

Revealing Smooth Structure of Visual Data by Permutation on Manifolds

Yi-Lei Chen¹

<https://sites.google.com/site/fallcolor/>

Chiou-Ting Hsu¹

<http://www.cs.nthu.edu.tw/~cthsu/candy.html>

¹ Department of Computer Science, National Tsing Hua University, Taiwan

Abstract

Advance in technology and commercial media has simplified the process of collecting large-scale visual data, but it also raises new challenges in data organization. In this paper, we propose to characterize data association by recovering an intrinsic order from an unorganized dataset. Our method is motivated by smooth manifold geometry. We advocate that the optimal data order should encode the shape of underlying manifold as well as the latent data association. Following the data order, we find a smooth path to visualize the latent topic of visual data with a perceptually reasonable transition. We develop an efficient algorithm *Permutation on Manifolds* (PoM) to solve this NP-hard permutation problem. Experiments on synthetic and real-world dataset demonstrate the potential of PoM to serve as a core technique of numerous applications.

1 Introduction

Collecting visual data becomes incredibly easy nowadays. With various media tools and widespread internet, many people prefer to record their daily lives with photos or videos, and quickly share their own interpretation of the world. A new wave on communicating with the world is emerging: click on one picture and share other's life experience readily. For example, we may upload a travel photo into internet and soon find more photos of nearby landscape we are interested. However, the ever-increasing scale of ubiquitous visual data also raises new challenges, especially when we try to use the unorganized dataset for subsequent analysis. More specifically, if no semantic prior or domain knowledge is available, how can we automate the *invisible* data association from the *visible* data collection?

In this paper, we address the issue of visual data organization by recovering an intrinsic order from an unorganized dataset. We observe that, although the sources of data collection are unpredictable, a collection of visual data usually shares similar topics (e.g., landmarks, personal albums) and most of them are created for some special events (e.g., sports games, family gathering). Assuming that there exists a strong relation among the unorganized visual content, we recast our problem as finding a smooth path to visualize the latent order of the visual data with a smooth transition. This technique could be adopted to numerous applications, such as photo grouping, image browsing, album organization, and so on. Some pioneer work, which was called photo sequencing, has been proposed to discover a set of still images for estimating the temporal order of a dynamic scene [1, 2] or

the spatial order of an instantaneous event [3]. These methods adopt different hypotheses when capturing images, such as similar view points or ring-like camera arrangement. Unlike these methods, our goal is to provide a general framework for visual data ordering without imposing additional constraints on data collection.

The proposed method is built on the inherent nature of manifold geometry. In [4], cognitive experiments have suggested that human visual system represents objects in the form of continuous manifolds under geometric and photometric changes. Under the observation that strong relation exists among visual content, we assume a visual dataset lies on a manifold and thus changes smoothly from point to point. By exploiting the linearity within nearby data points, our goal then becomes to visit all of the data points along a manifold-guided order and to characterize the specific manifold’s shape. We believe that this order should encode the underlying manifold as well as the latent data association. To the best of our knowledge, this is the first work that leverages manifold geometry to permute unorganized data for data analysis. In Sec. 2, we will detail our algorithm, *permutation on manifolds* (PoM), and show that even with increasing problem size, PoM can solve the permutation of high-dimensional data efficiently, whereas most off-the-shelf solvers fail. With the optimal data permutation, we can then organize a given dataset in terms of visual coherency for numerous applications.

1.1 Related Work

Our target task is related to photo organization [5, 6, 7] and image navigation [8, 9]. Existing work of photo organization focus on analysing specific type of images (e.g., human faces [5], urban scenes [6]) and thus these methods can exploit structure constraints on facial appearance or planar façade to organize the data. However, these methods usually require pre-processing like 3D registration and segmentation to align the image space. Health et al. [7] proposed a more general framework to increase the connectivity between common image regions by building a spectral graph. This method focused more on effective graph construction than on explicit data organization. On the other hand, research of image navigation seeks to explore data paths for browsing and rendering. For example, Snavely et al. [8] proposed to navigate a 3D scene by discovering a range of camera viewpoints; and Arpa et al. [9] introduced an angled graph to guide the navigation of crowd images with the user-specified direction. Note that, all the above-mentioned methods consider finding a path from only a subset of images as data organization. This problem can be solved efficiently by the shortest path algorithm. By contrast, our task is much more challenging because we attempt to recover the intrinsic order of the entire dataset.

The most similar work to ours is [3], where the authors assume that the images captured over an event should reveal a ring-topology in spatial ordering as well as in low-dimensional distribution. Different from [3], we impose no constraint on the shape of manifold and only exploit locally-smooth geometry through pseudo graph. Accordingly, our method is more general to comprehensive sources of visual data. In addition to [3], another similar idea was adopted for image restoration. In [10], the authors partitioned an image into overlapped patches and then reordered these patches by estimating a smooth path to facilitate image denoising and inpainting. Although this method [10] shares similar viewpoints to ours, it only focused on the smoothness of the “path” instead of treating the geometry formed by image patches as a manifold. Because the smoothness was measured in terms of one nearest neighbour in the graph principle, the method [10] finds the shortest

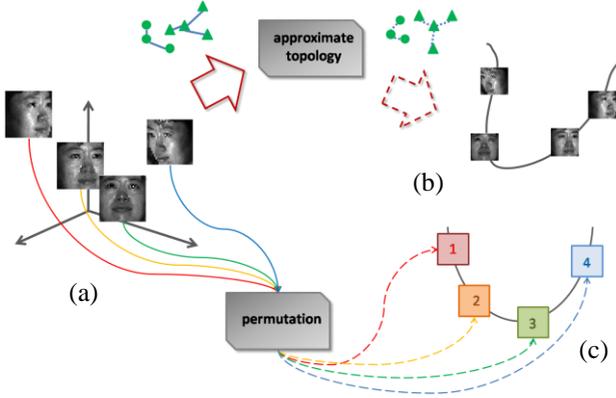


Fig. 1. Illustration of our proposed idea. (a) Data points in the original high-dimensional space; (b) the low-dimensional manifold approximated by classic methods; and (c) the smooth manifold approximated by permutation.

possible path by merely solving the traveling salesman problem (TSP). In Sec. 2.1, we will show that TSP is a special case of our problem formulation. Specifically, our method solves a special case of quadratic assignment problem (QAP) and this case is much more difficult than TSP in combinatorial optimization.

2. Method

2.1 Pseudo Graph

To characterize the manifold geometry, traditional methods first construct an explicit data affinity matrix under a graph representation. This data affinity matrix is then used to infer the intrinsic data dimension so that we can obtain a compact data distribution. Given an unorganized dataset, if the underlying manifold did exist, its low-dimensional distribution should form a specific shape. In other words, the manifold geometry should *constrain* the data implicitly even though the data organization is unknown. We are motivated by this reasoning and propose to explore the data organization from a manifold-guided perspective.

We use Figure 1 to clarify our idea. Because a manifold, though may be nonlinear, is highly probable to be smooth, exploring the shape of manifold is similar to visiting all the data points along a shape-guided order. To estimate the smooth ordering, we first *construct a template of smooth manifold*, and then *discover the optimal permutation between data points to fit this smooth shape*. We use a pseudo graph to build the template and define the corresponding Laplacian matrix $\mathbf{L} = \mathbf{D} - \mathbf{W}$ by

$$w_{ij} = \exp\left(-\frac{(i-j)^2}{\eta_g^2}\right) \text{ and } d_{ij} = \begin{cases} \sum_k w_{ik}, & \text{if } i = j \\ 0, & \text{if } i \neq j \end{cases}. \quad (1)$$

In Equation (1), the edge weight w_{ij} is simply measured according to the adjacency between indices to ensure the well-ordered smoothness. The parameter η_g roughly controls the size of effective neighbourhood. We fix $\eta_g = 2$ in all of our experiments and find that this simple setting already works well in general cases. Fig. 1(c) illustrates the template of the smooth manifold derived from the pseudo graph.

Note that our goal is essentially different from those of traditional methods. Rather than inferring the low-dimensional data distribution, we try to measure an explicit order indicator $i \mapsto \phi(i)$ (see Sec. 2.2 for detailed description) to reveal the data organization encoded by the manifold geometry. With the smooth template built on the pseudo graph, this indicator connects data smoothly and is believed to be the optimal permutation between data points. In addition, Equation (1) generalizes several related methods based on spectral graph theory (see Sec. 1.1). For example, when we fix the number of nearest neighbours to be one, the problem reduces to TSP [10]. If we further establish an edge link between starting and ending data points, the target manifold then embodies a ring topology, which has been introduced in [3] to characterize the spatial ordering of a temporal event.

2.2 Problem Formulation

Given N high-dimensional data points $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_i, \dots, \mathbf{x}_N]$ where \mathbf{x}_i denotes the feature descriptor of the i^{th} data point, our goal is to uncover the intrinsic data order $i \mapsto \phi(i)$ under a guidance of smooth manifold geometry. Specifically, we formulate the problem as

$$\min f(\mathbf{P}) = \min \sum_{1 \leq i, j \leq N} \|\mathbf{x}_{\phi(i)} - \mathbf{x}_{\phi(j)}\|_2^2 w_{\phi(i)\phi(j)} = \min \text{tr}(\mathbf{XPLP}^T \mathbf{X}^T), \quad (2)$$

where \mathbf{L} is the Laplacian matrix derived from a pseudo graph, $\text{tr}(\cdot)$ denotes the trace function, and $\mathbf{P} \in \Pi$ is a permutation matrix satisfying

$$\Pi = \{\mathbf{P} \mid p_{ij} \in \{0, 1\}, \sum_{j=1}^N p_{ij} = 1, \forall i, \sum_{i=1}^N p_{ij} = 1, \forall j\}. \quad (3)$$

Equation (2) is an NP-hard combinatorial optimization problem and remains very challenging without off-the-shelf solvers. Note that we can also rewrite the objective function $f(\mathbf{P})$ as a special case of quadratic assignment problem (QAP) [11]:

$$\min f(\mathbf{P}) = \min \text{tr}(\mathbf{XPLP}^T \mathbf{X}^T) = \min \text{tr}(\mathbf{PLP}^T \mathbf{X}^T \mathbf{X}) = \min \text{tr}(\mathbf{PAP}^T \mathbf{B}), \quad (4)$$

where \mathbf{A} and \mathbf{B} denote the weight matrix and the distance matrix defined in QAP, respectively. However, to the best of our knowledge, most QAP solvers fail to solve our problem for two reasons:

- When the problem size N increases (e.g., $N > 20$), exact QAP algorithms (e.g., branch-and-bound [12], cutting plane [13]) become computationally intractable.
- Few convex relaxation methods [14] are scalable to large N by relaxing the binary integer constraints into nonnegative constraints. Unfortunately, this relaxation strategy just leads to trivial solutions ($p_{ij} = 1/N, \forall i, j$) in our case because, in Equation (2), the nature of quadratic penalty encoded by \mathbf{L} encourages a smooth result.

By contrast, as we will detail in Sec. 2.3, the proposed algorithm can address the scalability issue and reach local optima of Equation (2) efficiently. As a byproduct, our method can even solve the symmetric cases of QAP provided that the weight matrix \mathbf{B} is decomposable (i.e., $\mathbf{B} = \mathbf{X}^T \mathbf{X}$).

2.3 Algorithm

We name our algorithm *Permutation on Manifolds* (PoM) in the following sections. To make Equation (2) tractable without reaching *wrong optima* (i.e., $p_{ij} = 1/N, \forall i, j$), we first use variable splitting technique and obtain

$$\min \text{tr}(\mathbf{X}_p \mathbf{L} \mathbf{X}_p^T) \quad \text{s.t. } \mathbf{X}_p = \mathbf{X} \mathbf{P}. \quad (5)$$

Equation (5) is a constrained optimization problem and can be solved efficiently by first-order method. Most importantly, the unknown \mathbf{P} is now splitting from the Laplacian matrix \mathbf{L} . Hence we circumvent the risk of relaxing binary integer constraints into nonnegative constraints [14]. In this work, we use augmented Lagrange multiplier (ALM) method to solve Equation (5). The convergence of ALM has been well-studied and recently shown its success in computer vision and machine learning community [15]. By introducing an augmented Lagrange function:

$$L(\mathbf{X}_p, \mathbf{P}, \mathbf{Y}, \mu) = \text{tr}(\mathbf{X}_p \mathbf{L} \mathbf{X}_p^T) + \langle \mathbf{Y}, \mathbf{X}_p - \mathbf{X} \mathbf{P} \rangle + \frac{\mu}{2} \|\mathbf{X}_p - \mathbf{X} \mathbf{P}\|_F^2, \quad (6)$$

we follow ALM to approximate \mathbf{P} iteratively by

$$\mathbf{X}_p^{(t+1)} = \arg \min_{\mathbf{X}_p} L(\mathbf{X}_p, \mathbf{P}^{(t)}, \mathbf{Y}^{(t)}, \mu^{(t)}), \quad (7)$$

$$\mathbf{P}^{(t+1)} = \arg \min_{\mathbf{P}} L(\mathbf{X}_p^{(t+1)}, \mathbf{P}, \mathbf{Y}^{(t)}, \mu^{(t)}), \quad (8)$$

$$\mathbf{Y}^{(t+1)} = \mathbf{Y}^{(t)} + \mu^{(t)}(\mathbf{X}_p - \mathbf{X} \mathbf{P}), \text{ and} \quad (9)$$

$$\mu^{(t+1)} = \rho \mu^{(t)}, \quad (10)$$

where t denotes the iteration index and ρ is a constraint-penalty parameter larger than one (we fix ρ as 1.1 in all of our experiments). Let \mathbf{I} be an $n \times n$ identity matrix, the analytic solutions of Equations (7) is readily derived by

$$\mathbf{X}_p^{(t+1)} = (\mu^{(t)} \mathbf{X} \mathbf{P}^{(t)} - \mathbf{Y}^{(t)}) (\mathbf{L} + \mathbf{L}^T + \mu^{(t)} \mathbf{I})^{-1}. \quad (11)$$

The core of our algorithm then boils down to

$$\mathbf{P}^{(t+1)} = \arg \min_{\mathbf{P}} \|\mathbf{X} \mathbf{P} - \mathbf{Z}\|_F^2 \quad \text{s.t. } \mathbf{P} \in \Pi, \quad (12)$$

where $\mathbf{Z} = \mathbf{X}_p^{(t+1)} + \mathbf{Y}^{(t)}/\mu^{(t)}$. Note that solving Equation (12) is not trivial. If we simply adopt convex relaxation with $0 \leq p_{ij} \leq 1$, we will obtain floating-point results on \mathbf{P} and still lose the problem precision after binary thresholding. The numerical solution should be carefully approximated to reach good local optima.

Because the matrix element of a desired \mathbf{P} is either zero or one, we rephrase Equation (12) by a sequence of additive costs and obtain

$$\mathbf{P}^{(t+1)} = \arg \min_{\mathbf{P}} \sum_{i=1}^N \sum_{j=1}^N \|\mathbf{x}_i - \mathbf{z}_j\|^2 p_{ij} \quad \text{s.t. } \mathbf{P} \in \Pi, \quad (13)$$

where \mathbf{z}_j is the j^{th} column of the augmented matrix \mathbf{Z} and p_{ij} denotes the $(i, j)^{\text{th}}$ element of \mathbf{P} . Now Equation (13) becomes a binary integer programming problem but remains intractable when N is very large. Fortunately, if we relax binary integer constraints into nonnegative constraints, the optimal solution of Equation (13) is guaranteed by linear programming solved with simplex method [16]. That is, in average cases, we probably solve Equation (13) in polynomial time; however, its worst case complexity is still exponential time. To further reduce the computational cost, we propose two guidelines to constrain our problem size:

- Each data point \mathbf{x}_i is only allowed a limited number of moves, and its moving destination is selected from its k nearest data points among $\mathbf{z}_1, \dots, \mathbf{z}_N$.
- The best number k , as small as possible, is determined by a binary search strategy so that every \mathbf{z}_i can be accessed by $\mathbf{x}_1, \dots, \mathbf{x}_N$ at least once.

The first criterion effectively limits the search range in simplex method, while the second criterion guarantees one-to-one mapping between \mathbf{X} and \mathbf{Z} . In practice, we observe that increasing k by a small constant makes our algorithm more stable and thus we use $k + 5$ as the neighbourhood size in all of our experiments.

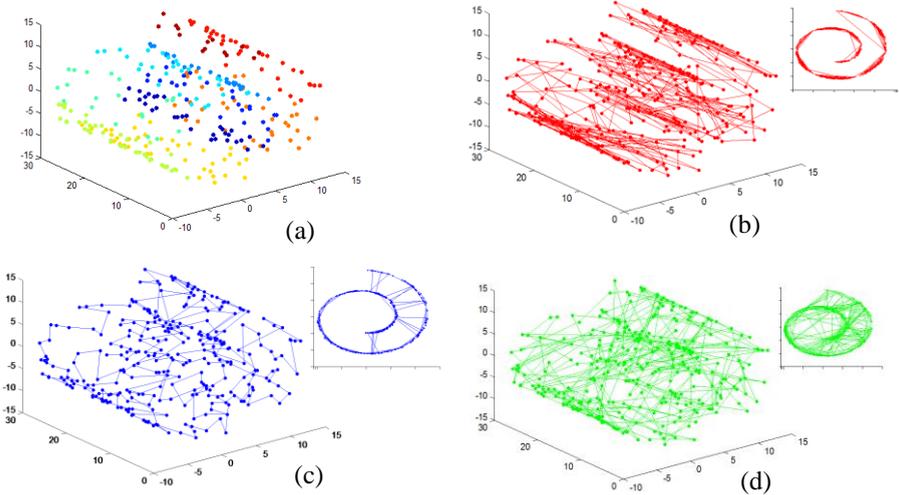


Fig. 2. (a) The 300 data points sampled from a Swiss roll; and the data-linking results obtained by (b) ground truth labels; (c) PoM; and (d) GNCGCP [19].

2.4 Extensions

In addition to data permutation, we study the potential of extending PoM Algorithm to different applications, especially for unsupervised tasks if the data order is essential to the underlying problem. For example, we can seamlessly embed PoM into linear dimension reduction methods when the data affinity matrix is unavailable. That is, we use the permuted pseudo-graph \mathbf{PLP}^T to approximate the affinity matrix under the graph-embedding framework [17]:

$$\hat{\mathbf{V}}, \hat{\mathbf{P}} = \arg \min \text{tr}(\mathbf{V}^T \mathbf{XPLP}^T \mathbf{X}^T \mathbf{V}) \quad \text{s.t. } \mathbf{P} \in \Pi. \quad (14)$$

The projection matrix \mathbf{V} and the permutation matrix \mathbf{P} are then iteratively refined until the objective cost is unchanged.

Equation (14) can be also modified and used as a regularization term in data completion problem. For example, a recent tensor completion method called STDC [18] was proposed to exploit within- or joint-manifold priors, under the assumption that these priors are known a priori and can be measured from the semantics of data. However, this assumption is sometimes impractical because we usually lack sufficient knowledge of incomplete data. To render STDC into a fully-unsupervised one, instead of defining the data affinity from data semantics, we can follow PoM and introduce the permutation matrix \mathbf{P} into manifold priors. Interested readers may refer to [18] for more technical details. We will provide preliminary simulation results in our experimental section.

3. Experiments

3.1 Validation on Synthetic Dataset

Because the Swiss roll has a smooth manifold shape and is a good candidate to justify our hypothesis, we sample 300 data points from the Swiss roll (see Fig. 2(a)) as the

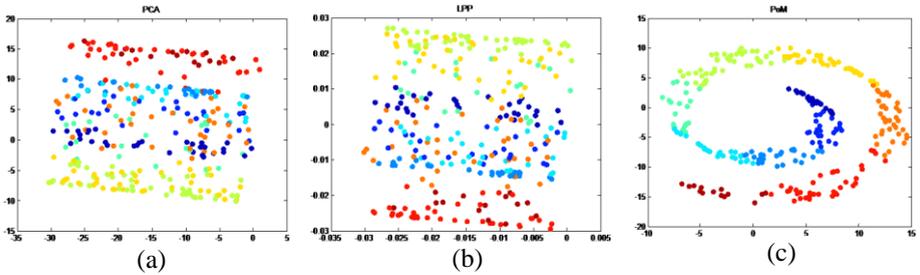


Fig. 3. The linear dimension reduction results of Fig. 2(a) obtained by (a) PCA [20]; (b) LPP [21]; and (c) PoM. Note the difference that PoM characterizes the shape of Swiss roll while both PCA and LPP fail.

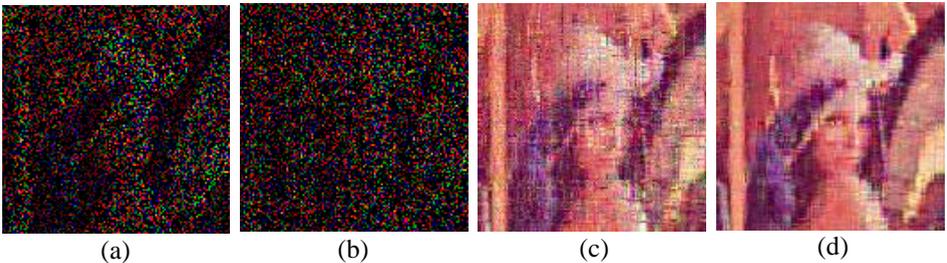


Fig. 4. The image “Lena” with 80% missing entries: (a) before and (b) after permutation (50% rows and columns are randomly reordered); and its image completion results obtained by (c) STDC (PSNR=17.61 dB) and (d) STDC-PoM (PSNR=20.95 dB).

synthetic dataset. To visualize the results, we apply the proposed PoM Algorithm to the dataset and link data points along the estimated order. In addition, we compare with a recently proposed QAP solver, called GNGCQP [19], by recasting Equation (2) into a QAP formulation, using the source code downloaded from the author’s website. Because GNGCQP is a first-order optimization method, it can handle large-scale cases if we tune its initialization (let $\mathbf{P} = \mathbf{I}$ rather than $p_{ij} = 1/N$) carefully.

Figure 2(b)-(d) demonstrates the ground truth and the data-linking results of GNGCQP and PoM. In comparison with the ground truth, where the order is determined by labels, the proposed PoM approximately recovers the order along the manifold’s shape. In contrast, GNGCQP shows poor linking and incurs high computational overhead (about 4.5 hours) for convergence, whereas our method takes less than five minutes.

3.2 Regularization as Manifold Priors

We first show that the proposed PoM can be used for linear dimension reduction. We solve Equation (14) to reduce the data dimension of Fig. 2(a) from three to two. Figure 3 demonstrates the two-dimensional subspaces in comparison with two classic algorithms PCA [20] and LPP [21], where the data affinity matrices are constructed from global distribution and local topology, respectively. Note that the focus here is not to unroll the Swiss roll (infeasible for linear methods) but to justify the significance of intrinsic order. Both PCA and LPP characterize the first two principal dimensions where a few data points overlap disorderly. By contrast, our method characterizes the first and the third principal dimension and thus preserve the manifold’s shape as well as the label coherency.

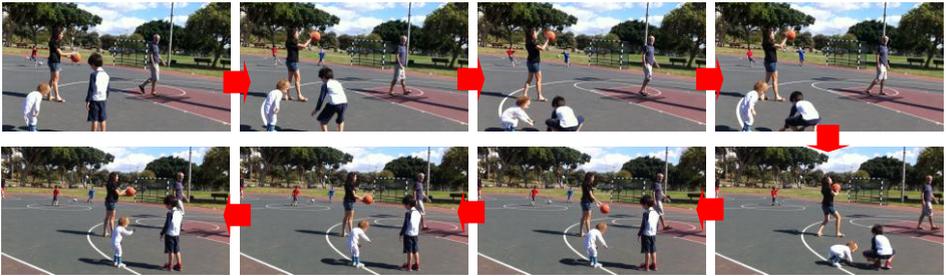


Fig. 5. The photo ordering results of the image set “Basketball”.



Fig. 6. The photo ordering results of the image set “Slide”.

Next, we incorporate the proposed PoM into STDC [18] framework using the permutation matrices as a set of augmented variables. This modification enables an unsupervised STDC even when the within- or joint-manifold priors are not known a priori. Fig. 4 demonstrates an application on image completion, where Fig. 4(a) and Fig. 4(b) show the incomplete images before and after randomly permuting rows and columns, respectively. The data missing rate is 80% and the permutation rate is 50%. We give the image completion results of the original method [18] and its PoM extension in Fig. 4(c)-(d). Note that the role of PoM here is to characterize the smooth manifold priors for improving data completion rather than to infer the original row and column arrangements absolutely. For better visualization, we rearrange the rows and columns of two results with correct order to show the improvements. Because the local smoothness of 2-D image no longer holds after permutation, STDC obtains a quite noisy result and loses fine-scale structure. By contrast, STDC-PoM generates a visually-pleasant image as well as higher PSNR.

3.3 Visual Data Organization

We use two image sets “Basketball” and “Slide” released from [1] to verify our method. The “Basketball” dataset contains eight images captured by a pair of mobile phones, while the “Slide” dataset consists of tens images captured by five different cameras. The images of each dataset are initially unorganized and exhibit viewpoint changes. Because our goal is to evaluate the feasibility of the PoM algorithm, we include no pre-processing step (e.g., 3D calibration) and just adopt a simple descriptor “image signature” (IS) [22] for feature representation. Other advanced features, for example, SIFT or SURF as suggested in [3], can be used if one considers a more challenging dataset.

Following [22], each image is downsampled to the size of 36x64 and encoded by IS descriptor. We then vectorize the feature maps of three color channels and concatenate the feature vectors into \mathbf{x} . Fig. 5 and Fig. 6 demonstrate the results of photo ordering measured by our algorithm. As shown in Fig. 5, after we permute images by PoM, one can observe a

smooth change of the position of children which somewhat reveals the spatial order of cameras. Likewise, in Fig. 6, the images with similar viewpoints are grouped together, although our method cannot determine which image is the best starting point. From the results, we can conclude that the proposed PoM algorithm is able to approximate a perceptually reasonable order from a given image set, provided that a proper feature representation is chosen. The estimated order organizes image data implicitly and enables the possibility of numerous applications, such as visualization, image browsing, photo grouping, and so on.

4. Conclusion

In this paper, we propose to organize visual data with a perceptually reasonable order, which corresponds to the shape of manifold where the data lie on. This new perspective, posing no hypothesis on local topology of observed data, is simply built on the smoothness prior of manifold geometry. By constructing a template of smooth manifold, we recast our problem into binary integer programming and propose a scalable algorithm to solve it efficiently. We summarize our contributions from three aspects: 1) to the best of our knowledge, this is the first work that leverages smooth manifold geometry to organize visual data; 2) the proposed pseudo graph generalizes the use of spectral graph theory and bypasses the measurement of data affinity in traditional methods; and 3) from our experiments, the proposed PoM algorithm shows its potential to serve as a core technique of numerous applications.

References

- [1] T. Basha, Y. Moses, and S. Avidan. Photo Sequencing. In *Proc. ECCV*, 2012.
- [2] T. Basha, Y. Moses, and S. Avidan. Space-Time Tradeoffs in Photo Sequencing. In *Proc. ICCV*, 2013.
- [3] H. Averbuch-Elor and D. Cohen-Or. RingIt: Ring-ordering Casual Photos of a Temporal Event. In *Proc. SIGGRAPH*, 2015.
- [4] J. DiCarlo and D. Cox. Untangling Invariant Object Recognition. *Trends in Cognitive Sciences*, 11(8): 333-341, 2007.
- [5] I. Kemelmacher-Shlizerman, E. Shechtman, R. Garg, and S. M. Seitz. Exploring Photobios. *ACM Trans. on Graphics*, 30(4), 2011.
- [6] G. Wan, N. Snavely, D. Cohen-Or, Q. Zheng, B. Chen, and S. Li. Sorting Unorganized Photo Sets for Urban Reconstruction. *Graphical Models*, 74: 14-28, 2012.
- [7] K. Heath, N. Gelfand, M. Ovsjanikov, M. Anajaneva, L. and J. Guibas. Image Webs: Computing and Exploiting Connectivity in Image Collections. In *Proc. CVPR*, 2010.
- [8] N. Snavely, R. Garg, S. M. Seitz, and R. Szeliski. Finding Paths through the World's Photos. *ACM Trans. on Graphics*, 27(3), 2008.
- [9] A. Arpa, L. Ballan, R. Sukthankar, G. Taubin, M. Pollefeys, and R. Raskar. CrowCam: Instantaneous Navigation of Crowd Images using Angled Graph. In *3DTV*, 2013.
- [10] I. Ram, M. Elad, and I. Cohen. Image Processing Using Smooth Ordering of Its Patches. *IEEE Trans. Image Processing*, 22(7): 2764-2774, 2013.

-
- [11] R. E. Burkard, E. Cela, P. M. Pardalos, and L. S. Pitsoulis. The Quadratic Assignment Problem. SFB-Report 126, Inst. Mathematik B, Tech. Univ. Graz, Austria, 1998.
- [12] P. M. Pardalos, K. G. Ramakrishnan, M. G. C. Resende, Y. Li. Implementation of a Variable Reduction Based Lower Bound in a Branch and Bound Algorithm for the Quadratic Assignment Problem. *SIAM J. Optimization*, 7: 280-294, 1997.
- [13] M. S. Bazaraa and H. D. Sherali. On the Use of Exact and Heuristic Cutting Plane Methods for the Quadratic Assignment Problem. *J. Operation Research Society*, 33: 991-1003, 1982.
- [14] C. Schellewald, S. Roth, and C. Schnorr. Evaluation of a Convex Relaxation to a Quadratic Assignment Matching Approach for Relational Object Views. *J. Image and Vision Computing*, 25: 1301-1314, 2007.
- [15] D. Bertsekas. *Constrained Optimization and Lagrange Multiplier Method*. Academic Press, 1982.
- [16] G. Nemhauser and L. Wolsey. *Integer and Combinatorial Optimization*. Wiley, 1988.
- [17] S. Yan, D. Xu, B. Zhang, H. Zhang, Q. Yang, and S. Lin. Graph Embedding and Extensions: A General Framework for Dimensionality Reduction. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 29(1): 40-51, 2007.
- [18] Y. L. Chen, C. T. Hsu, and H. Y. Mark Laio. Simultaneous Tensor Decomposition and Completion Using Factor Priors. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 36(3): 577-591, 2014.
- [19] Z. Y. Liu and H. Qiao. Graduated NonConvexity and Concavity Procedure for Partial Graph Matching. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 36(6): 1258-1267, 2014.
- [20] M. Turk and A. Pentland. Face Recognition Using Eigenfaces. In *Proc. CVPR*, 1991.
- [21] X. He, S. Yan, Y. Hu, P. Niyogi, and H. Zhang. Face Recognition Using Laplacianfaces. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 27(3): 328-340, 2005.
- [22] X. Hou, J. Harel, and C. Koch. Image Signature: Highlighting Sparse Salient Regions. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 34(1): 194-201, 2012.