Visual Manifold Sensing

Irina Burciu, Adrian Ion-Mărgineanu, Thomas Martinetz, and Erhardt Barth

University of Lübeck, Institute for Neuro- and Bioinformatics, Ratzeburger Allee 160, D-23562 Lübeck, Germany

ABSTRACT

We present a novel method, Manifold Sensing, for the adaptive sampling of the visual world based on manifolds of increasing but low dimensionality that have been learned with representative data. Because the data set is adapted during sampling, every new measurement (sample) depends on the previously acquired measurements. This leads to an efficient sampling strategy that requires a low total number of measurements. We apply Manifold Sensing to object recognition on UMIST, Robotics Laboratory, and ALOI benchmarks. For face recognition, with only 30 measurements - this corresponds to a compression ratio greater than 2000 - an unknown face can be localized such that its nearest neighbor in the low-dimensional manifold is almost always the actual nearest image. Moreover, the recognition rate obtained by assigning the class of the nearest neighbor is 100%. For a different benchmark with everyday objects, with only 38 measurements - in this case a compression ratio greater than 700 - we obtain similar localization results and, again, a 100% recognition rate.

Keywords: Manifold Sensing, Locally Linear Embedding, Compressed Sensing, efficient sensing, adaptive sensing

1. INTRODUCTION

This work takes some inspiration from Compressed Sensing (CS)[1] and derives an efficient sensing scheme in the context of Active Vision. The objective is to develop appropriate learning schemes for adaptive sensing and to increase the efficiency of visual sensing. From a different perspective, we are dealing with the problem of how to efficiently sample the world under the constraint of a limited bandwidth. In human vision the bandwidth is limited, for example, by the capacity of the optic nerve, and in technical systems by the performance and cost of hardware. The method Manifold Sensing (MS) proposed here is inspired by the sampling model of biological systems, which efficiently sense the information required for a particular task.

CS is a coding principle which is based on the existence of an image representation that provides a sparse code.[2] The method does not adapt to particular datasets. It was shown that CS can be applied successfully on natural images, which can be encoded sparsely. The idea is that sensing with a random sensing matrix can significantly reduce the bandwidth required for an exact representation and reconstruction of the image.[3] Our MS method is related to CS, in the sense that every measurement (note that we use the expressions *measurement* and *sample* as synonyms to denote the sensing values, not to be confused with a data sample) is a weighted sum of the original unknown signal (the world to be sensed). MS, however, involves a two-fold adaptation process: (i) the algorithm adapts to particular datasets, and (ii) every new measurement depends on the previously acquired measurements.

MS is based on the assumption that, due to redundancies, natural images lie on non-linear low-dimensional manifolds.[4,5] Images trace out non-linear curves embedded in the image space. In case there are changes in scale, illumination and other sources of continuous variability, then the images lie on low-dimensional manifolds rather than on the simple one-dimensional curves.[4] MS is based on manifolds of low dimensions and step by step we increase the dimension in which we learn a new low-dimensional manifold. This means that considering the information that we already have, we move into a higher dimension and we take a number of samples and so on. The number of samples we need for solving a particular recognition task should be as small as possible. An alternative hierarchical sensing scheme is presented by Schütze et al.[6]

Further author information: (Send correspondence to Irina Burciu)
Irina Burciu: E-mail: irina.burciu@inb.uni-luebeck.de
Published in Proceedings SPIE Electronic Imaging, Vol. 9014, Human Vision and Electronic Imaging XIX, 2014.

2. THE APPROACH

This section presents the main steps of the MS method which is currently based on the iterative Locally Linear Embedding (LLE) algorithm but could be used with alternative manifold learning algorithms.

2.1 The main steps of MS

We consider a given data set $\mathcal{D}$ with P points of dimension D. For the given data set we learn a manifold of dimension N_i, typically 2 or 3, by using the LLE[7,8] algorithm. This corresponds to the learning step shown in Figure 1. LLE has only one free parameter, the number of nearest neighbors (for each data point of $\mathcal{D}$) given here by r. Any new data point, i.e., a test point outside $\mathcal{D}$, is first projected on the learned manifold. This is represented in Figure 1 by the projection step. The adaptive data set $\mathcal{D}_k$ (k denotes the current iteration) is a subset of the original data set embedded in the same dimension N_i and it is illustrated in Figure 1 as part of the adaptation step. $\mathcal{D}_k$ is given by:

$$\mathcal{D}_k : \mathcal{D} \rightarrow \mathcal{D}' \rightarrow \mathcal{D}'' \rightarrow \dots \tag{1}$$

The process of embedding the adaptive data set $\mathcal{D}_k$ in a manifold of dimension N_i is repeated n_i times with different sizes of $\mathcal{D}_k$, as it can be seen in the second row of Figure 1. Thus, n_i denotes how often we iterate the embedding in a particular dimension N_i, before moving to a manifold with higher dimension N_g, $N_g > N_i$, and until we reach a predefined maximal dimension $N_{\max}$.

Starting with the LLE algorithm and considering the steps presented before, we implement an iterative LLE algorithm, which is fundamental for the MS method.

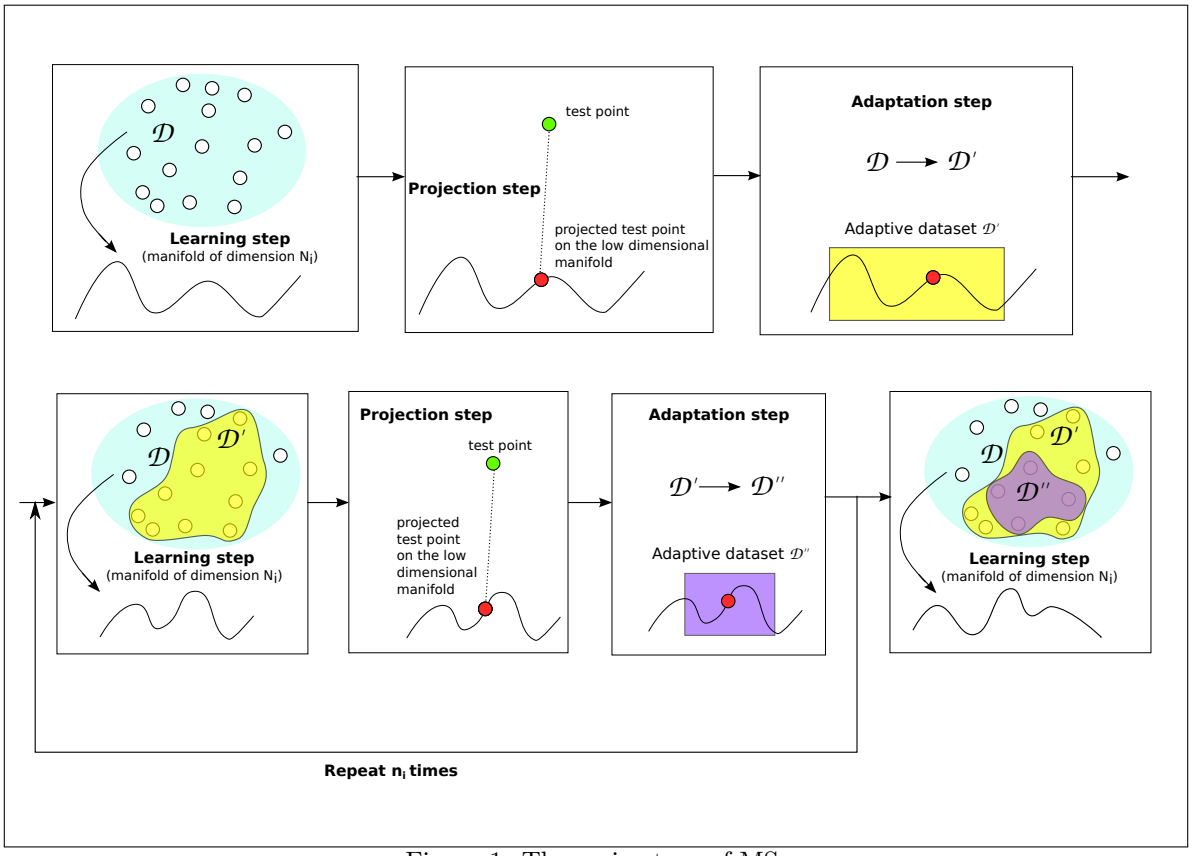


Figure 1: The main steps of MS.

2.2 Iterative LLE algorithm

This section explains how the MS method works based on the implementation of an iterative LLE algorithm. We focus on the steps presented in Figure 1 and in Algorithm 1. We introduce the following notations for Algorithm 1: Y represents the low dimensional embedding coordinates of the data points in $\mathcal{D}_k$; y^j denotes the j-th element of Y, i.e., the coordinates of a particular point in $\mathcal{D}_k$; A^+ is the pseudo-inverse matrix; Y_{test}: projected test data on the low dimensional manifold; y^l_{test}: the l-th element of Y_{test}, i.e., the coordinates of a particular point in X_{test}; $d(y^l_{\text{test}}, y^j)$: the Euclidean distance between y^l_{test} and y^j.

Algorithm 1 MS - Iterative LLE

Input: $\mathcal{D}_k$ - adaptive data set (Equation 1 where $\mathcal{D}$ is the data set represented in a Haar basis)
r - number of nearest neighbors for each data point
N_i - dimension of the manifold
X_{test} - test data
n_i - number of iterations computed for dimension N_i
Learning manifold
$Y \leftarrow \text{LLE}(\mathcal{D}_k, r, N_i)$[7]
Learning pseudo-inverse matrix
$A^+ \leftarrow Y \cdot \mathcal{D}_k^{-1}$
Projecting test data
$Y_{\text{test}} \leftarrow A^+ \cdot X_{\text{test}}$
Adapting data set
compute all $d(y^l_{\text{test}}, y^j)$
select neighborhood $\mathcal{D}_k$, size$(\mathcal{D}_k) = q$ (Equation 2)
repeat n_i times

We consider the adaptation step where $\mathcal{D}_k$ has a decreasing number q of data points. This is graphically shown in Figure 1 by: $\mathcal{D} \rightarrow \mathcal{D}' \rightarrow \mathcal{D}''$. In order to learn the embeddings of the data for all $\mathcal{D}_k$, one would require an hierarchal partitioning of the data set. Currently, we do not use such a partitioning but define the subsets by the size q of the neighborhood and use the following heuristics for q:

$$q = (n - k + 2) \cdot r \tag{2}$$

Thus, the number of data points in $\mathcal{D}_k$ depends on: (i) the number r of neighbors that we select for the LLE algorithm, (ii) the current iteration k and the total number of iterations n that we use. Note that, at each iteration k, $k > 1$, the number of points for $\mathcal{D}_k$ is reduced.

Finally, the parameters of MS are: (i) the number r of the nearest neighbors, (ii) the (decreasing) size q of the adaptive data set $\mathcal{D}_k$, and (iii) the dimensions N_i of the manifolds.

Currently, learning the manifolds with LLE and sensing with the pseudo-inverse are performed with the data represented in an Haar basis in order to have a sparser representation of the data.

3. EXPERIMENTS

We worked under the assumption that the goal of sensing is to acquire information for a particular task and not for representing the world. We therefore applied the MS method for both face recognition and object recognition. We evaluated the performance of MS on three benchmarks, two for face-recognition and one for the recognition of everyday objects. We measured the performance of MS by computing: (i) the Signal to Noise Ratio (SNR) between the test images and the corresponding nearest images in $\mathcal{D}$ and (ii) the recognition rate.

Regarding the SNR, note that the selection of the nearest image is based on the distances in the low dimensional embedding but the actual SNR is computed in the original image space as a measure for how much information is retained in the low-dimensional embedding. The recognition rate is based on a classifier that assigns each low-dimensional y^l_{test} to the class of the nearest y^j in the low-dimensional data set.

We compared MS with: (i) Principal Component Analysis (Non-iterative PCA) presented in Algorithm 2: new data points are projected on the M principal components of $\mathcal{D}$, and (ii) Iterative PCA presented in Algorithm 3: the Algorithm follows the same steps as MS (see Algorithm 1) but any new data point is simply projected on the $N_i \cdot n_i$ principal components of $\mathcal{D}_k$, which correspond to the $N_i \cdot n_i$ largest eigenvalues of $\mathcal{D}_k$. The notations used in Algorithm 2 and Algorithm 3 are the following: princomp($\mathcal{D}$) performs principal component analysis on $\mathcal{D}$ and returns the principal component coefficients stored in the matrix U of size $(D \times D)$; Y are the given data from the adaptive data set $\mathcal{D}_k$ projected on the $N_i \cdot n_i$ principal components of $\mathcal{D}_k$ that correspond to the $N_i \cdot n_i$ largest eigenvalues of $\mathcal{D}_k$; y^j is the j-th element of Y; Y_{test} are the test data projected on the M (respective $N_i \cdot n_i$) principal components of $\mathcal{D}$ (respective $\mathcal{D}_k$) that correspond to the M (respective $N_i \cdot n_i$) largest eigenvalues of $\mathcal{D}$ (respective $\mathcal{D}_k$); y^l_{test} is the l-th element of Y_{test}; $\mathcal{D}_k^{\text{T}}$ is the transpose of $\mathcal{D}_k$; $d(y^l_{\text{test}}, y^j)$ denotes the Euclidean distance between y^l_{test} and y^j.

Algorithm 2 Non-iterative PCA

Input: $\mathcal{D}$ - data set of size $(P \times D)$
$M = \sum n_i \cdot N_i$ - total number of measurements required for MS (Algorithm 1)
X_{test} - test data
Learning Karhunen-Loeve matrix
1: $U \leftarrow \text{princomp}(\mathcal{D})$
Projecting test data
2: $Y_{\text{test}} \leftarrow U(1 : M, :) \cdot X_{\text{test}}$

Algorithm 3 Iterative PCA

Input: $\mathcal{D}_k$ - adaptive data set (Equation 1)
N_i - dimensions of the manifolds used in MS (Algorithm 1)
X_{test} - test data
n_i - number of iterations computed for dimension N_i (same as in Algorithm 1)
Learning Karhunen-Loeve matrix
1: $U \leftarrow \text{princomp}(\mathcal{D}_k)$
2: $Y \leftarrow U(1 : N_i, :) \cdot \mathcal{D}_k^{\text{T}}$
Projecting test data
3: $Y_{\text{test}} \leftarrow U(1 : N_i, :) \cdot X_{\text{test}}$
Adapting data set
4: compute all $d(y^l_{\text{test}}, y^j)$
5: select neighborhood $\mathcal{D}_k$, size$(\mathcal{D}_k) = q$ (Equation 2)
6: repeat n_i times

Any new data point is projected on the learned manifold by using the pseudo-inverse matrix that minimizes the mean projection error on $\mathcal{D}$ and after that we define a neighborhood (considering the location of the new data point) and by that, a subset of the original data set, which can be embedded in the same dimension as the manifold, that is N_i. The pseudo-inverse is needed for the projection of the test points because the test images are unknown, and thus we cannot use LLE to find the low-dimensional coordinates of the test points.

The goal of doing iterations in the same dimension is to obtain a better local linear approximation of the manifold and thus a better projection. The process of iterating in one particular dimension and then moving to the next higher dimension is illustrated in Figure 2.

Currently, we do not have any formal criterion for how to proceed from lower- to higher-dimensional manifolds, and for how to choose r and q. In order to evaluate the MS method, we consider a vector of N_i components which corresponds to the number of dimensions of the learned manifolds. Each component of the vector tells us how many n_i iterations we have to compute for each dimension N_i. For simplicity, we call this vector representation a combination. Figure 2 shows a particular case of a combination with five components. For each iteration we compute a different number of iterations, n_i, $i = 1 : 5$.

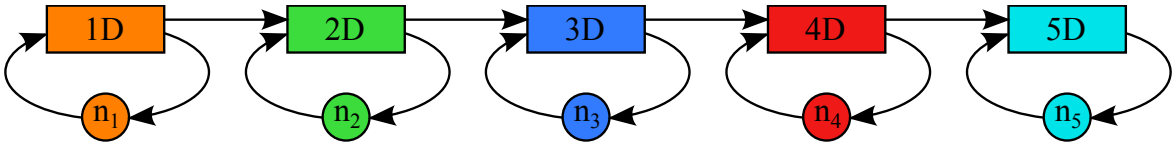


Figure 2: Graphical representation of the different combinations used to evaluate MS: we iterate n_1 times in the dimension $D = 1$, then n_2 times in the dimension $D = 2$, and so on up to $D = 5$.

In our simulations all the combinations have five components (dimensions) and for each dimension we compute maximum five iterations, $N_{\max} = 5$. We define n as the total number of iterations, $\sum n_i$, with n_i the number of iterations in dimension N_i, $N_i = 1 : N_{\max}$. The total number of measurements is $M = \sum N_i \cdot n_i$. As an example, we consider the combination 0 0 1 3 5 presented in Table 1.

Table 1: Example for the combination 0 0 1 3 5

N_i (dimensions)	5
$N_{\max}$ (maximum iterations)	5
n_i (number of iterations in dimension N_i)	$n_1 = 0$, $n_2 = 0$, $n_3 = 1$, $n_4 = 3$, $n_5 = 5$
n (total number of iterations)	$0 + 0 + 1 + 3 + 5 = 9$
M (number of measurements)	$1 \cdot 0 + 2 \cdot 0 + 3 \cdot 1 + 4 \cdot 3 + 5 \cdot 5 = 40$

For the combination illustrated in Table 1 we compute the following steps: first we compute one iteration in the 3rd dimension, we consider the result obtained after the first step and we go on to the 4th dimension and we compute one iteration. We proceed in the same way with the result obtained after the second step and we compute another iteration in the same dimension, the 4th dimension. We go on and compute another iteration in the same dimension, the 4th dimension. With the result obtained before we move to the 5th dimension and step by step we compute $n_5 = 5$ iterations.

To explore the potential of MS we applied the iterative LLE algorithm on different combinations using different values for r starting from 30 to 80 neighbors with steps of 10. The decreasing size of the transferred neighborhoods, q, is an internal parameter of the iterative LLE algorithm. It is defined by Equation (2) in such a way that the size of the adaptive data set will always be larger than the number of neighbors used in LLE.

4. RESULTS

The first benchmark we considered for evaluating our MS method was the UMIST[9] database with faces (twenty different persons in different poses and a total of 1000 images of size 256×256 pixel). We evaluated MS by computing (i) the Signal to Noise Ratio (SNR) between the test image and the image that was the nearest neighbor on the learned manifold, and (ii) the person recognition rate (test images assigned to the class of the nearest neighbor). With only 30 measurements, i.e. a compression ratio greater than 2000, we obtained an average SNR over 20 test images (one per class) of 22.70 dB (the best possible average SNR for the database, when the correct nearest neighbor was identified in all cases, was 22.73 dB). On the same data, PCA with 30 components yielded an average SNR of 22.60 dB. The recognition rate for MS and PCA with 30 measurements was 100%, and for Iterative PCA 85 %. We obtained this result for the combination 0 2 1 2 3 with 40 neighbors. Figure 4 shows how the pseudo-inverse matrix evolved after each iteration computed with the MS method for the image test from Figure 3.

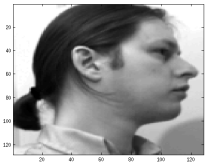


Figure 3: Image test - UMIST[9] database.

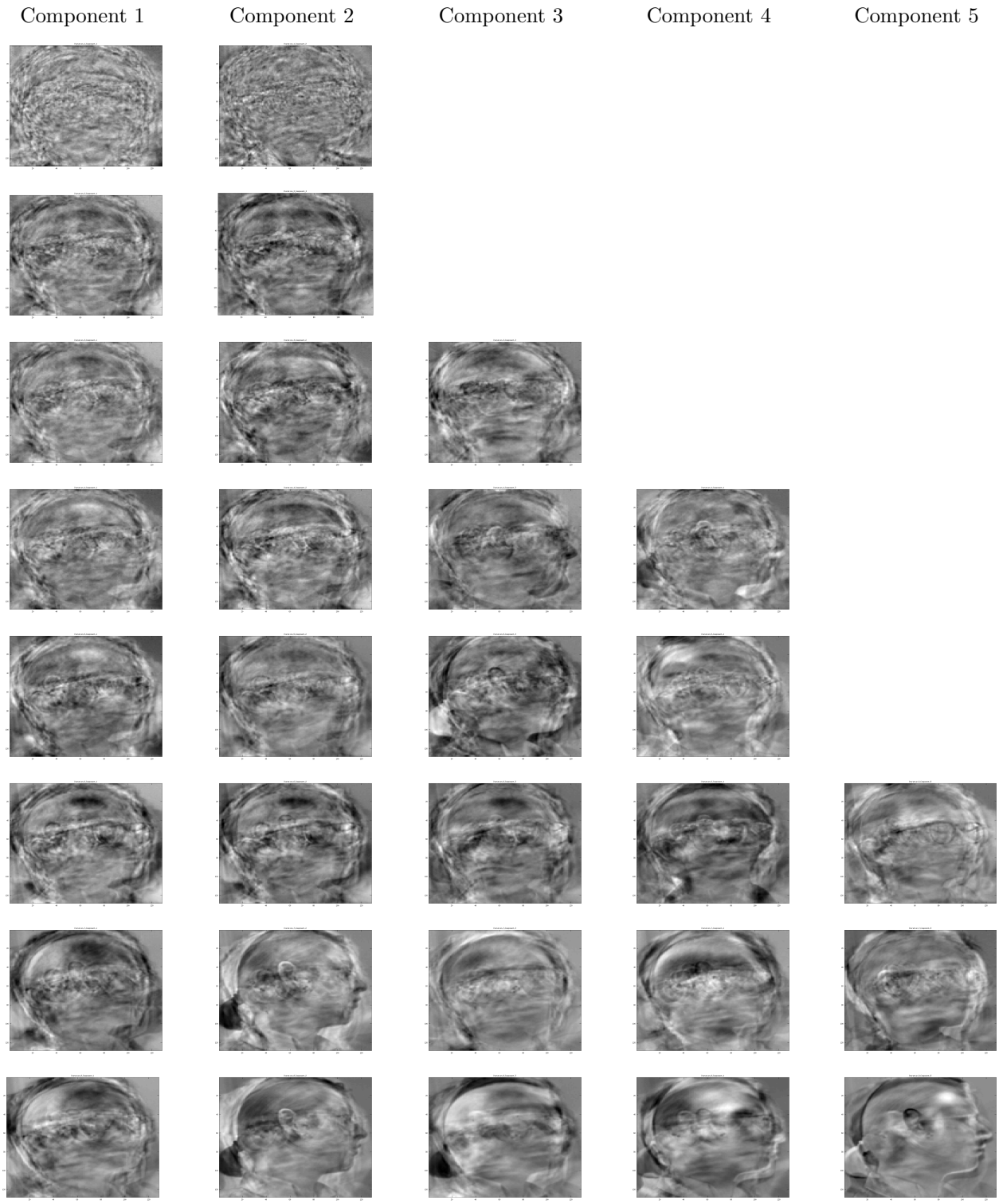


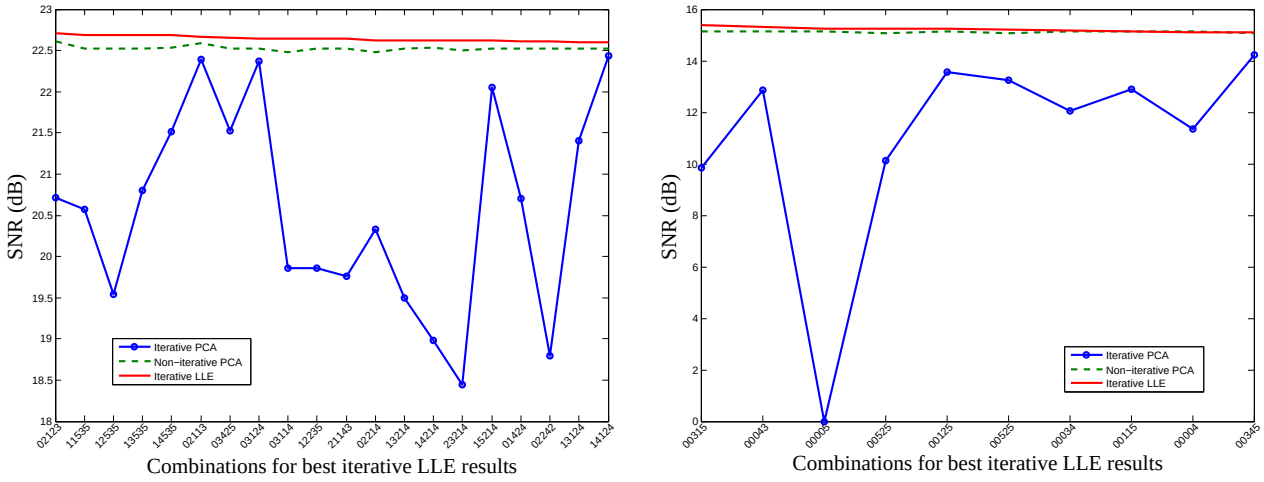
Figure 4: Learned pseudo-inverse matrices for the combination 02123 with 40 neighbors using the MS method. Each row shows the components (in descending order of the singular values) of the pseudo-inverse matrix for each iteration. These images represent the basis in which MS is actually sensing. Note the evolution from rather random to more specific templates (projection axes).

For the combination 0 0 1 0 0, i.e. with only 3 measurements, the recognition rate was 75%. We obtained similar results on a different database from Robotics Laboratory,[10] which contains face images of twenty different persons in different poses (a total of 1300 images of size 640 x 480 pixel).

We also evaluated MS for object recognition on the ALOI (Amsterdam Library of Object Images)[11] database. We considered, from this database, a subset of images of the first twenty objects. These are everyday objects images with different rotation angles resulting in a total of 1400 images of size 192×144 pixel. With only 38 measurements, i.e. a compression ratio greater than 700, we obtained an average SNR of 15.39 dB (in this case the best possible SNR was 15.49 dB, PCA with 38 components yielded 15.15 dB) and a 100% recognition rate. The results we obtained by applying MS on the two different benchmarks for object and face recognition are summarized in Table 2. For each benchmark we mentioned the combination for which we got the best result and we also compared the average SNR obtained for MS with the values obtained for PCA and with the best possible SNR for the respective benchmark. A detailed comparison between MS, Iterative PCA, and Non-iterative PCA for both face and object recognition is shown in Figure 5. Note that the MS method yielded better results than both, PCA and Iterative PCA.

Table 2: Best results obtained by applying MS for UMIST[9] and ALOI[11] benchmarks.

Benchmark	Combination	r	Average SNR (dB)	Recognition rate	M	Compression ratio
UMIST (256×256)	0 2 1 2 3	40	MS: 22.70 dB PCA: 22.60 dB Best possible: 22.73 dB	100 %	30	> 2000
ALOI (192×144)	0 0 3 1 5	80	MS: 15.39 dB PCA: 15.15 dB Best possible: 15.49 dB	100 %	38	> 700



(a) (b)

Figure 5: Iterative LLE (MS) vs. PCA for (a) UMIST[9] and (b) ALOI[11] database.

5. CONCLUSIONS AND DISCUSSION

We presented a novel method, Manifold Sensing, for sampling natural images based on learning manifolds of increasing dimensions. MS is a hierarchical method that iteratively localizes data points in low-dimensional manifolds of increasing dimensions. The sampling strategy is learned for a particular data set, a procedure that reduces the number of required samples. By data adaptation, every new sample depends on all previously

acquired samples. We evaluated the performance of MS on three benchmarks, two for face-recognition and one for the recognition of everyday objects. Thus, the information gathered during sensing was quantified by the recognition performance that it enabled. In other words, the acquired samples were mainly assessed by how much they contributed to a particular task and not by how accurate they represented the world. However, we also evaluated the distance to the nearest neighbor in the data set, a measure that corresponds to a reconstruction error.

6. FUTURE WORK

Future development of the MS method for sampling natural images includes various improvements, e.g. different algorithms for learning a manifold, a better method for projecting a new data point on the learned manifold, extension to dynamic scenes, etc. Besides providing an effective sensing strategy for known environments, e.g., in robotics, we expect these results provide new insights to visual processes such as retinal and cortical projections,[12] peripheral vision, gist, and eye movements. The traditional view of visual processing is that first the information is processed locally and after that the local features are integrated to a global percept. "Gist" is a more recent and alternative approach, describing a strategy which first performs a global and fast recognition of the scene and then proceeds to more refined sampling and recognition.[13] One could argue that human vision employs similar strategies since we are often capable of providing a "gist" of the scene before processing the details. In this context, it seems particularly striking that acceptable face recognition is possible with only three measurements, i.e., with only three samples of the visual world.

ACKNOWLEDGMENTS

The research is funded by the DFG Priority Programme SPP 1527, grant number MA 2401/2-1, by the ESF, POSDRU/86/1.2/S/61756, and by the Graduate School for Computing in Medicine and Life Sciences funded by Germany's Excellence Initiative [DFG GSC 235/2].

REFERENCES

[1] Candès, E. J. and Wakin, M., "An introduction to compressive sampling," *IEEE Signal Processing Magazine* **25**(2), 21–30 (2008).
[2] Olshausen, B. A. and Field, D. J., "Natural image statistics and efficient coding," *Network: Computation in Neural Systems* **7**(2), 333–339 (1996).
[3] Donoho, D. L., "Compressed sensing," *IEEE Transactions on Information Theory* **52**(4), 1289–1306 (2006).
[4] Seung, S. and Lee, D., "The manifold ways of perception," *Science* **290**(5500), 2268–2269 (2000).
[5] Zetzsche, C., Barth, E., and Wegmann, B., "The importance of intrinsically two-dimensional image features in biological vision and picture coding," in [*Digital Images and Human Vision*], Watson, A. B., ed., 109–38, MIT Press (Oct. 1993).
[6] Schütze, H., Barth, E., and Martinetz, T., "An adaptive hierarchical sensing scheme for sparse signals," in [*Human Vision and Electronic Imaging XIX*], Rogowitz, B. E. and de Ridder, T. N. P. H., eds., *Proc. of SPIE Electronic Imaging* **this volume** (2014).
[7] Roweis, S. and Saul, L., "Nonlinear Dimensionality Reduction by Locally Linear Embedding," *Science* **290**(5500), 2323–2326 (2000).
[8] L. K. Saul, S. T. R., "Think globally, fit locally: Unsupervised learning of nonlinear manifolds," *Journal of Machine Learning Research* **4**, 119–155 (2003).
[9] "UMIST database." http://www.sheffield.ac.uk/eee/research/iel/research/face.
[10] "Database for face recognition - Robotics Laboratory." http://robotics.csie.ncku.edu.tw/Database.htm.
[11] Geusebroek, J. M., "Amsterdam Library of Object Images (ALOI)." http://staff.science.uva.nl/ aloi/.
[12] W. Coulter, C. Hillar, G. I. and Sommer, F., "Adaptive compressed sensing: A new class of self-organizing coding models for neuroscience," *IEEE International Conference on Acoustics Speech and Signal Processing* **5370**, 5494–5497 (2010).
[13] Oliva, A. and Torralba, A., "Modelling the shape of the scene: A holistic representation of the spatial envelope," *International Journal of Computer Vision* **42**, 145–175 (2001).