Manifold Integration with Markov Random Walks

Heeyoul Choi[†] and **Seungjin Choi**[‡] and **Yoonsuck Choe**[†]
[†]Department of Computer Science, Texas A&M University
3112 TAMU, College Station, TX 77843-3112, USA
{hchoi, choe}@cs.tamu.edu
[‡]Department of Computer Science, Pohang University of Science and Technology
San 31, Hyoja-dong, Pohang, 790-784, Korea
seungjin@postech.ac.kr

Abstract

Most manifold learning methods consider only one similarity matrix to induce a low-dimensional manifold embedded in data space. In practice, however, we often use multiple sensors at a time so that each sensory information yields different similarity matrix derived from the same objects. In such a case, manifold integration is a desirable task, combining these similarity matrices into a compromise matrix that faithfully reflects multiple sensory information. A small number of methods exists for manifold integration, including a method based on reproducing kernel Krein space (RKKS) or DISTATIS, where the former is restricted to the case of only two manifolds and the latter considers a linear combination of normalized similarity matrices as a compromise matrix. In this paper we present a new manifold integration method, Markov random walk on multiple manifolds (RAMS), which integrates transition probabilities defined on each manifold to compute a compromise matrix. Numerical experiments confirm that RAMS finds more informative manifolds with a desirable projection property.

Introduction

Manifold learning involves inducing a smooth nonlinear low-dimensional manifold from a set of data points drawn from the manifold. Various methods (for example see Seung and Lee; Tenenbaum, de Silva, and Langford; Saul and Roweis 2000; 2000; 2003) have been developed and their wide applications have drawn attention in pattern recognition and signal processing. Isomap is a representative isometric mapping method. It extends metric multidimensional scaling (MDS), considering Dijkstra's geodesic distance (shortest path) on a weighted graph, instead of Euclidean distance (Tenenbaum, de Silva, and Langford 2000). Isomap can be interpreted as a kernel method (Ham et al. 2004; Choi and Choi 2007), where one dissimilarity matrix corresponds to one kernel matrix, representing one manifold of the data set.

In practice, however, we as humans use multiple sensors at a time, and each sensor generates data set on one manifold, then our brains combine multiple manifolds to identify objects. For example, when we analyze a speech source, we use both ears at the same time and merge the difference in information from the two ears. As another example, suppose there are colored discs in a room as in Fig. 1, and N persons enter the room separately and each measures the dissimilarity of the discs in their own ways. One person might measure the dissimilarity mainly based on the disc size, while another might take the measure based on color only. Based on these measures, we want to reconstruct the true locations of the discs from the N persons' different opinions, described by the dissimilarities. These examples give us more than one dissimilarity matrix and require us to find the one dissimilarity matrix, which is interpreted as one manifold.

So far, there have been few approaches that merge multiple dissimilarity matrices. A simple method to make one dissimilarity matrix is to combine them through substraction or division (Laub and Müller 2004). Alternatively, we can use the reproducing kernel Krein space (RKKS) instead of reproducing kernel Hilbert space (RKHS) as in Ong et al. (2004). RKKS uses negative eigenvalues in the kernel matrix, while conventional kernel methods use positive semi-definite kernel matrix. Recently, Abdi et al. (2005) suggested an algorithm, DISTATIS, based on linear sum of manifolds through principal component analysis (PCA).

In this paper, we suggest a new algorithm, *random walk on multiple manifolds* (RAMS), utilizing the fact that the distance in Euclidean space can be translated into transition probability. As in previous examples, it is natural to assume that two dissimilarities are measured independently. From this, we can calculate one transition probability from multiple transition probabilities, and then obtain a statistical distance from the final transition probability. This statistical distance is a nonlinear sum of the multiple distances, contrary to DISTATIS. RAMS can be applied to the case of more than two manifolds, contrary to RKKS which uses the sign of the eigenvalues. Moreover, since RAMS uses kernel Isomap after getting one dissimilarity matrix, RAMS inherits the desirable projection property from kernel Isomap (Choi and Choi 2007).

Related Work

Although it was not suggested in Laub and Müller (2004) and Ong et al. (2004) that manifolds be merged, we can utilize their approaches in our problem. To our knowledge,

Abdi et al. (2005) is the first to try to merge manifolds to find one manifold (*compromised manifold*), where they introduced DISTATIS. These methods work well in certain cases, but they still have some problems in nonlinear cases as shown in "Experiments". In this section, we look into Ong et al. (2004) and Abdi et al. (2005) in more detail (Laub and Müller 2004 is more-or-less straight forward).

Using reproducing kernel Krein space (RKKS)

Most kernel methods use positive semi-definite kernel matrix to guarantee that the dissimilarity matrix is the Euclidean representation in the embedded manifold. In those methods, any negative eigenvalue for the kernel matrix means error or noise. However, as in Isomap, usually the kernel matrix from dissimilarity of points does not satisfy the semi-positiveness. Moreover, as stated in Laub and Müller (2004), negative eigenvalues might indicate important features.

Ong et al. (2004) showed that non-positive kernels are meaningful just as positive kernels and suggested using RKKS to generalize reproducing kernel Hilbert space (RKHS) for both positive semi-definite kernel and negative kernel. With two RKHSs, $\mathcal{H}_+$ and $\mathcal{H}_-$, if a Krein space $\mathcal{K}$ is on RKKS with $\mathcal{K} = \mathcal{H}_+ \ominus \mathcal{H}_-$ defined by $\langle f, g\rangle_{\mathcal{K}} = \langle f_+, g_+\rangle_{\mathcal{H}_+} - \langle f_-, g_-\rangle_{\mathcal{H}_-}$, where $\langle,\rangle_{\mathcal{K}}$, $\langle,\rangle_{\mathcal{H}_+}$ and $\langle,\rangle_{\mathcal{H}_-}$ represent inner products in $\mathcal{K}$, $\mathcal{H}_+$ and $\mathcal{H}_-$, respectively, and $f \in \mathcal{K}$, $f_+ \in \mathcal{H}_+$, $f_- \in \mathcal{H}_-$ as do g, g_+ and g_-, then there exists two positive kernels k_+ and k_- such that

$$k = k_+ - k_-. \tag{1}$$

From this, if we have two manifolds that are orthogonal, then we can merge them using Eq. (1). From two dissimilarity matrices, we can calculate two kernel matrices and then apply Eq. (1). However, this assumption about orthogonality between two manifolds is not satisfied in the general case, so when we apply this approach, there is some distortion in the compromised manifold. Moreover, this approach is only for two dissimilarity measures. To apply this method to more than two dissimilarity measures, we must apply it iteratively. However, it is unnatural and the order of merger should be determined properly. See "Experiments" for results comparing the orthogonal and the nonorthogonal cases.

DISTATIS

Abdi et al. (2005) calculated kernel matrices $\boldsymbol{S}^{(k)}, k = 1, \cdots, C$ for each dissimilarity matrix $\boldsymbol{D}^{(k)}$, as in kernel Isomap (Choi and Choi 2007).

$$\widetilde{\boldsymbol{S}}^{(k)} = -\frac{1}{2}\boldsymbol{H}\boldsymbol{D}^{(k)2}\boldsymbol{H}, \tag{2}$$

$$\boldsymbol{S}^{(k)} = \lambda_1^{-1}\widetilde{\boldsymbol{S}}^{(k)}, \tag{3}$$

where λ_1 is the first eigenvalue of $\widetilde{\boldsymbol{S}}^{(k)}$, $\boldsymbol{D}^{(k)2} = [D_{ij}^{(k)2}]$ means the element-wise square of the distance matrix $\boldsymbol{D}^{(k)} = [D_{ij}^{(k)}]$, and $\boldsymbol{H}$ is the centering matrix, given by $\boldsymbol{H} = \boldsymbol{I} - \frac{1}{N}\boldsymbol{e}_N\boldsymbol{e}_N^\top$ for $\boldsymbol{e}_N = [1 \ldots 1]^\top \in \mathbb{R}^N$. Then, each kernel matrix is converted into one vector $\boldsymbol{s}_i$, to produce a matrix $\boldsymbol{X} = [\boldsymbol{s}_1, \boldsymbol{s}_2, \cdots, \boldsymbol{s}_T]$, which corresponds to a matrix of manifolds using our terminology. The principal components are then calculated using the inner product of $\boldsymbol{X}$ as in MDS or Isomap. The first eigenvector corresponding to the largest eigenvalue is the compromised matrix, which serves as the target manifold to project the data set. Actually, this final compromised manifold is expressed using the kernel matrix, $\boldsymbol{S}_+$.

$$\boldsymbol{S}_+ = \sum_{k=1}^{C} \alpha^{(k)}\boldsymbol{S}^{(k)}, \tag{4}$$

where α is the first eigenvector of $\boldsymbol{N}^{-\frac{1}{2}}\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{N}^{-\frac{1}{2}}$, and $\boldsymbol{N}$ is the diagonal matrix with diagonal terms of $\boldsymbol{X}^T\boldsymbol{X}$.

The remaining part is the same as MDS or Isomap. Projection to this compromised manifold $\boldsymbol{Y}$ is executed by Eq. (5) after eigen-decomposition, $\boldsymbol{S}_+ = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^T$,

$$\boldsymbol{Y} = \boldsymbol{V}\boldsymbol{\Lambda}^{\frac{1}{2}}, \tag{5}$$

where columns of $\boldsymbol{V}$ and diagonal elements of $\boldsymbol{\Lambda}$ are eigenvectors and eigenvalues of $\boldsymbol{S}_+$, respectively.

Here, Abdi et al. tried to find the best space to project the data set. However, because they used just the first linear principal component of $\boldsymbol{X}$, they lose some information residing in the other components. Especially, when the principal component is nonlinear, it will lose more information. We experiment with the same data set as in Abdi et al. (2005), and discuss the differences with our approach in "Discussion".

The Algorithm: RAMS

To overcome the limits stated above, we need a more general method for nonlinear cases or for more than two manifolds. We introduce a new approach, *random walk on multiple manifolds* (RAMS), to find one manifold from multiple measurements. The connection between random walk and manifold has been mentioned in many papers (Szummer and Jaakkola 2002; Ham et al. 2004; Kondor and Lafferty 2002). Let $\boldsymbol{G}$ be a weighted graph with N nodes $\boldsymbol{V}$ and edges $\boldsymbol{E}$, to represent a manifold. Then, the distance between two nodes on the kth manifold, $D_{ij}^{(k)}$, can be transformed into probability $P_{ij}^{(k)}$ which is the transition probability from the ith node to the jth node on the kth manifold, which can be given by

$$P_{ij}^{(k)} = \frac{1}{Z_i^{(k)}} e^{-\frac{D_{ij}^{(k)2}}{\sigma^{(k)2}}}, \tag{6}$$

where $Z_i^{(k)}$ is a normalization term so that the sum of transition probabilities from the ith node to all its neighbors on the kth manifold is 1, and $\sigma^{(k)}$ is a parameter representing the variance.

Given C dissimilarity matrices, $\boldsymbol{D}^{(1)}, \boldsymbol{D}^{(2)}, \cdots, \boldsymbol{D}^{(C)}$, we can get C probability matrices, $\boldsymbol{P}^{(1)}, \boldsymbol{P}^{(2)}, \cdots, \boldsymbol{P}^{(C)}$. We assume that these dissimilarities are measured independently. Note that we are not assuming that these dissimilarities make up orthogonal manifolds. Based on these assumptions, the compromised probability matrix $\boldsymbol{P}^*$ is calculated

as

$$P_{ij}^{*} = \frac{1}{Z_i^{*}} \prod_{k=1}^{C} P_{ij}^{(k)}, \tag{7}$$

where Z_i^{*} is a normalization term given by $Z_i^{*} = \sum_j P_{ij}^{*}$. Eq. (7) represents the probability for transition from the ith node to the jth node on the target (or compromised) manifold. From $\boldsymbol{P}^{*}$, we reconstruct the compromised dissimilarity $\boldsymbol{D}^{*}$ again.

$$D_{ij}^{*} = \sigma^{*} \sqrt{-\log(P_{ij}^{*})}, \tag{8}$$

which is the *statistical distance* from all the individual manifolds. Here, we have $C + 1$ parameters, $\sigma^{(1)}, \sigma^{(2)}, \cdots, \sigma^{(C)}, \sigma^{*}$, and we obtain these parameters using the following equations:

$$\sigma^{(k)} = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} D_{ij}^{(k)}, \tag{9}$$

$$\sigma^{*} = \frac{1}{C} \sum_{k=1}^{C} \sigma^{(k)}. \tag{10}$$

After getting $\boldsymbol{D}^{*}$, the rest is the same as kernel Isomap (Choi and Choi 2007). We calculate the kernel matrix from the compromised dissimilarity matrix.

$$\boldsymbol{K} = -\frac{1}{2} \boldsymbol{H} \boldsymbol{D}^{*2} \boldsymbol{H}. \tag{11}$$

As in kernel Isomap, we make the kernel matrix positive definite by adding a constant, c.

$$\widetilde{\boldsymbol{K}} = \boldsymbol{K}(\boldsymbol{D}^{*2}) + 2c\boldsymbol{K}(\boldsymbol{D}^{*}) + \frac{1}{2} c^2 \boldsymbol{H}, \tag{12}$$

where c is the largest eigenvalue of the matrix

$$\begin{bmatrix} \boldsymbol{0} & 2\boldsymbol{K}(\boldsymbol{D}^{*2}) \\ -\boldsymbol{I} & -4\boldsymbol{K}(\boldsymbol{D}^{*}) \end{bmatrix}. \tag{13}$$

Eq. (12) implies substituting $\widetilde{\boldsymbol{D}}^{*}$ for $\boldsymbol{D}^{*}$ in Eq. (11), which is given by

$$\widetilde{D}_{ij}^{*} = D_{ij}^{*} + c(1 - \delta_{ij}), \tag{14}$$

which makes the matrix $\boldsymbol{K}$ to be positive semi-definite. The term δ_{ij} is the Kronecker delta. Finally, projection mapping $\boldsymbol{Y}$ is given by Eq. (15) after eigen-decomposition, $\widetilde{\boldsymbol{K}} = \boldsymbol{V} \boldsymbol{\Lambda} \boldsymbol{V}^{T}$.

$$\boldsymbol{Y} = \boldsymbol{V} \boldsymbol{\Lambda}^{\frac{1}{2}}. \tag{15}$$

Relation to Previous Work

In RKKS, in the case of two distance matrices, the compromised distance matrix $\boldsymbol{D}^{*}$ is obtained implicitly from Eq. (1) and is given by

$$D_{ij}^{*} = \sqrt{D_{ij}^{(1)2} - D_{ij}^{(2)2}}. \tag{16}$$

It is not easy to compare Eq. (16) directly to DISTATIS or RAMS because it uses negative eigenvalues even though $\boldsymbol{D}^{*2}$ is a linear sum of $\boldsymbol{D}^{(1)2}$ and $\boldsymbol{D}^{(2)2}$ as in DISTATIS.

From the RKKS theory, however, if $D^{(1)}$ is orthogonal to $D^{(2)}$, which means two distances are uncorrelated, RKKS works, otherwise it has some distortion in the compromised manifold. Generally, two measurements will be correlated and this approach has some distortion on the final manifold. On the other hand, RAMS does not assume that two measurements are orthogonal. One assumption in RAMS is that two distances are measured independently, which is generally true. So, even with correlated distance matrices, RAMS works well.

In DISTATIS, the compromised distance matrix, $\boldsymbol{D}^{*}$, is obtained implicitly from Eq. (4) and given by

$$D_{ij}^{*} = \sqrt{\sum_{k=1}^{C} \alpha^{(k)} \frac{D_{ij}^{(k)2}}{\lambda_1^{(k)}}}. \tag{17}$$

Eq. (17) shows that $\boldsymbol{D}^{*2}$ is a linear sum of each squared distance matrix as in the RKKS method, where $\alpha^{(k)}$ gives the weight of the normalized distance matrix. In other words, it represents the importance of the individual manifold.

On the other hand, in RAMS, before adding the constant c in Eq. (14), the compromised distance matrix $\boldsymbol{D}^{*}$ is obtained explicitly as in Eq. (8) and is given by

$$D_{ij}^{*} = \frac{1}{C} \sum_{t=1}^{C} \sigma^{(t)} \sqrt{\log Z_i^{*} + \sum_{k=1}^{C} \log Z_i^{(k)} + \frac{D_{ij}^{(k)2}}{\sigma^{(k)2}}}. \tag{18}$$

In Eq. (18), the distance is calculated in a statistical way, where some variables such as $\log Z_i^{*} + \sum_{k=1}^{C} \log Z_i^{(k)}$ reflect how much related the ith point is to others on the compromised manifold. To compare with Eq. (17), if the normalization terms are 1, which means data points are already distributed in a well normalized form, then $\boldsymbol{D}^{*2}$ is a linear sum of each squared distance matrix as in Eq. (17).

RAMS is generally better than DISTATIS, especially when the compromised manifold is nonlinear. For example, let A, B, and C be three points with two distance matrices and the distance between A and C is longer than the distance between B and C in both distance matrices. However, if A and B on each individual manifold are located on a perpendicular line from the compromised manifold, then the distance between A and C will be the same as that between B and C on the compromised manifold represented by $\boldsymbol{D}^{*}$. But in RAMS, if one distance is longer than the other in both distance matrices, the former will be definitely longer than the latter on the compromised manifold. The reason for this difference is that DISTATIS uses a linearly compromised manifold that discards the other components except for the first principal component, whereas RAMS uses statistical distance as in Eq. (8).

Usually, the manifold made by a distance matrix is curved in high dimensional space and then the first components of the manifolds is not enough to contain the proper information. This is the reason why DISTATIS has some problems

with nonlinear manifolds. However, RAMS does not depend on the linear structure of distance matrices but depends on the statistical structure. This makes RAMS robust even with nonlinear manifolds.

Projection Property

Since RAMS uses kernel Isomap after obtaining the compromised distance matrix, RAMS inherits the projection property of kernel Isomap (Choi and Choi 2007). The generalization property (or projection property) of the kernel Isomap involves the embedding of test data points in the associated low-dimensional space. In other words, generalization property means the ability to determine a point $\boldsymbol{y}_l$ embedded in the low-dimensional space, given a test data point $\boldsymbol{x}_l$. The generalization property naturally emerges from the fact that $\widetilde{\boldsymbol{K}}$ (geodesic kernel with the constant-shifting employed in kernel Isomap) is a Mercer kernel. Due to limited space, we do not derive the equations for projection here. The derivations are similar to those in the previous section and in Choi and Choi (2007).

Experiments

In order to show the useful behavior of our method, we carried out experiments with three different data sets: (a) disc data set made of 100 discs with different colors and sizes; (2) head-related impulse response (HRIR) data (Algazi et al. 2001); and (3) face data (Abdi et al. 2005).

Discs

We made an artificial data set to show the differences between the three methods: (1) RKKS, (2) DISTATIS, and (3) RAMS. Fig. 1 shows 100 discs, where the horizontal axis represents color and the vertical axis represents size. Actually, 100 points were generated on a 10×10 lattice with added random noise in location.

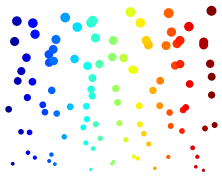


Figure 1: Disc data set

Let $\boldsymbol{X} \in \mathbb{R}^{2\times 100}$ be the discs' locations. The first row $\boldsymbol{X}_1$ and the second row $\boldsymbol{X}_2$ are the coordinates for color and size, respectively. From this disc data set, we made 3 pairs of measurements: (1) Orthogonal and linear case, (2) correlated but still linear case and (3) orthogonal but nonlinear case. Each case is calculated by the following equations.

(1) Orthogonal and linear: Each distance matrix is obtained by only color or size, respectively,

$$\boldsymbol{D}_{ij}^{(k)} \quad = \quad dist(X_{k,i}, X_{k,j}),$$

where $k = 1, 2$, and $dist(x, y)$ is the Euclidean distance function between two points x and y.

(2) Nonorthogonal and linear: one distance matrix is based on only color, and the other is based on both color and size;

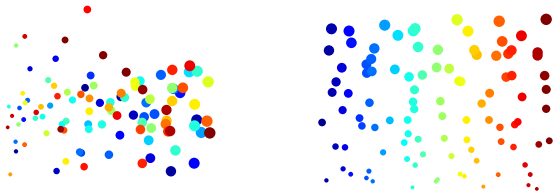
$$\boldsymbol{D}_{ij}^{(k)} \quad = \quad dist(X_{1:k,i}, X_{1:k,j}).$$

(3) Orthogonal and nonlinear: Each distance matrix is obtained by only color or size, respectively, and squared.

$$\boldsymbol{D}_{ij}^{(k)} \quad = \quad dist(X_{k,i}, X_{k,j})^2,$$

Orthogonal and linear case All three methods found 2 dominant eigenvalues. RKKS uses positive and negative values. Due to limited space, no figure is presented, but all three methods had no problem in finding the relative locations of the discs. In this case, two distance matrices are uncorrelated and each distance is obtained linearly. So, RKKS and DISTATIS also work well as well as our RAMS.

Nonorthogonal or nonlinear cases When the distances are measured nonorthogonaly or nonlinearly, RKKS and DISTATIS failed to find true locations, even though they are able to estimate along one coordinate, either size or color, but not both. Fig. 2 is the result of correlated but linear case. As we expected, RKKS found just one coordinate (size) and failed to find the color axis. On the other hand, RAMS still found proper coordinates as DISTATIS also did.



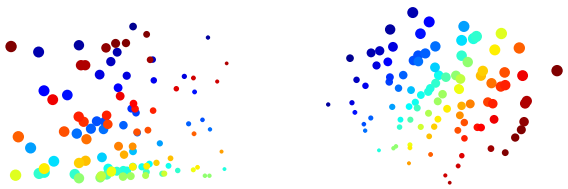
(a) (b)

Figure 2: Projections with (a) RKKS and (b) RAMS, in the case of nonorthogonal and linear distances. RKKS finds the size axis but fails to find the color axis, whereas RAMS finds both successfully.

Fig. 3 is the result of nonlinear but orthogonal case. In contrast to the linear and nonorthogonal case, DISTATIS found the size axis but failed to find the color axis, whereas RAMS found proper coordinates as RKKS also did.

Head-Related Impulse Responses (HRIR)

In this experiment, we used the public-domain CIPIC HRIR data set. Detailed description of the data can be found in Algazi et al. (2001). It was recently shown in Duraiswami and Raykar (2005) that the low-dimensional manifold of HRIRs could encode perceptual information related to the direction of sound source. However, there has been no attempt to use both the left and the right HRIRs at the same time. We applied the kernel Isomap to both left and right, and then tried to merge the two dissimilarity matrices into one with the three methods (RKKS, DISTATIS, and RAMS).



(a) (b)

Figure 3: Projections with (a) DISTATIS and (b) RAMS, in the case of orthogonal and nonlinear distances. DISTATIS finds the size axis but fails to find the color axis, whereas RAMS finds both successfully.

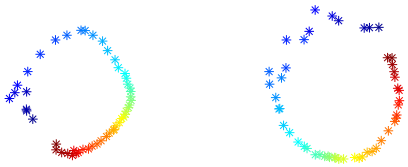
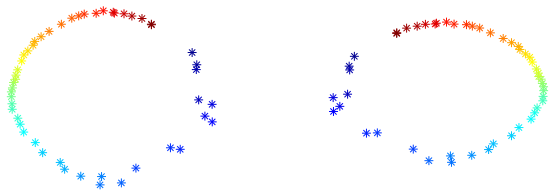


Figure 4: Two-dimensional manifolds of HRIRs: (left) Left HRIR; (right) Right HRIR. Each ear has similar information with a little distortion. This is the result from the 10th subject.

Two-dimensional manifolds of HRIRs are shown in Fig. 4 for kernel Isomap, where each ear has similar information, with a little distortion. RAMS and DISTATIS found two dominant eigenvalues of HRIRs because the manifold of each ear's HRIR is almost linear (see Choi and Choi 2007), whereas RKKS failed because these two pieces of information from the two ears are strongly correlated. Fig. 5 shows the embedded manifolds of the two methods and confirms the result. RKKS did not work here, so the results are not shown.



(a) (b)

Figure 5: Projected results on compromised 2-dimensional manifold from (a) DISTATIS and (b) RAMS. Both work well.

Faces

We used the face data set from Abdi et al. (2005), to compare RAMS with DISTATIS. Abdi et al. (2005) made 4 distance matrices and found a 2-dimensional space to represent 6 face images.

Table 1: Efficiency of eigenvalues.

Methods	First Eigenvalue	Second Eigenvalue
DISTATIS	47.75%	20.74%
RAMS	63.67%	19.04%

Table 1 compares the 2 top eigenvalues from the two methods, DISTATIS and RAMS, showing that RAMS is more efficient. Fig. 6 shows the 2-dimensional spaces from the two methods. Even though they have similar results, RAMS produces a little more spread map, because RAMS uses all information whereas DISTATIS uses only the first component of the compromised manifold and discards the rest.

(a) DISTATIS

(b)RAMS

Figure 6: Projections of faces.

Projection Property

Since RAMS inherits the projection property of kernel Isomap, the result of this experiment is promising. We just checked with 3 cases of discs above. We used 100 training data points as in Fig. 1, and 100 test data points generated in the same way as the training points. In Fig. 7 (a), blue crosses are training data points and colored and different sized discs are test data points in the original space. We

projected the test data points into the compromised manifold of the training data based on the two distance matrices. Fig. 7 (b),(c) and (d) show the result of projection in RAMS in all three cases. As expected, RAMS successfully projected the test data points on the compromised manifold in all three cases.

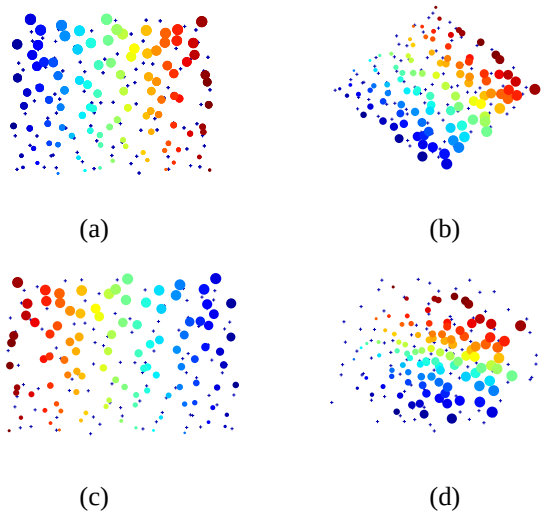


Figure 7: Training data set (blue crosses) and test data set (colored and differently sized discs) are shown. Test of the projection property of RAMS with (a) orthogonal and linear, (b) nonorthogonal and linear, (c) orthogonal and nonlinear, and (d) nonorthogonal and nonlinear cases show good mapping along both the color and the size axes.

Discussion

In RKKS, even though two measurements are obtained independently, if their values are affected by each other, then the final manifold is distorted because RKKS assumes that the two Hilbert spaces are orthogonal. However, in RAMS, even though the measurements are related to each other, it gives a proper mapping because the two dissimilarity matrices are measured in a statistically independent way. Another interesting issue is that in RKKS, it is not straightforward to extend the number of measurements to more than 2, while RAMS can be naturally extended.

DISTATIS uses the principal component of the manifolds to project data sets into the compromised manifold. That results in loss of information that could be found in the other components. When the distances are measured nonlinearly, such as in squareness, DISTATIS fails to find a proper mapping, because nonlinear manifold does not guarantee that one linear component of manifolds can contain most of the information. However, in RAMS, the final manifold is not a linear combination but rather a probabilistic combination, so it is robust even with nonlinear manifolds.

More interestingly, RKKS and DISTATIS are not concerned with the projection property, while RAMS is. So, RAMS can be considered as the generalization of RKKS and DISTATIS.

Conclusion

Many algorithms have been developed in the past to find a low dimensional manifold from data sets. However, techniques for finding one manifold from multiple measurements have not been fully explored. In this paper, we proposed a new algorithm, *RAMS*, to address this gap and compared it with other existing methods: RKKS and DISTATIS. Those previous algorithms have some limitations or need to be generalized. Our algorithm, RAMS, is a generalization of previous methods and also possesses a desirable projection property as in kernel Isomap. Experimental results confirmed the performance and the desirable projection property. Also, RAMS can be applied to finding one feature set from several similar feature sets. An interesting direction for future work would be to compare with other data fusion theories and extend this algorithm.

Acknowledgments

This work was supported by KOSEF Grant-D00115 and by Korea Ministry of Knowledge Economy under the ITRC support program supervised by the IITA (IITA-2008-C1090-0801-0045)

References

Abdi, H.; Valentin, D.; O'Toole, A. J.; and Edelman, B. 2005. DISTATIS: The analysis of multiple distance matrices. In *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition*, 42–47.

Algazi, V. R.; Duda, R. O.; Thompson, D. M.; and Avendano, C. 2001. The CIPIC HRTF database. In *Proc. 2001 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, 99–102.

Choi, H., and Choi, S. 2007. Robust Kernel Isomap. *Pattern Recognition* 40(3):853–862.

Duraiswami, R., and Raykar, V. C. 2005. The manifolds of spatial hearing. In *Proc. IEEE Int'l Conf. Acoustics, Speech, and Signal Processing*, 285–288.

Ham, J.; Lee, D. D.; Mika, S.; and Schölkopf, B. 2004. A kernel view of the dimensionality reduction of manifolds. In *Proc. Int'l Conf. Machine Learning*, 369–376.

Kondor, R. I., and Lafferty, J. 2002. Diffusion kernels on graphs and other discrete structures. In *Proc. Int'l Conf. Machine Learning*.

Laub, J., and Müller, K. R. 2004. Feature discovery in non-metric pairwise data. *Journal of Machine Learning Research* 5:801–818.

Ong, C.; Mary, X.; Canu, S.; and Smola, A. 2004. Learning with non-positive kernels. In *Proc. Int'l Conf. Machine Learning*, 639–646.

Saul, L., and Roweis, S. T. 2003. Think globally, fit locally: Unsupervised learning of low dimensional manifolds. *Journal of Machine Learning Research* 4:119–155.

Seung, H. S., and Lee, D. D. 2000. The manifold ways of perception. *Science* 290:2268–2269.

Szummer, M., and Jaakkola, T. 2002. Partially labeled classification with markov random walks. In *Advances in Neural Information Processing Systems*, volume 14. MIT Press.

Tenenbaum, J. B.; de Silva, V.; and Langford, J. C. 2000. A global geometric framework for nonlinear dimensionality reduction. *Science* 290:2319–2323.