Visual Exploration of Pathology Images by a Discrete Wavelet Transform preprocessed Locally Linear Embedding

Claudio Varini[1,2], Birgit Lessmann[1,2], Andreas Degenhard[1], Volkmar Hans[3] and Tim Nattkemper[2]

[1]Theoretical Physics, University of Bielefeld, Bielefeld, Germany
[2]Applied Neuroinformatics Group, University of Bielefeld, Bielefeld, Germany,
[3]Institute of Neuropathology, Evangelischesn Krankenhaus, Bielefeld, Germany
Email: varini@physik.uni-bielefeld.de

Abstract. The information content of large collections of histopathological images can be explored utilizing computer-based techniques that can help the user to explore the similarity between different brain tumor types. To visually inspect the degree of similarity between different tumors, we propose a combined approach based on the Discrete Wavelet Transform (DWT) and Locally Linear Embedding (LLE). The former is employed as a preprocessing utility, the latter achieves the dimensional reduction required for visualization.

1 Introduction

The automatic exploration of large collections of histopathological images is an actual field of research in computer science applied to medicine. Computer-based techniques can allow the user to find relations and similarities in the data. In particular the visual exploration of a dataset containing different types of tumor can provide valuable information concerning classification.

Techniques for dimensional data reduction offer appealing characteristics for the visual exploration of collections of histopathological images and of high-dimensional data in general. These techniques compute a projection of the data into a lower-dimensional space while best preserving the information content. The visualization of the low-dimensional space can reveal hidden structures in the data, e. g. the presence of meaningful patterns. A recently proposed algorithm for dimensional data reduction is Locally Linear Embedding (LLE) [1]. It attempts to reduce the dimension of the data while preserving the relationships between neighboring data points. i. e. neighboring data points in the high-dimensional space are projected by LLE into neighboring data points in the dimensional reduced space. In other words, the closeness of two data points in the dimensional reduced space reflects the similarity of the two respective data points in the high-dimensional space.

This property of neighborhood preservation in LLE can be useful for the visual exploration of a collection of histopathological images. Specifically, the

image space $I = \Re^N$, where N is the number of pixels of an image and each image represents a data point belonging to this space, can be reduced down to two or three dimensions in order to visualize the total information content. Images with similar structural characteristics represent neighboring data points in the image space that are expected to be mapped onto nearby points in the dimensional reduced space. In this way the similarity between several different images can be visualized at once in a customized display.

On the other hand, the image space is extremely high-dimensional ($10^4 - 10^6$) while the number of images, i. e. the number of points, is typically limited ($10^2 - 10^3$). This means the image space is actually very sparse and it may probe problematic for LLE to compute a faithful projection of such space. Indeed, for this purpose LLE must find a global hidden structure in the data and this in turn requires the data to be sufficiently dense. Ideally, the number of data points should grow exponentially with the dimension of the data space and this fact is known as "curse of dimensionality"in the machine learning literature [2]. To reduce the sparsity of the image space there are typically two possibilities, either to increase the number of data points (i. e. images) or to reduce the dimension of the image space itself. While the first possibility is often not achievable because of practical reasons, e. g. the limited number of patients and images, the second possibility can be done by appropriate techniques of signal processing such as the Discrete Wavelet Transformation (DWT). This is a technique for signal processing used to access the localized and scale-dependent information in signals or images. By computing wavelet based image features we are able to generate a lower-dimensional feature vector that encodes the image information.

In this work we present the visual exploration of a dataset of histopathological images of different brain tumors based on a combination of DWT and LLE. Each image is preprocessed by DWT and transformed into a feature vector. This DWT pre-processing step is used to reduce the dimension of the image space, thereby reducing its sparsity. These sets of feature vectors represent the LLE input that is reduced down to two dimensions. The dimensionally reduced dataset is finally visualized with customized colors encoding the type of tumor. This allows the user to explore the similarity among the images of the dataset.

2 Methods

2.1 Locally Linear Embedding

Given V D-dimensional vectors $\{\mathbf{X}_i\}$ as input, the LLE algorithm comprises three steps:

Step 1: it consists in assigning each data point $\mathbf{X}_i$ a predetermined number n of neighbors, typically according to the Euclidean distance.

Step 2: by minimizing the following error function

$$\Psi(W) = \sum_{i=1}^{V} |\mathbf{X}_i - \sum_{j=1}^{n} W_{ij}\mathbf{X}_j|^2 \qquad \text{with} \quad \sum_{j=1}^{n} W_{ij} = 1 \tag{1}$$

one computes the weights $\{W_{ij}\}$ that combined with the neighbors best approximate each data point $\mathbf{X}_i$.

Step 3: the weights are used to map each data point $\mathbf{X}_i$ into a d-dimensional vector $\mathbf{Y}_i$ with $d < D$, such that the following error function is minimized:

$$\Phi(Y) = \sum_{i=1}^{V} |\mathbf{Y}_i - \sum_{j=1}^{n} W_{ij}\mathbf{Y}_j|^2 \quad \text{with} \quad \frac{1}{V}\sum_{i=1}^{V} \mathbf{Y}_i\mathbf{Y}_i^T = I\,, \sum_{i=1}^{V} \mathbf{Y}_i = 0 \quad (2)$$

2.2 Discrete Wavelet Transform (DWT)

The DWT analysis enables the assessment of localized and scale-dependent information in signals and images [3]. A signal f is decomposed into a basis of shifted and dilated versions of a *mother wavelet* ψ [4]

$$f(x) = \sum_{(j,k)} d_{j,k}\psi_{j,k}(x)\,, \qquad \text{with} \quad \psi_{j,k}(x) = 2^{-j/2}\psi(2^{-j}x - k)\,. \quad (3)$$

The index j indicates the dilation or *scaling step* while k refers to translation or shifting. The wavelet coefficients $d_{j,k}$ are given by the scalar product $d_{j,k} = \langle f(x), \psi_{j,k}(x)\rangle$ or $d_{j,k} = \langle f(x), \tilde{\psi}_{j,k}(x)\rangle$ in the case of *biorthogonal wavelets* with the dual wavelet $\tilde{\psi}$ [4]. An efficient calculation of these coefficients is accomplished by the *Fast Wavelet Transform* (FWT), an algorithm allowing the coefficients to be calculated in a stepwise manner. To perform a FWT a scaling function $\phi(x)$ is required such that [4]

$$\phi(x) = \sqrt{2}\sum_{k} h(k)\phi(2x - k) \qquad \text{and} \qquad \psi(x) = \sqrt{2}\sum_{k} g(k)\phi(2x - k)\,. \quad (4)$$

The coefficients $h(k)$ and $g(k)$ are termed *Filter coefficients*. On the first scale the signal is decomposed into its *details* and the remaining signal, i.e. the *approximation*, reflecting the particular scale. The details are described by the wavelet coefficients of this scale while the approximation is represented by scaling coefficients corresponding to the scaling function. The procedure can be iterated by a further decomposition of the approximation into details and approximation of the next coarser scale. In two dimensions the DWT can be applied to each dimension separately, resulting in wavelet coefficients describing the horizontal, the vertical and the diagonal details.

3 Data and Experiments

The experimental dataset contains 84 histopathological images from 21 different cases of meningioma WHO grade I, a benign brain tumor. Each case relates to one of four meningioma subtypes, namely meningotheliomatous, fibroblastic, psammomatous and transitional, the latter showing intermediate features of the characteristics of the former classes. Each pathological image comprises 1300×1050 pixels at 400x magnification. Dividing each image into 16 subimages with 256×256 pixels, we derive a data set containing 1344 subimages.

The LLE is performed on two different input data: the first one is the set of all subimages where each subimage is treated as a point in a 256×256-dimensional data space; the second one is the set of texture describing features based on DWT. These features are computed as follows. Firstly, an average absolute coefficient is computed for each color channel and each of the finest six scales:

$$f_1(j) = \sum_o \sum_{k_x,k_y} |d_{j,o}(k_x,k_y)|, \quad j = 1,..,6, \quad \text{and} \quad o = o_1, o_2, o_3. \tag{5}$$

Here j is again the scale index, while index o indicates the orientation in the image. As explained in section 2.2 The indices o_1 and o_2 indicate coefficients in vertical or horizontal direction, while o_3 indicates the diagonal details. Secondly, we introduce a feature set f_2 describing a preferred orientation in the image.

$$f_2(j) = \frac{1}{f_1(j)}\Big(\sum_{k_x,k_y} |d_{j,o_1}(k_x,k_y)| - \sum_{k_x,k_y} |d_{j,o_2}(k_x,k_y)| + c\sum_{k_x,k_y} |d_{j,o_3}(k_x,k_y)|\Big). \tag{6}$$

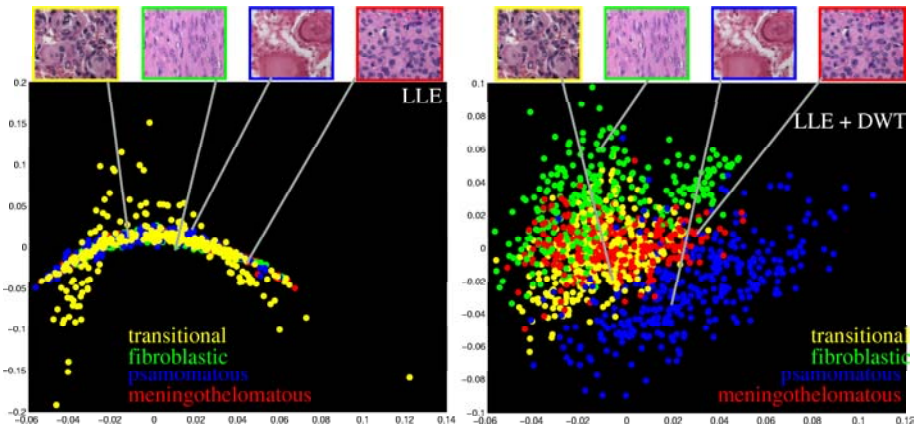
with $j = 1,..,6$. These computations result in two types of features for each of the finest six scales and each of the three color channels, leading to a set of 36 features for each subimage. Thus, the input of LLE comprises 1344 36-dimensional feature vectors.

Both input data are projected into a two-dimensional space and each data point is colored according to its respective histopathological diagnosis. The overlap among the clusters in the two embeddings has been quantified by computing for each data point the percentage of its nearest neighbors which belong to the same class. The average value computed over the entire dataset quantifies the degree of overlap among the various clusters. The value is comprised between 0 and 1. The closer to 1, the less the clusters overlap.

4 Results

The LLE projections of the subimages and of the DWL coefficients are shown in Fig. 1. Direct application of LLE to the subimages results in overlapping clusters lacking a clear structure. On the contrary, the DWT preprocessing of the images allows LLE to detect the presence of four different clusters. The meningotheliomatous, fibroblastic and psammomatous data points are clearly localized and quite well separated in the projected space. This is in agreement with the histopathological characteristics of the various tissue types. Specifically, psammomatous tissue is characterized by round calcifications that differ significantly from the round structures in meningothelomatous tissue and the elongated structures (cells and cell nuclei) in fibroblastic tissue (see the tissue samples in Fig. 1. At the same time, the transitional subtype cluster largely overlaps with meningotheliomatous and fibroblastic, in agreement with the fact that transitional meningiomas show features incorporating characteristics from the

Fig. 1. Two-dimensional projections obtained by LLE of the histopathological dataset without (left) and with (right) DWT processing.



other classes. The matching between the data visualization and the prelabelling of the data ensures that significant information has been preserved by the DWT preprocessing approach. The overlap between the clusters has been quantified as described above and the experimental values are 0.387 for the LLE projection and 0.567 for the LLE projection with the DWT preprocessing.

5 Discussion

Our results clearly show that LLE benefits from the DWT-based preprocessing. Specifically, applying LLE directly to the 256×256-dimensional space of the subimages proves problematic because of the sparsity of the space itself. By computing wavelet based features, the information of each subimage can be compacted into a lower dimensional feature vector. In this way the dimension of the data is significantly reduced and, in turn, the resulting feature space is more dense. This allows LLE to detect clinical relevant data structures in a more efficient manner.

References

1. Roweis ST, Saul LK. Nonlinear Dimensionality Reduction by Locally Linear Embedding. Science 2000;290:2323–2326.
2. Bellmann R. Adaptive Control Processes: A guided Tour. Princeton University Press; 1961.
3. Mallat S, Zhong S. Characterization of Signals from Multiscale Edges. Transactions on Pattern Analysis and Machine Intelligence 1992;14(7):710–732.
4. Daubechies I. Ten Lectures on Wavelets. CBMS-NFS Series Appl. Math., SIAM; 1991.