Permutation Diffusion Maps (PDM) with Application to the Image Association Problem in Computer Vision

Deepti Pachauri, Risi Kondor, Gautam Sargur, Vikas Singh

http://pages.cs.wisc.edu/~pachauri/pdm/

The University of Wisconsin Madison
The University of Chicago

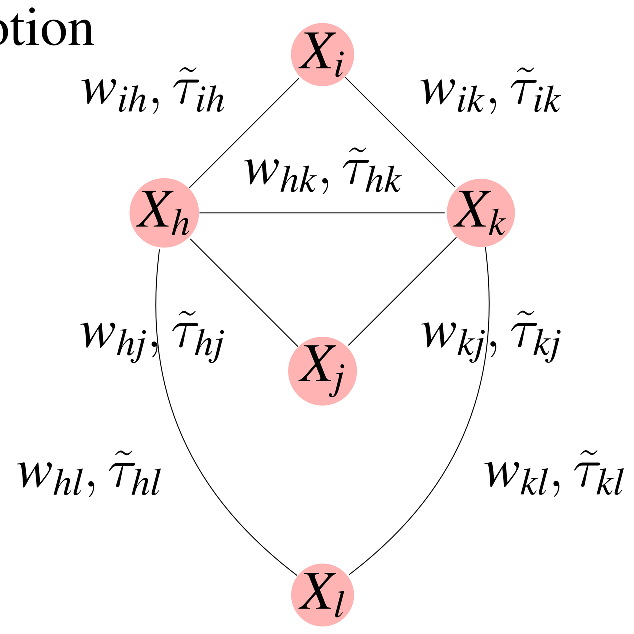
Problem Statement

- **Data:** Given m sets $X_1, X_2, \dots, X_m$ comprised of items from a ground set U
- **Observations: Coherently noisy** pairwise transformations $\tilde{\tau}_{ij}$ relating the sets X_i and X_j such that

$$\tilde{\tau}_{ij} \Rightarrow X_i \sim \tilde{\tau}_{ij}(X_j) \quad \forall 1 \le i, j \le m$$

- **Find:** Identify **correct** pairwise transformations between the sets by reasoning globally with noisy pairwise information of transformations
- **Why?** Many applications in machine learning and computer vision: Ranked data analysis and structure from motion

Strategy. Run diffusion process on this graph with transformation-attributed edges



This Paper

We present an algorithm for diffusion maps on this graph. In particular if:

- The transformation τ of interest is a member of a compact group G
- That is w.r.t the natural definition of multiplication, we have

$$(\tau'\tau)(i) := (\tau'(\tau(i))) \quad \tau, \tau' \in G,$$

We show that the invariance properties of functions defined on the group G provide a "certificate" of "consistent" association between $X_1, X_2, \dots, X_m$.

Structure from Motion

- Images are "related" by the matching between features i.e., $\tilde{\tau}$ is a permutation
- Set of all possible permutations of order n corresponds to a group – the symmetric group $\mathbb{S}_n$
- Presence of large repetitive structure introduce coherent errors and downstream analysis yields highly unsatisfactory results
- For example – image match matrix of a cup with 180° visual symmetry has large self-consistent erroneous match pairs.

Figure : Representative images.

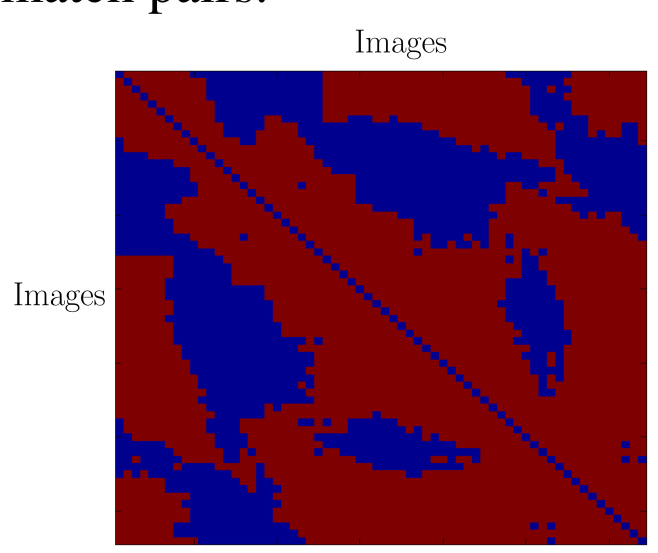


Figure : Visual cue image match matrix.

Synchronization by Permutation Diffusion

Proposition 1. Let G be any compact group with identity e and $\rho\colon G \to \mathbb{C}^{d_\rho \times d_\rho}$ be a unitary representation of G. Then given an array of possibly noisy and unsynchronized group elements (g_{ji}) and corresponding positive confidence weights (w_{ji}), any synchronization loss function (assuming $g_{ii} = e \forall i$)

$$\mathcal{E}(h_1, \dots, h_m) = \sum_{i,j=1}^{m} w_{ji} \|\rho(h_j h_i^{-1}) - \rho(g_{ji})\|^2_{\text{Frob}} \qquad h_1, \dots, h_m \in G$$

can be written in the form $\mathcal{E}(h_1, \dots, h_m) = V^\dagger \mathcal{L} V$, where

$$V = \begin{pmatrix} \rho(h_1) \\ \vdots \\ \rho(h_m) \end{pmatrix}, \qquad \mathcal{L} = \begin{pmatrix} d_1 I & -w_{21}\,\rho(g_{21}) & \dots & -w_{m1}\,\rho(g_{m1}) \\ \vdots & & \ddots & \vdots \\ -w_{1m}\,\rho(g_{1m}) & -w_{2m}\,\rho(g_{2m}) & \dots & d_m I \end{pmatrix}.$$

For **SfM with large repetitive structure**, we instantiate this proposition with

1. The observed but noisy matchings between images $\tilde{\tau}_{ij} \in \mathbb{S}_n$
2. The appropriate unitary representation of the $\mathbb{S}_n$

Uncertain Matches and Diffusion Distance

- Each block of the $V(i)$ correspond to a distribution $p_i(\sigma)$ over base permutations characterized by

1. assignments of landmarks that are seen in image $\mathcal{I}_i$
2. assignments of landmarks that are not seen in image $\mathcal{I}_i$

- $p_i(\sigma)$ is agnostic with respect to the assignment of occluded landmarks

Occlusion induces **an invariant structure** in $p_i(\sigma)$

Invariance Based Certificate of Association

We used auto-correlation function to represent the invariances of $p_i(\sigma)$

Permutation domain	Fourier domain
$a_i(\sigma) = \sum_{\mu \in \mathbb{S}_n} p_i(\sigma\mu)\, p_i(\mu).$	$\hat{a}_i(\rho) = \hat{p}_i(\rho)\, \hat{p}_i^\dagger(\rho).$

For some unitary matrix C, $V(i)$ can be expressed as direct sum of Fourier components of $p_i(\sigma)$

$$V(i) = C^\dagger \big[\bigoplus_{\lambda \in \Lambda} \hat{p}_i(\lambda) \big] C$$

Re-writing auto-correlation in the new basis

$$\hat{a}_i(\rho) := C^\dagger \big[\bigoplus_{\lambda \in \Lambda} \hat{a}_i(\lambda) \big] C = C^\dagger \big[\bigoplus_{\lambda \in \Lambda} \hat{p}_i(\lambda)\, \hat{p}_i(\lambda)^\dagger \big] C = V(i)\, V(i)^\top$$

Define **P**ermutation **D**iffusion **A**ffinity (**PDA**), which compares the invariance structure between some $\mathcal{I}_i$ and $\mathcal{I}_j$

$$\Pi(i,j) = \Big(\sum_{\sigma \in \mathbb{S}_n} a_i(\sigma)\, a_j(\sigma) \Big)^{1/2} = \operatorname{tr}\big(V(i)\, V(i)^\top V(j)\, V(j)^\top\big)^{1/2}$$

SFM – Cup Dataset Reconstruction

Figure : Visual cue.

Figure : PDA.

Experimental Results

OATS Dataset

Figure : Representative images.

Figure : Visual cue.

Figure : PDA.

STREET Dataset

Figure : Representative images.

Figure : Visual cue.

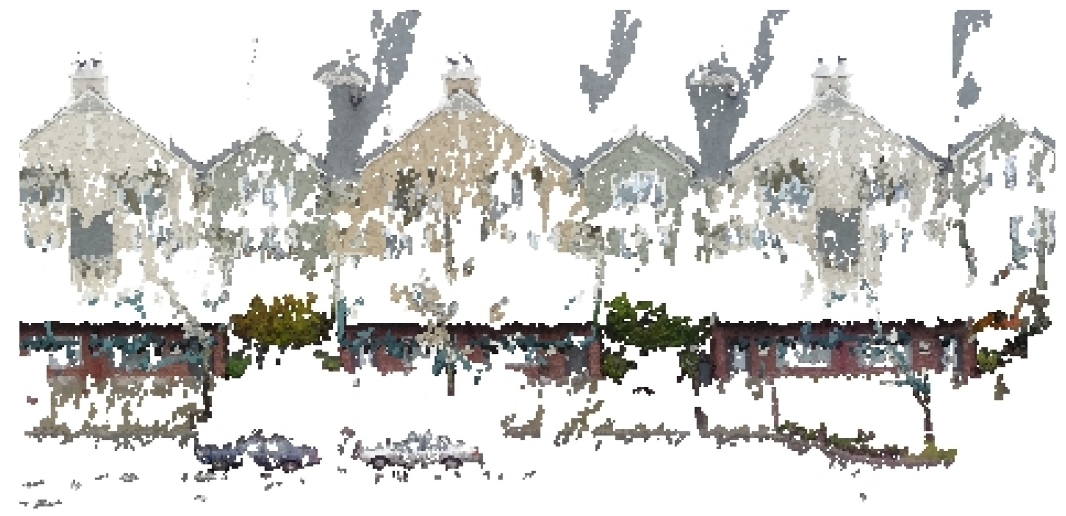


Figure : PDA.

NIPS 2014

This work was supported by NSF–1320344 and NSF–1320755.