IMAGE MASKING SCHEMES FOR LOCAL MANIFOLD LEARNING METHODS

Hamid Dadkhahi and Marco F. Duarte

Dept. of Electrical and Computer Engineering, University of Massachusetts, Amherst, MA 01003

ABSTRACT

We consider the problem of selecting a subset of the dimensions of an image manifold that best preserves the underlying local structure in the original data. We have previously shown that masks which preserve the data neighborhood graph are well suited to global manifold learning algorithms. However, local manifold learning algorithms leverage a geometric structure beyond that captured by this neighborhood graph. In this paper, we present a mask selection algorithm that further preserves this additional structure by designing an extended data neighborhood graph that connects all neighbors of each data point, forming local cliques. Numerical experiments show the improvements achieved by employing the extended graph in the mask selection process.

Index Terms— Dimensionality Reduction, Manifold Learning, Locally Linear Embedding, Masking

1. INTRODUCTION

Recent advances in sensing technology have enabled a massive increase in the dimensionality of data captured from digital sensing systems. Naturally, the high dimensionality of data affects various stages of the digital systems, from acquisition to processing and analysis of the data. To meet communication, computation, and storage constraints, in many applications one seeks a low-dimensional embedding of the high-dimensional data that shrinks the size of the data representation while retaining the information we are interested in capturing. This problem of *dimensionality reduction* has attracted significant attention in the signal processing and machine learning communities.

For high-dimensional data, the process of data acquisition followed by a dimensionality reduction method is inherently wasteful, since we are often not interested in obtaining the full-length representation of the data. This issue has been addressed by *compressed sensing*, a technique to simultaneously acquire and reduce the dimensionality of sparse signals in a randomized fashion [2]. Compressed sensing provides a good match to the requirements of cyber-physical systems, where power constraints are paramount.

Novel designs of imaging sensors for low-power applications follow a pixel-level addressable architecture, yielding a power expense that is proportional to the number of pixels sensed [1, 3]. Thus, to achieve the benefits of compressive sensing with such architectures, we require an *a priori* selection of a subset of the image pixels to monitor for information extraction purposes. This necessitates a change in the premise of CS from projection-based embeddings to mask-based embeddings. The resulting *data-dependent image masking schemes* are therefore to be designed with the goal of preserving the information of interest present in the original data. Our work focuses in particular on nonlinear manifold models for images, which are commonly used in computer vision applications. Hence, in this context the information of interest refers to the geometric structure of the underlying manifold-modeled dataset. In practice, a compressed imaging system applies the proposed masking scheme in the following way. In the training stage, where power resources are not constrained, full-length data is collected. The masking pattern obtained via the proposed scheme is then used to only sense the pixels contained in the mask for subsequent captures in order to reduce the power consumption.

Previously, we have proposed an algorithm [4] for the derivation of an optimal masking pattern for Isomap [5], a global manifold learning algorithm. The performance of manifold learning from data masked by such a pattern has been shown to be not only superior to baseline masking schemes, but also similar to the performance of common projection-based dimensionality reduction schemes for sufficiently large mask sizes. However, there exist alternative manifold learning algorithms that are local in nature (e.g., Locally Linear Embedding, LLE [6]) and target a different type of geometric structure underlying the manifold model. We have seen that applying our previously designed masks with these algorithms does not accurately preserve the relevant local structure of the manifold. Thus, in this paper we consider the problem of designing masking patterns that preserve the geometric structure relevant to a local manifold learning algorithm, where we focus on LLE as a landmark example of the class of local manifold learning algorithms.

2. BACKGROUND

Manifolds and Linear Dimensionality Reduction: A set of data points $\mathcal{X} = \{x_1, x_2, \ldots, x_n\}$ in a high-dimensional am-

bient space $\mathbb{R}^d$ that have been generated by an ℓ-dimensional parameter correspond to a sampling of an ℓ-dimensional *manifold* $\mathcal{M} \subset \mathbb{R}^d$. Given the high-dimensional data set $\mathcal{X}$, we would like to find the parameterization that has generated the manifold. One way to discover this parametrization is to *embed* the high-dimensional data on the manifold to a low-dimensional space $\mathbb{R}^m$ so that the local geometry of the manifold is preserved. This process is known as *dimensionality reduction*, since $m \ll d$.

When the dimensionality reduction embedding is *linear*, it is defined by a matrix $\Phi \in \mathbb{R}^{m \times d}$ that maps the data in the ambient space $\mathbb{R}^d$ into a low-dimensional space $\mathbb{R}^m$. One such popular scheme is *principal component analysis* (PCA) [7], defined as the orthogonal projection of the data onto a linear subspace of lower dimension m such that the variance of the projected data is maximized.

Nonlinear Manifolds and Manifold Learning: Unfortunately, PCA fails to preserve the geometric structure of a *nonlinear manifold*, i.e., a manifold where the map from the parameter space to the data space is nonlinear. Particularly, since PCA arbitrarily distorts individual pairwise distances, it can significantly change the local geometry of the manifold. Fortunately, *nonlinear manifold learning (or embedding) methods* can successfully embed the data into a low-dimensional space while preserving the local geometry of the manifold, measured by a neighborhood-preserving criteria that varies depending on the specific method, in order to simplify parameter estimation.

The *Isomap* method aims to preserve the pairwise *geodesic distances* between data points [5]. The geodesic distance $d_G(x_i, x_j)$ is defined as the length of the shortest path between two data points $x_i, x_j \in \mathcal{M}$ along the surface of the manifold $\mathcal{M}$. Isomap first finds an approximation to the geodesic distances between each pair of data points by constructing a neighborhood graph in which each point is connected only to its k nearest neighbors; the edge weights are equal to the corresponding pairwise distances. For neighboring pairs of data points, the Euclidean distance provides a good approximation for the geodesic distance, i.e., $d_G(x_i, x_j) \approx \|x_i - x_j\|_2$ for $x_j \in \mathcal{N}_k(x_i)$, where $\mathcal{N}_k(x_i)$ designates the set of k nearest neighbors to the point $x_i \in \mathcal{X}$. For non-neighboring points, the length of the shortest path along the neighborhood graph is used to estimate the geodesic distance. Then, multidimensional scaling (MDS) [8] is applied to the resulting geodesic distance matrix to find a set of low-dimensional points that best match such distances.

As an alternative, the *locally linear embedding* (LLE) method retains the geometric structure of the manifold as captured by locally linear fits [6]. More precisely, LLE computes coefficients of the best approximation to each data point by a weighted linear combination of its k nearest neighbors. The weights $W = [w_{ij}]$ are found such that the squared

Euclidean approximation error is minimized:

$$W = \arg\min_W \sum_{i=1}^n \left\| x_i - \sum_{j: x_j \in \mathcal{N}_k(x_i)} w_{ij} x_j \right\|_2^2 \quad (1)$$

$$\text{subject to} \quad \sum_{j=1}^k w_{ij} = 1, \quad i = 1, \ldots, n.$$

LLE then finds a set of points in an m-dimensional space that minimizes the error of the local approximations given by the weights W. More precisely, LLE finds the set $\mathcal{Y} = \{y_1, y_2, \ldots, y_n\} \subset \mathbb{R}^m$ that minimizes the squared Euclidean error function

$$\mathcal{Y} = \arg\min_{\{y_i\}} \sum_{i=1}^n \left\| y_i - \sum_{j: x_j \in \mathcal{N}_k(x_i)} w_{ij} y_j \right\|_2^2 \quad (2)$$

$$\text{subject to} \quad \sum_{i=1}^n y_i = 0, \quad \frac{1}{n} \sum_{i=1}^n y_i y_i^T = I,$$

where the first and second constraints are to remove the degrees of freedom due to translation and scaling of the coordinates, in order to obtain a unique solution for the embedding. Note that LLE is considered as a local method, since the manifold structure is determined only by neighboring data points.

Linear Dimensionality Reduction for Manifolds: An alternative linear embedding approach to PCA is the method of *random projections*, where the entries of the linear dimensionality reduction matrix are drawn independently following a standard probability distribution such as normal Gaussian or Rademacher. One can show that such random projections preserve the relevant pairwise distances with high probability [9, 10], and so they preserve the structure relevant for manifold learning methods. Unfortunately, random embeddings are independent of the geometric structure of the data, and thus cannot take advantage of training data.

Manifold Masking for Isomap: A masking index set $\Omega = \{\omega_1, \ldots, \omega_m\}$ of cardinality m is defined as a subset of the dimensions $[d] := \{1, 2, \ldots, d\}$ of the high-dimensional space containing the original dataset. In [4], we developed schemes to obtain masking patterns for manifold-modeled data that preserve the global structure leveraged by Isomap. Such schemes rely on a set of *secants*, i.e., pairwise data point differences that have been normalized to lie on the unit sphere. Roughly speaking, the aim of the proposed manifold mask design was to minimize the distortion incurred by secants of neighboring data points, i.e.,

$$\mathcal{S}_k = \left\{ \frac{x_i - x_j}{\|x_i - x_j\|_2} : i \in [n], x_j \in \mathcal{N}_k(x_i) \right\}.$$

This gives rise to an integer program that minimizes the average or maximum secant norm distortion with respect to the expected value $\sqrt{m/d}$ of the normalized secants' norms after masking. This integer program is intractable but can

be approximated by a fast greedy algorithm, referred to as *Manifold-Aware Pixel Selection for Isomap (MAPS-Isomap)*. At iteration i of the algorithm, MAPS-Isomap finds a new dimension that, when added to the existing dimensions in Ω, causes the squared norm of the masked secant to match the expected value of $\frac{i}{d}$ as closely as possible. However, for a local manifold learning algorithm such as LLE, such a masking selection process is not effective. This is due to the fact that the LLE weights, in contrast to Isomap, depend not only on the distances between neighbors but also on the angles determined by each point and each pair among its neighbors.

3. MASKING STRATEGIES FOR LOCAL MANIFOLD LEARNING METHODS

We propose a greedy algorithm for selection of an LLE-aware masking pattern that attempts to preserve the weights w_{ij} obtained from the optimization in (1). Preserving these weights would in turn maintain the embedding $\mathcal{Y}$ found from (2) through the image masking process.

The rationale behind the proposed algorithm is as follows. The weights w_{ij} for $x_j \in \mathcal{N}_k(x_i)$ are preserved if both the lengths of the secants involving x_i (up to a scaling factor) and the angles between these secants are preserved. Geometrically, this can be achieved if the distances between all pairs of points in the set $\mathcal{C}_{k+1}(x_i) := \mathcal{N}_k(x_i) \cup \{x_i\}$ are preserved up to a scaling factor. Therefore, we define the *secant clique* for x_i as $\mathcal{S}_{k+1}(x_i) := \{x_{j_1} - x_{j_2} : x_{j_1}, x_{j_2} \in \mathcal{C}_{k+1}(x_i)\}$; our goal for LLE-aware mask selection is to preserve the norms of these secants up to a scaling factor.

To formulate our algorithm, we define a $3-$dimensional array B of size $c \times d \times n$, where c denotes the number of elements in each secant clique $\mathcal{S}_{k+1}(x_i)$. The array has entries $B(\ell, j, i) = s_\ell^i(j)^2$, where s_ℓ^i denotes the ℓth secant contained in $\mathcal{S}_{k+1}(x_i)$. In words, the array encodes the structure of the secant cliques: every 2-D slice of B, denoted by $B_i := B(:,:,i)$, contains the squared entries for the secants contained in the clique $\mathcal{S}_{k+1}(x_i)$, and the ℓth row of B_i corresponds to the ℓth secant in $S_{k+1}(x_i)$.

We now derive our LLE-aware masking algorithm. Define the d-dimensional mask indicator vector z as $z(j) = 1$ if $j \in \Omega$, and zero otherwise. The vector $\alpha = B_i z$ contains the squared norms of the masked secants from $S_{k+1}(x_i)$ as its entries. Similarly, the vector $\beta = B_i 1_d$ will contain the squared norms of the full secants in the same set. To measure the similarity between the original and masked secant norms (up to scaling), we use a normalized inner product commonly referred to as *cosine similarity measure*, defined as $\text{sim}(\alpha, \beta) := \frac{\langle \alpha, \beta \rangle}{\|\alpha\|_2 \|\beta\|_2}$. Maximizing the cosine similarity measure $\text{sim}(\alpha, \beta)$ promotes these two vectors being a scaled version of one another, i.e., the norms of the masked secants to approximately be equal to a scaling of the full secant norms. Note that since LLE is a local algorithm, the value of this scaling can vary over data points without incurring distortion of the manifold structure. In order to incor-

---

Algorithm 1 MAPS-LLE

1: **Inputs:** clique secant array B, mask size m
2: **Outputs:** masking index set Ω
3: **Initialize:** $\Omega \leftarrow \{\}$
4: $\alpha \leftarrow \sum_{j \in [d]} B(:,j,:)$
5: **for** $i = 1 \rightarrow m$ **do**
6: $\quad \theta \leftarrow \sum_{j \in \Omega} B(:,j,:)$
7: $\quad$ **for** $j \in \Omega^C$ **do**
8: $\quad\quad \beta \leftarrow \theta + B(:,j,:)$
9: $\quad\quad \lambda(j) \leftarrow \sum_{t \in [n]} \frac{\langle \alpha(:,t), \beta(:,t) \rangle}{||\alpha(:,t)||_2 ||\beta(:,t)||_2}$
10: $\quad$ **end for**
11: $\quad \omega \leftarrow \arg\max_{j \in \Omega^C} \lambda(j)$
12: $\quad \Omega \leftarrow \Omega \cup \{\omega\}$
13: **end for**

---

porate the cosine similarity measure for all data points, we maximize the sum of the aforementioned similarities for all data points as follows:

$$\hat{z} = \arg\max_z \sum_{i=1}^{n} \text{sim}(B_i z, B_i 1_d) \qquad (3)$$
$$\text{subject to} \quad 1_d^T z = m, z \in \{0,1\}^d.$$

The constraints above allow for z to yield an indicator function for the selection of m pixels from the image. Finding an optimal solution for z from the integer program (3) has a combinatorial (exponential) time complexity. An approximation can be obtained by greedily selecting the masking elements that maximize the value of the objective function of (3), one at a time. Thus, we propose the Manifold-Aware Pixel Selection for LLE (MAPS-LLE) algorithm, as shown in Algorithm 1. Note that lines $7-9$ of MAPS-LLE evaluate the integer program objective function (3) for each candidate pixel $j \in \Omega^C$ to be added to the mask Ω.

The MAPS-Isomap algorithm attempts to preserve the norms of the secants in the neighborhood graph up to a compaction factor of $\frac{m}{d}$. This in turn preserves the geodesic distances between all point pairs – as considered in Isomap – up to the same compaction factor. However, the aforementioned preservation is not suitable for LLE manifold learning, since the angles between neighboring secants may vary arbitrarily. The addition of clique secants in MAPS-LLE ensures that such angles cannot change arbitrarily. However, applying the same compaction factor as MAPS-Isomap imposes an unnecessay constraint on the masking algorithm. This can be easily seen by noting that LLE weights from (1) are invariant to a constant scaling factor, which could be different for each neighborhood. As such, the proposed MAPS-LLE algorithm relaxes the constraint of a global compaction factor and instead picks the best factor for each neighborhood, in an adaptive fashion. Hence, the compaction factor chosen by the MAPS-LLE algorithm could vary along the manifold. This intuitively agrees with local manifold learning algorithms, where only the local structure of the manifold is relevant.
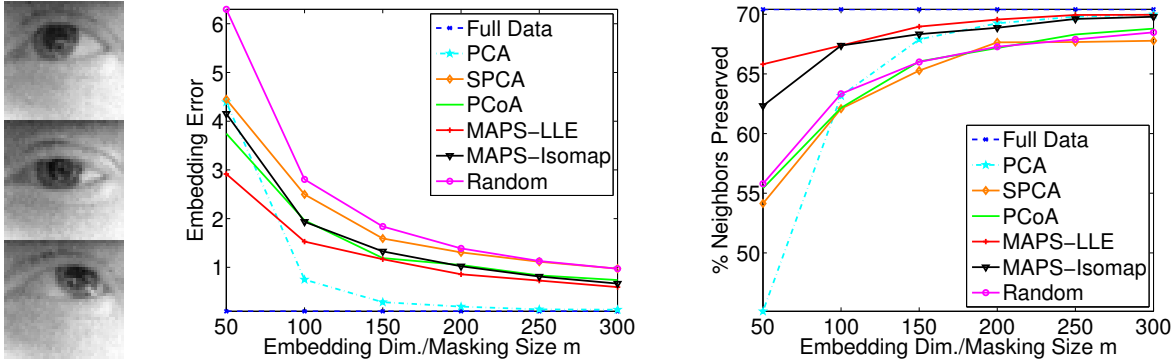
Fig. 1. Left: Example images from Eyeglasses dataset. Right: Performance comparison for linear embeddings (dashed lines) and masking algorithms (solid lines) with respect to original full-length data. Columns from left to right: residual variance as a function of m; percentage of preserved nearest neighbors for 20 neighbors.

4. NUMERICAL EXPERIMENTS

In this section, we present a set of experiments that compare the performance of the proposed algorithms to those in the existing linear embedding literature, in terms of preservation of the low-dimensional structure of a nonlinear manifold. For our experiments, we use a custom eye-tracking dataset from a computational eyeglass prototype. The *Eyeglasses* dataset corresponds to 40×40-pixel captures from a prototype implementation of computational eyeglasses that use the imaging sensor array of [3]. The dataset contains $n = 929$ images and the dimensionality of the learned manifold is $\ell = 2$.

We evaluate the methods described in Section 3, together with three baseline methods: *random masking*, where we pick an m-subset of the d data dimensions uniformly at random; *principal coordinate analysis (PCoA)*, where we select the indices of the m dimensions with the highest variance across the dataset [11]; and Sparse PCA (SPCA) [12, 13]. SPCA is a variation of PCA in which sparsity is enforced in the principal components. Since the support of the SPCA principal components need not be the same, we define the SPCA mask from support of the top SPCA principal component.

The algorithms are tested for linear embeddings of dimensions $m = 50, 100, 150, 200, 250, 300$; for the masking algorithms, m provides the size of the mask (number of pixels preserved), while for the linear embedding algorithms of Section 2, m provides the embedding dimensionality. Note that since the linear embeddings employ all d dimensions of the original data, the latter algorithms have an intrinsic performance advantage against the former. The performance of random masking is averaged over 100 independent draws in each case.

In order to match the practical application of masking schemes, we employ the following experimental setup. We divide the dataset into training and testing subsets of (almost) equal size in a random fashion. For each tested masking/linear embedding algorithm, we learn the mask/embedding using the training subset. We then apply LLE directly on the

masked/ embedded and the original (full) testing images. In order to remove the dependence of the experiments on the random division of training/testing data, we repeat this experiment for $T = 10$ such training/testing random subsets of the dataset and average the results of the experiments.

We measure the performance of the manifold embedding obtained from the masked dataset using two different performance metrics. First, we consider the following *embedding error*. Suppose the pairs $(\mathcal{X}, \mathcal{Y})$ and $(\mathcal{X}', \mathcal{Y}')$ designate the ambient and embedded set of vectors for full and masked data, respectively. Having found the weights $w_{i,j}$ from the full data via (1), we define the embedding error for the masked data as

$$e = \sum_{i=1}^{n} \left\| y_i' - \sum_{j:x_j \in \mathcal{N}_k(x_i)} w_{ij} y_j' \right\|_2^2. \quad (4)$$

The rationale behind this definition of the embedding error is that, ideally, the embedded vectors y_i' obtained from masked images should provide a good linear fitting using the neighborhood approximation weights obtained from the original (full) images. In other words, (4) finds the amount of deviation of $\mathcal{Y}'$ from $\mathcal{Y}$, which minimizes the value of this score, cf. (2). Second, we use the percentage of preserved nearest neighbors, similar to [14]. More precisely, for a given neighborhood of size k, we obtain the fraction of the k-nearest neighbors in the full d-dimensional data that are among the k-nearest neighbors when the masked images are considered.

We display the results of manifold learning in Figure 1. We observe that MAPS-LLE significantly outperforms random sampling, SPCA, and PCoA. Furthermore, MAPS-LLE outperforms MAPS-Isomap for small and moderate mask sizes of up to $m = 200$. Surprisingly, we see that for sufficiently large values of m the performance of MAPS approaches or matches that of the linear embedding algorithms, even though the embedding feasible set for masking methods is significantly reduced.

Additional simulations and datasets are presented in [11]. A Matlab toolbox implementing these simulations is available online [15].

5. REFERENCES

[1] A. Mayberry, P. Hu, B. Marlin, C. Salthouse, and D. Ganesan, "iShadow: Design of a wearable, real-time mobile gaze tracker," in *Int. Conf. Mobile Systems, Applications and Services (MobiSys)*, Bretton Woods, NH, June 2014, pp. 82–94.

[2] D. L. Donoho, "Compressed sensing," *IEEE Trans. Info. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[3] Centeye, Inc., "Stonyman and Hawksbill vision chips," Available online at http://centeye.com/products/current-vision-chips-2, Nov. 2011.

[4] H. Dadkhahi and M. F. Duarte, "Masking schemes for image manifolds," in *IEEE Statistical Signal Proc. Workshop (SSP)*, Gold Coast, Australia, June 2014, pp. 256–259.

[5] J. B. Tenenbaum, V. de Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, Dec. 2000.

[6] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, Dec 2000.

[7] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*, Springer-Verlag, Secaucus, NJ, 2006.

[8] T. Cox and M. Cox, *Multidimensional Scaling*, Boca Raton : Chapman & Hall/CRC, 2001.

[9] C. Hegde, M. Wakin, and R. Baraniuk, "Random projections for manifold learning," in *Neural Info. Proc. Systems (NIPS)*, Vancouver, BC, Dec. 2007, pp. 641–648.

[10] Y. Freund, S. Dasgupta, M. Kabra, and N. Verma, "Learning the structure of manifolds using random projections," in *Neural Info. Proc. Systems (NIPS)*, Vancouver, BC, Dec. 2007, pp. 473–480.

[11] H. Dadkhahi and M. F. Duarte, "Masking strategies for image manifolds," *Preprint, Available online at* `http://www.ecs.umass.edu/~mduarte/images/MSIM2014.pdf`, Sept. 2014.

[12] H. Zou, T. Hastie, and R. Tibshirani, "Sparse principal component analysis," *J. Computational and Graphical Statistics*, vol. 15, no. 2, pp. 262–286, 2006.

[13] A. d'Aspremont, L. El Ghaoui, M. Jordan, and G. Lanckriet, "A direct formulation of sparse PCA using semidefinite programming," *SIAM Review*, vol. 49, no. 3, pp. 434–448, July 2007.

[14] C. Hegde, A. Sankaranarayanan, W. Yin, and R. Baraniuk, "NuMax: A convex approach for learning near-isometric linear embeddings," *Submitted to IEEE Trans. Signal Proc.*, 2014.

[15] H. Dadkhahi, "Masking strategies for image manifolds toolbox," `http://www.ecs.umass.edu/~mduarte/Software.html`, Sep. 2014.