cardiac patch embedding and the ECG trace is 38. Figure 6 (a) and (c) show the correlation coefficient between the bellows trace and the embedding for all patches, for two different subjects. Similarly, the residual error obtained when regressing the embedding with R-wave peak
locations are shown in 6(b) and (d). It can be seen that the non-cardiac patches more closely correlate to the respiratory signal and cardiac patches more closely correlates to the cardiac signal.

4. CONCLUSIONS

The traditional application of manifold learning to whole images suppresses local differences present in the data. We have presented a new framework for localised, patch-based manifold learning using Laplacian Eigenmaps. The techniques described create a coherent joint embedding of all patches in an image resulting in embeddings which vary according to location in the image. By using the joint embedding technique presented, we have been able to discover and separate local respiratory and cardiac motions within the abdominal and thoracic cavity, without the use of any prior information or training. The results correlate well with data obtained from respiratory bellows and ECG. Additionally, we have described methods for multi-scale analysis by successively dividing patches while maintaining the alignment to the current level.

REFERENCES

Figure 6. Patch-wise correlation coefficient with respiratory trace (a) (c) and regression residual with cardiac trace (b) (d) for two different subjects