Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence

A Local Sparse Model for Matching Problem

Bo Jiang[1], **Jin Tang**[1], **Chris Ding**[2,1] and **Bin Luo**[1]
[1]School of Computer Science and Technology, Anhui University, Hefei, 230601, China
[2]CSE Department, University of Texas at Arlington, Arlington, TX 76019, USA
jiangbo@ahu.edu.cn, ahhftang@gmail.com, chqding@uta.edu, luobin@ahu.edu.cn

Abstract

Feature matching problem that incorporates pairwise constraints is usually formulated as a quadratic assignment problem (QAP). Since it is NP-hard, relaxation models are required. In this paper, we first formulate the QAP from the match selection point of view; and then propose a local sparse model for matching problem. Our local sparse matching (LSM) method has the following advantages: (1) It is parameter-free; (2) It generates a local sparse solution which is closer to a discrete matrix than most other continuous relaxation methods for the matching problem. (3) The one-to-one matching constraints are better maintained in LSM solution. Promising experimental results show the effectiveness of the proposed LSM method.

Many problems of interest in computer vision and pattern recognition can be formulated as a problem of finding consistent correspondences between two sets of features. Since it is NP-hard, approximate relaxation methods are usually required (Albarelli et al. 2009; Gold and Rangarajan 1996; Cho, Lee, and Lee 2010; Zhou and la Torre 2012).

Recently, many relaxation methods have been developed to the feature matching problem (Albarelli et al. 2009; Cho, Lee, and Lee 2010; Leordeanu and Hebert 2005; Leordeanu, Hebert, and Sukthankar 2009; Rodolà et al. 2013; Torresani, Kolmogorov, and Rother 2008; Pachauri et al. 2012). Among these methods, one kind of popular relaxation methods is to solve the problem under different norm constraints (Albarelli et al. 2009; Leordeanu and Hebert 2005; Rodolà et al. 2012; Liu and Yan 2010; Rodolà et al. 2013). These methods can be well interpreted or inspired from **match selection** view, i.e., aim to select the subset of matches $\mathcal{S}^*$ with some constraint (*e.g.* one-to-one or one-to-many) from the potential match space $\mathcal{C}$ such that the total match score is maximized. Leordeanu et al. (Leordeanu and Hebert 2005) proposed a relaxed match selection model (spectral matching, SM) under a ℓ_2-norm constraint. The benefit of SM is that it has a closed-form optimal solution. As a post-selection step, SM further uses a greedy selection strategy to select the optimal matches $\mathcal{S}^*$ with mapping constraint. One drawback is that this post-selection step is independent with the matching problem (or match selection) objective, and thus may lead to weak local optimum. Inspired by game theory, Albarelli et al (Albarelli et al. 2009) proposed a relaxation matching method by replacing the ℓ_2 norm constraint in SM with ℓ_1-norm. Recently, this model has been widely used in many matching tasks (Albarelli, Rodolà, and Torsello 2012; Liu and Yan 2010; Rodolà et al. 2012). One important feature is that it can generate a sparse solution for the problem due to the ℓ_1 norm constraint, and thus has a desired selective behavior in nature. However, one main limitation is that the sparsity of its solution is generally uncontrolable and usually too high to select desired number of matches from the potential match space. Rodolà et al (Rodolà et al. 2013) recently provided a relaxation method with elastic net constraint and thus combined SM and game-theoretic matching simultaneously. The accuracy and sparsity can be balanced by a weighting parameter in this model (Rodolà et al. 2013).

Inspired by these works, our aim in this paper is to propose a new match selection method by exploring a local sparse model. Our local sparse matching (LSM) is motivated by the observations that: (1) the potential match space $\mathcal{C}$ has an inherent partition property; (2) For the one-to-one matching task, the optimal selected matches $\mathcal{S}^*$ must be uniformly distributed in all parts of $\mathcal{C}$. Our LSM is motivated by the desire to sparsely select matches from each part of $\mathcal{C}$. An effective algorithm can be derived to solve LSM problem. Promising experimental results on several matching tasks demonstrate the effectiveness of LSM method.

Problem Formulation and Related Works

Given two feature sets $\mathcal{M}$, containing n_M model features and $\mathcal{D}$, containing n_D data features, a corresponding mapping is a set $\mathcal{C}$ of potential matches (or candidate matches) (i, i'), where $i \in \mathcal{D}$ and $i' \in \mathcal{M}$. Generally, the size of $\mathcal{C}$ depends mainly on how discriminative the descriptors of the features are. When the features are non-discriminative, such as 2D or 3D points, all possible matches should be kept as candidate correspondences in $\mathcal{C}$. If the features are discriminative, such as SIFT descriptor, only small fraction of all possible matches (i, i') are kept in $\mathcal{C}$ (Leordeanu and Hebert 2005). For each match pair (a, b) in $\mathcal{C}$, where $a = (i, i')$ and $b = (j, j')$, there is an affinity $\mathbf{W}(a, b)$ that measures how compatible the feature pair (i, j) in $\mathcal{D}$ are with the feature pair (i', j') in $\mathcal{M}$.

From the point of view of **match selection**, the aim of the matching problem is to find or select the subset of matches $\mathcal{S}^*$ from $\mathcal{C}$ that maximizes the following total matching score,

$$\mathcal{S}^* = \arg\max_{\mathcal{S}\in\Omega} \sum_{a,b\in\mathcal{S}} \mathbf{W}(a,b) \tag{1}$$

where Ω denotes some mapping constraint such as one-to-one or one-to-many. In this paper, we focus on the one-to-one matching problem. In this case, the affinity for any conflict match pair, such as $a = (i,j)$ and $b = (i,k)$ $(j \neq k)$, can be penalized with smaller value (Cho, Lee, and Lee 2010; Leordeanu and Hebert 2005). We can represent the subset $\mathcal{S}$ by an indicator matrix $\mathbf{X}$, such that $\mathbf{X}_{ii'} = 1$ implies that feature i in $\mathcal{D}$ corresponds to feature i' in $\mathcal{M}$, and $\mathbf{X}_{ii'} = 0$ otherwise. Then, the objective of matching problem can be formulated as finding the indicator matrix $\mathbf{X}^*$ that maximizes the following objective score,

$$\max_{\mathbf{X}\in\Omega,\mathbf{X}_{ij}\in\{0,1\}} \sum_{ij,kl} \mathbf{W}_{ij,kl}\mathbf{X}_{ij}\mathbf{X}_{kl} = \mathrm{vec}\,(\mathbf{X})^{\mathrm{T}}\mathbf{W}\,\mathrm{vec}\,(\mathbf{X}) \tag{2}$$

where $\mathrm{vec}\,(\mathbf{X}) = (\mathbf{X}_{11}...\mathbf{X}_{1n},...,\mathbf{X}_{m1}...\mathbf{X}_{mn})^{\mathrm{T}} \in \mathbb{R}^{mn\times 1}$ is the vector form of $\mathbf{X}$, and $\mathbf{X}$ is called as the matrix form of vector $\mathrm{vec}\,(\mathbf{X})$ in the following. Ω denotes the mapping constraint.

The above problem Eq.(2) with integer constraint is NP-hard, and relaxation models are required. Here, we briefly review some popular models that are closely to our work.

Spectral matching (SM): By relaxing both integer and mapping constraints, Leordeanu et al. (Leordeanu and Hebert 2005) proposed a relaxed problem as follows,

$$\max_{\mathbf{X}} \ \mathrm{vec}\,(\mathbf{X})^{\mathrm{T}}\mathbf{W}\,\mathrm{vec}\,(\mathbf{X}) \quad \mathrm{s.t.} \ \ \|\mathbf{X}\|_2 = 1, \tag{3}$$

where $\|\mathbf{X}\|_2 = (\sum_{ij}\mathbf{X}_{ij}^2)^{1/2}$. It has a closed-form solution which is the leading eigenvector of $\mathbf{W}$.

Game-theoretic matching (GameM): From game-theoretic perspective, Albarelli et al. (Albarelli et al. 2009) have proposed a matching model by replacing the ℓ_2 norm constraint in SM with ℓ_1 norm, i.e.,

$$\max_{\mathbf{X}} \ \mathrm{vec}\,(\mathbf{X})^{\mathrm{T}}\mathbf{W}\,\mathrm{vec}\,(\mathbf{X}) \quad \mathrm{s.t.} \ \|\mathbf{X}\|_1 = 1, \mathbf{X} \geq 0. \tag{4}$$

where $\|\mathbf{X}\|_1 = \sum_{ij}|\mathbf{X}_{ij}|$. It can generate a sparse solution for the problem due to the ℓ_1 norm constraint, and thus has a desirable match selective behavior in nature. However, one drawback is that the selectivity (sparsity of the solution) demonstrated by this method is usually uncontrolable and too strong to select desired number of the matches.

Elastic net matching (EnetM): Rodolà et al. (Rodolà et al. 2013) recently introduced an elastic net model for match selection problem by combining both SM and GameM simultaneously as follows,

$$\begin{aligned} &\max_{\mathbf{X}} \ \ \mathrm{vec}\,(\mathbf{X})^{\mathrm{T}}\mathbf{W}\,\mathrm{vec}\,(\mathbf{X}) \quad (5)\\ &\mathrm{s.t.} \ \ (1-\alpha)\|\mathbf{X}\|_1 + \alpha\|\mathbf{X}\|_2^2 = 1, \mathbf{X} \geq 0, \end{aligned}$$

where $\alpha \in [0,1]$ is balanced parameter. One important benefit of EnetM method is that the optimality of SM and sparsity of GameM can be balanced via the weighting parameter α in this model.

Local Sparse Matching

LSM model

In this section, we present our local sparse matching (LSM) model, which is motivated by the following observations.

(1) The potential assignment space $\mathcal{C}, \mathcal{C} = \{(i,i')|i \in \mathcal{D}, i' \in \mathcal{M}\}$ can be divided as,

$$\mathcal{C} = \bigcup_{i=1}^{m} \mathcal{C}_i, \tag{6}$$

where $\mathcal{C}_i = \{(i,k)|i \in \mathcal{D}, k \in \mathcal{M} \ \text{ and } \ k = 1,2\cdots n\}$ and $\mathcal{C}_i \cap \mathcal{C}_j = \emptyset(i \neq j)$.

(2) For the one-to-one matching task, the optimal selected matches in $\mathcal{S}^*$ must be uniformly distributed in each subset $\mathcal{C}_i(i = 1,2\cdots m)$.

In both GameM (Albarelli et al. 2009) and EnetM (Rodolà et al. 2013), they just globally select $\mathcal{S}^*$ from $\mathcal{C}$ and thus ignore the uniform distribution constraint of $\mathcal{S}^*$. Our aim in this paper is to further incorporate this inherent distribution constraint in match selection. This motivates us to develop a local sparse model for the problem. Specifically, from the solution perspective, our LSM is motivated by the desire to encourage each **row** of converged solution $\mathbf{X}^*$ to be sparse while maximize the total matching score. This is achieved by using a mixed norm constraint.

By imposing a $\ell_{1,2}$-norm constraint on solution $\mathbf{X}$, our LSM can be formulated as,

$$\max_{\mathbf{X}} \ \ \mathrm{vec}\,(\mathbf{X})^{\mathrm{T}}\mathbf{W}\,\mathrm{vec}\,(\mathbf{X}) \ \ \mathrm{s.t.} \ \ \|\mathbf{X}\|_{1,2} = 1, \mathbf{X} \geq 0. \tag{7}$$

where $\|\mathbf{X}\|_{1,2} = \big(\sum_i(\sum_j|\mathbf{X}_{ij}|)^2\big)^{1/2}$. Note that the ℓ_1-norm on each row of $\mathbf{X}$ encourages sparsity in each row. Also, the ℓ_2-norm on rows of matrix encourages that there is no zero row in the solution matrix. Thus, the optimal solution $\mathbf{X}^*$ of LSM can be local sparse due to both ℓ_1 and ℓ_2 norms, which is desirable for the matching problem. Our LSM has three main benefits: (1) comparing with EnetM, it is a parameter-free balanced model between SM and GameM. Thus, the selectivity (sparsity of the solution) can be more stable and controlable than GameM; (2) It incorporates more mapping constraint in the match selection; (3) An simple and effective update algorithm can be derived to solve it.

Computational algorithm

The proposed LSM can be effectively solved by an iterative algorithm. The algorithm iteratively updates a current solution $\mathbf{X}^{(t)}$ as follows,

$$\mathbf{X}_{ij}^{(t+1)} = \mathbf{X}_{ij}^{(t)}\sqrt{\frac{\mathbf{K}_{ij}^{(t)}}{\lambda\sum_j \mathbf{X}_{ij}^{(t)}}}, \tag{8}$$

where matrix $\mathbf{K}^{(t)} \in \mathbb{R}^{m\times n}$ is the matrix form of the vector $[\mathbf{W}^{(t)}\,\mathrm{vec}\,(\mathbf{X}^{(t)})]$, and λ is computed as,

$$\lambda = \mathrm{vec}\,(\mathbf{X}^{(t)})^{\mathrm{T}}\mathbf{W}\,\mathrm{vec}\,(\mathbf{X}^{(t)}). \tag{9}$$

The iteration starts with an initial $\mathbf{X}^{(0)}$ and is repeated until convergence. Since $\mathbf{W}$ is a real nonnegative matrix for the matching problem, the nonnegativity of $\mathbf{X}^{(t)}$ is guaranteed in each iteration.

Theoretical analysis

The optimality and convergence of the algorithm is guaranteed by Theorem 1 and Theorem 2, respectively.

Theorem 1 *Update rule of Eq.(8) satisfies the first-order Karush-Kuhn-Tucker (KKT) optimality condition.*

Proof Since the constraint $\|\mathbf{X}\|_{1,2} = 1$ can be replaced by $\|\mathbf{X}\|_{1,2}^2 = 1$, the Lagrangian function is

$$\mathcal{L}(\mathbf{X}) = \mathrm{vec}\,(\mathbf{X})^{\mathrm{T}}\mathbf{W}\,\mathrm{vec}\,(\mathbf{X}) - \lambda(\sum_i(\sum_j \mathbf{X}_{ij})^2 - 1). \quad (10)$$

The KKT complementary slackness condition is

$$\frac{\partial \mathcal{L}}{\partial \mathbf{X}_{kl}}\mathbf{X}_{kl} = 2\Big[(\mathbf{W}\,\mathrm{vec}\,(\mathbf{X}))_{kl} - \lambda\sum_j \mathbf{X}_{kj}\Big]\mathbf{X}_{kl} = 0. \quad (11)$$

Summing over index l, we obtain,

$$2\sum_l \big[(\mathbf{W}\,\mathrm{vec}\,(\mathbf{X}))_{kl}\mathbf{X}_{kl}\big] - 2\lambda(\sum_l \mathbf{X}_{kl})^2 = 0. \quad (12)$$

Because $\sum_k(\sum_l \mathbf{X}_{kl})^2 = 1$, summing over index k, we can obtain the value for λ as,

$$\lambda = \mathrm{vec}\,(\mathbf{X})^{\mathrm{T}}\mathbf{W}\,\mathrm{vec}\,(\mathbf{X}). \quad (13)$$

Note that, at convergence of update rule Eq.(8),

$$\mathbf{X}^*_{kl} = \mathbf{X}^*_{kl}\sqrt{\frac{\mathbf{K}^*_{kl}}{\lambda\sum_l \mathbf{X}^*_{kl}}}, \quad (14)$$

or $\big(\mathbf{K}^*_{kl} - \lambda\sum_l \mathbf{X}^*_{kl}\big)\mathbf{X}^*_{kl} = 0$, where $\mathbf{K}$ is the matrix form of $[\mathbf{W}\,\mathrm{vec}\,(\mathbf{X})]$. This is the KKT condition of Eq.(11). Substituting λ from Eq.(13), we can obtain the update rule Eq.(8).

Theorem 2 *Under the update rule Eq.(8), the Lagrangian function $\mathcal{L}(X)$ Eq.(10) is monotonically increasing.*

Proof We use the auxiliary function method (Ding, Li, and Jordan 2010; Lee and Seung 2001). An auxiliary function function $\mathcal{Z}(\mathbf{X}, \mathbf{X}')$ of Lagrangian function $\mathcal{L}(\mathbf{X})$ satisfies

$$\mathcal{Z}(\mathbf{X}, \mathbf{X}) = \mathcal{L}(\mathbf{X}), \mathcal{Z}(\mathbf{X}, \mathbf{X}') \leq \mathcal{L}(\mathbf{X}). \quad (15)$$

In the following, we define

$$\mathbf{X}^{(t+1)} = \arg\max_{\mathbf{X}} \mathcal{Z}(\mathbf{X}, \mathbf{X}^{(t)}). \quad (16)$$

Then, we have

$$\mathcal{L}(\mathbf{X}^{(t)}) = \mathcal{Z}(\mathbf{X}^{(t)}, \mathbf{X}^{(t)}) \leq \mathcal{Z}(\mathbf{X}^{(t+1)}, \mathbf{X}^{(t)}) \leq \mathcal{L}(\mathbf{X}^{(t+1)}) \quad (17)$$

This proves that $\mathcal{L}(\mathbf{X}^{(t)})$ is monotonically increasing.

We rewrite Eq.(10) as

$$\mathcal{L}(\mathbf{X}) = \sum_{ij}\sum_{kl}\mathbf{W}_{ij,kl}\mathbf{X}_{ij}\mathbf{X}_{kl} - \lambda(\mathbf{e}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{X}\mathbf{e} - 1), \quad (18)$$

where $\mathbf{e} = (1, 1, \cdots 1)^{\mathrm{T}} \in \mathbb{R}^{n\times 1}$. We can show that one auxiliary function of $\mathcal{L}(\mathbf{X})$ is

$$\mathcal{Z}(\mathbf{X}, \mathbf{X}') = \sum_{ij}\sum_{kl}\mathbf{W}_{ij,kl}\mathbf{X}'_{ij}\mathbf{X}'_{kl}\Big(1 + \log\frac{\mathbf{X}_{ij}\mathbf{X}_{kl}}{\mathbf{X}'_{ij}\mathbf{X}'_{kl}}\Big) - \lambda\sum_i\sum_j \frac{(\mathbf{X}'\mathbf{e}\mathbf{e}^{\mathrm{T}})_{ij}\mathbf{X}^2_{ij}}{\mathbf{X}'_{ij}}. \quad (19)$$

Using the inequality $z \geq 1 + \log z$ and setting $z = \frac{\mathbf{X}_{ij}\mathbf{X}_{kl}}{\mathbf{X}'_{ij}\mathbf{X}'_{kl}}$, the first term in Eq.(19) is a lower bound of the first term in Eq.(18). Also, for any positive matrices $\mathbf{A}, \mathbf{B}, \mathbf{S}$ and $\mathbf{S}'$ with $\mathbf{A}$ and $\mathbf{B}$ symmetric, the following inequality always holds (Ding, Li, and Jordan 2010),

$$\sum_i\sum_j \frac{(\mathbf{A}\mathbf{S}'\mathbf{B})_{ij}\mathbf{S}^2_{ij}}{\mathbf{S}'_{ij}} \geq \mathrm{Tr}(\mathbf{S}^{\mathrm{T}}\mathbf{A}\mathbf{S}\mathbf{B}) = \mathrm{Tr}(\mathbf{S}\mathbf{B}\mathbf{S}^{\mathrm{T}}\mathbf{A}). \quad (20)$$

Using this general equality, we can have,

$$\sum_i\sum_j \frac{(\mathbf{X}'\mathbf{e}\mathbf{e}^{\mathrm{T}})_{ij}\mathbf{X}^2_{ij}}{\mathbf{X}'_{ij}} \geq \mathrm{Tr}(\mathbf{X}\mathbf{e}\mathbf{e}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}) = \mathbf{e}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{X}\mathbf{e}. \quad (21)$$

Thus, the second term in Eq.(19) is a lower bound of the second term in Eq.(18). Therefore, $\mathcal{Z}(\mathbf{X}, \tilde{\mathbf{X}})$ is an auxiliary function of $\mathcal{L}(\mathbf{X})$.

According to Eq.(16), we need to find the global maximum of $\mathcal{Z}(\mathbf{X}, \mathbf{X}')$ for $\mathbf{X}$. The gradient is

$$\frac{\partial \mathcal{Z}(\mathbf{X}, \mathbf{X}')}{\partial \mathbf{X}_{kl}} = 2\Big[\frac{(\mathbf{W}\,\mathrm{vec}\,(\mathbf{X}'))_{kl}\mathbf{X}'_{kl}}{\mathbf{X}_{kl}} - \lambda\frac{(\mathbf{X}'\mathbf{e}\mathbf{e}^{\mathrm{T}})_{kl}\mathbf{X}_{kl}}{\mathbf{X}'_{kl}}\Big]$$

Note that $\mathbf{W} = \mathbf{W}^{\mathrm{T}}$ and the second derivative is

$$\frac{\partial^2 \mathcal{Z}(\mathbf{X}, \mathbf{X}')}{\partial \mathbf{X}_{kl}\partial \mathbf{X}_{ij}} = -\Big[(\mathbf{W}\,\mathrm{vec}\,(\mathbf{X}'))_{kl}\frac{\mathbf{X}'_{kl}}{\mathbf{X}^2_{kl}} + \lambda\frac{(\mathbf{e}\mathbf{e}^{\mathrm{T}}\mathbf{X}')_{kl}}{\mathbf{X}'_{kl}}\Big]\delta_{ki}\delta_{lj} \leq 0$$

where $\delta_{uv} = 1$ if $u = v$, othewise 0. Therefore, $\mathcal{Z}(\mathbf{X}, \mathbf{X}')$ is a concave function in $\mathbf{X}$. Thus, it has a global maximum, which can be obtained by setting the first derivative to 0, i.e.,

$$\mathbf{X}_{kl} = \mathbf{X}'_{kl}\sqrt{\frac{\big(\mathbf{W}\,\mathrm{vec}\,(\mathbf{X}')\big)_{kl}}{\lambda(\mathbf{X}'\mathbf{e}\mathbf{e}^{\mathrm{T}})_{kl}}}. \quad (22)$$

Because $(\mathbf{X}'\mathbf{e}\mathbf{e}^{\mathrm{T}})_{kl} = \sum_l \mathbf{X}'_{kl}$, thus we obtain update Eq.(8) by setting $\mathbf{X}^{(t+1)} = \mathbf{X}$ and $\mathbf{X}^{(t)} = \mathbf{X}'$.

Sparsity and Desired Selectivity

As discussed before, one important benefit of LSM is that it can generate a local sparse solution and thus performs match selection naturally. Figure 1 (c) shows solution $\mathbf{X}^{(t)}$ (top row) and associated selected matches $\mathcal{S}^{(t)}$ (bottom row) across different iterations under LSM algorithm Eq.(8). All possible assignments are included in $\mathcal{C}$ in this example. The ground truth selection and its matrix solution are shown in Figure 1(a) and (b), respectively. Here, the selected matches $\mathcal{S}^{(t)}$ are obtained by follows. Firstly, set threshold $\delta_t = 0.001 \times \mathrm{mean}(\mathbf{X}^{(t)})$. Then, define $\mathcal{S}^{(t)}$ as $\mathcal{S}^{(t)} = \{(i, j)|i \in \mathcal{D}, j \in \mathcal{M} \text{ and } \mathbf{X}^{(t)}_{ij} > \delta_t\}$. Intuitively, we note that as the iteration increases, the solution $\mathbf{X}^{(t)}$ of LSM becomes more and more sparse and approximates the ground truth matrix solution more and more closely, which clearly indicates the ability of LSM to select the desired matches $\mathcal{S}^*$ from the potential match space $\mathcal{C}$, as shown in bottom row of Figure 1(c).

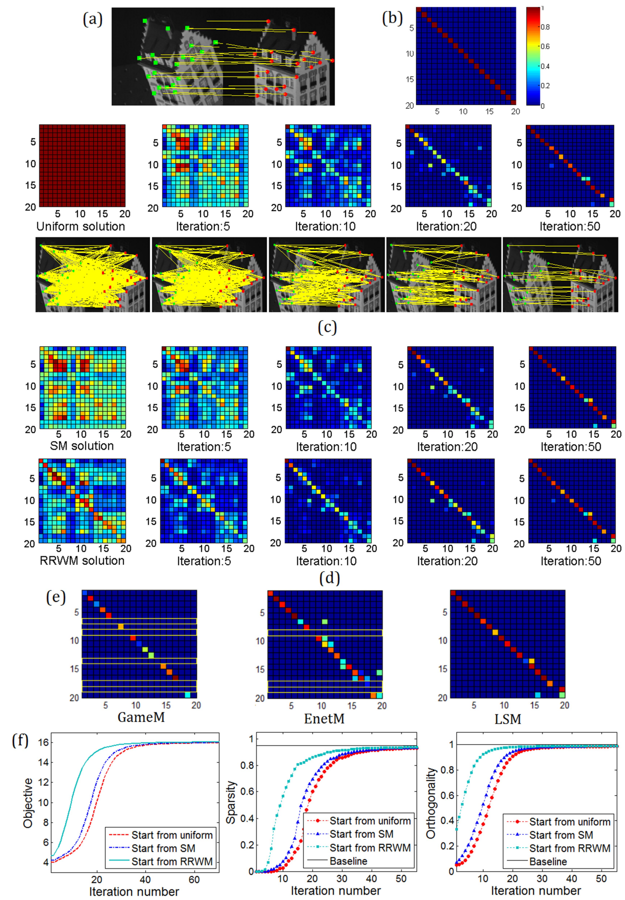

Figure 1: (a-b) Correct match, and its solution. (c) solution $\mathbf{X}^{(t)}$ (upper row) and associated matches $\mathcal{S}^{(t)}$ (lower row) at 5 iterations; (d) solution $\mathbf{X}^{(t)}$ at several iterations; TOP: starting from SM solution; BOTTOM: starting from RRWM solution; (e) converged solution $\mathbf{X}^*$ of GameM, EnetM and LSM, respectively. In GameM and EnetM, some rows of $\mathbf{X}^*$ become zero (undesirably), marked by yellow rectangles; (f) objective, sparsity and orthogonality of $\mathbf{X}^{(t)}$ for LSM.

Corollary 3 *Under LSM algorithm, as the iteration increases, the selected matches $\mathcal{S}^{(t)}$ satisfies,*

$$\mathcal{C} = \mathcal{S}^{(0)} \supseteq \cdots \mathcal{S}^{(t)} \supseteq \mathcal{S}^{(t+1)} \cdots \supseteq \mathcal{S}^*, t = 1, 2, \cdots .$$

Proof Assume that there exists a match (i_0, j_0), where $(i_0, j_0) \in \mathcal{S}^{(t+1)}$ and $(i_0, j_0) \notin \mathcal{S}^{(t)}$, i.e., $\mathbf{X}^{(t)}_{i_0 j_0} \to 0$ and $\mathbf{X}^{(t+1)}_{i_0 j_0} \nrightarrow 0$. From update Eq.(8), one can note that if $\mathbf{X}^{(t)}_{i_0 j_0} \to 0$, then $\mathbf{X}^{(t+1)}_{i_0 j_0} \to 0$, i.e., $(i_0, j_0) \notin \mathcal{S}^{(t+1)}$. This leads to a contradiction. Thus, we have $\mathcal{S}^{(t)} \supseteq \mathcal{S}^{(t+1)}$.

Figure 1 (d) shows solution $\mathbf{X}^{(t)}$ across iterations with different initializations. Figure 1 (e) shows the converged solution $\mathbf{X}^*$ of the match selection method GameM, EnetM and LSM, respectively. Comparing with GameM and EnetM, we can note that (1) the number of non-zero elements (sparsity) in LSM solution $\mathbf{X}^*$ is desirable in general, indicating that LSM can select the desired number of the matches. (2) More importantly, the non-zero elements in LSM solution $\mathbf{X}^*$ are generally located in each row of $\mathbf{X}^*$, which demonstrates that the selected matches are uniformly distributed in each $\mathcal{C}_i, i = 1, 2, \cdots m$, as discussed before. For further illustration, we introduce two measurements, namely sparsity and orthogonality (Ding, Li, and Jordan 2010).

Sparsity measures the percentage of zero (close-to-zero) elements in $\mathbf{A}$. Firstly, set the threshold $\delta = 0.001 \times \mathrm{mean}(\mathbf{A})$, then renew $\mathbf{A}_{ij} = 0$ if $\mathbf{A}_{ij} \leq \delta$. Finally, the percentage of zero elements in the renewed $\mathbf{A}$ is calculated as sparsity.

Orthogonality measures the level of the orthogonality of $\mathbf{A}$. Firstly, compute the normalized matrix $\mathbf{Q} = \mathbf{C}^{-1/2}\mathbf{A}\mathbf{A}^{\mathrm{T}}\mathbf{C}^{-1/2}$, where $\mathbf{C} = \mathrm{diag}(\mathbf{A}\mathbf{A}^{\mathrm{T}})$. Then, the orthogonality is defined as: $\mathrm{Orthogonality}(\mathbf{A}) = 1 - \Delta(\mathbf{Q})$, where $\Delta(\mathbf{Q})$ denotes the average of the off-diagonal in $\mathbf{Q}$. Figure 1 (f) shows the objective, sparsity and orthogonality of solution $\mathbf{X}^{(t)}$ under LSM algorithm Eq.(8) with different initializations. We can observe that: (1) regardless of initialization, the objective of $\mathbf{X}^{(t)}$ increases and converges after some iterations, demonstrating the convergence of LSM algorithm. (2) The sparsity of $\mathbf{X}^{(t)}$ increases and converges to the baseline ($\mathrm{Sparsity} = 380/400 = 0.95$ in this example) after some iterations, indicating the ability of LSM to select the desired number of matches. (3) The orthogonality of $\mathbf{X}^{(t)}$ increases and almost converges to an orthogonal matrix, i.e., $\mathrm{Orthogonality}(\mathbf{X}^*) \to 1$. This is a bit interesting because the orthogonal constraint which encodes one-to-one mapping has been entirely ignored in LSM model, although the affinity between two conflict matches has been penalized (see Section 2). These observations are generally consistent with the intuitive results shown in Figure 1 (c, d), and will further be quantified in experiments.

Experiments

In this section, we apply LSM to some matching tasks. We compare our LSM with the match selection method SM (Leordeanu and Hebert 2005), GameM (Albarelli et al. 2009), and EnetM (Rodolà et al. 2013). The parameter α in EnetM was set to obtain the solution which has the similar sparsity with LSM solution. Also, we compare LSM with some other matching methods including IPFP (Leordeanu, Hebert, and Sukthankar 2009), SMAC (Cour, Srinivasan, and Shi 2006) and RRWM (Cho, Lee, and Lee 2010).

Synthetic data

Following the experimental setting (Cho, Lee, and Lee 2010; Leordeanu and Hebert 2005), we have randomly generated data sets of n_M 2D model point set $\mathcal{M}$. The range of the x-y point coordinates is $\sqrt{n_M/10}$ to enforce a constant density of 10 points over a 1×1 region. We obtain the corresponding n_D data features in $\mathcal{D}$ by transforming the whole data set with a random rotation and then adding Gaussian noise $N(0, \sigma)$ to the n_M point positions from $\mathcal{M}$. In addition to the deformation noise, we also add n_{out} outlier points in both $\mathcal{M}$ and $\mathcal{D}$ at random positions. The affinity matrix $\mathbf{W}$ is computed by $\mathbf{W}_{ij,kl} = \exp(-\|\mathbf{r}^{\mathcal{D}}_{ik} - \mathbf{r}^{\mathcal{M}}_{jl}\|^2_F/\sigma_r)$, where σ_r was set to 0.05 in this experiment, and $\mathbf{r}^{\mathcal{D}}_{ik}$ is the Euclidean

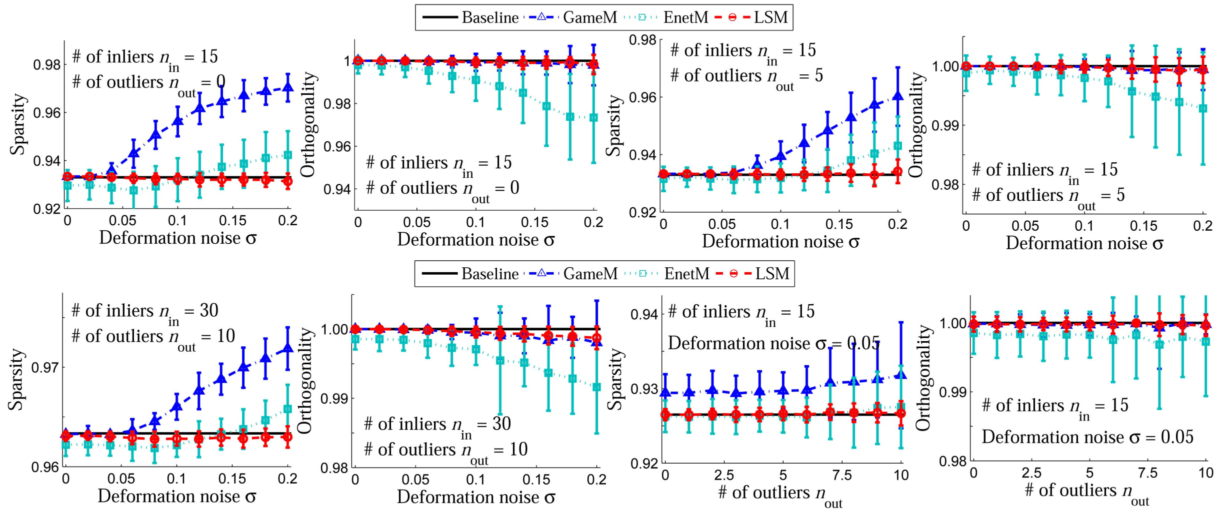

Figure 2: Comparison results on sparsity and orthogonality on synthetic data. Error bars correspond to standard errors.

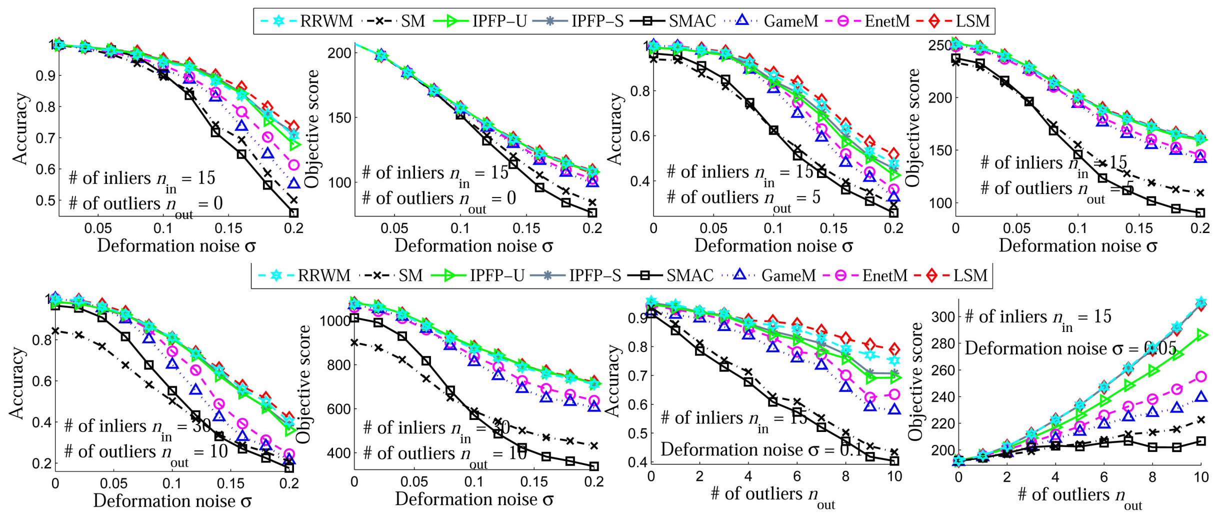

Figure 3: Comparison results on accuracy and objective score on synthetic data.

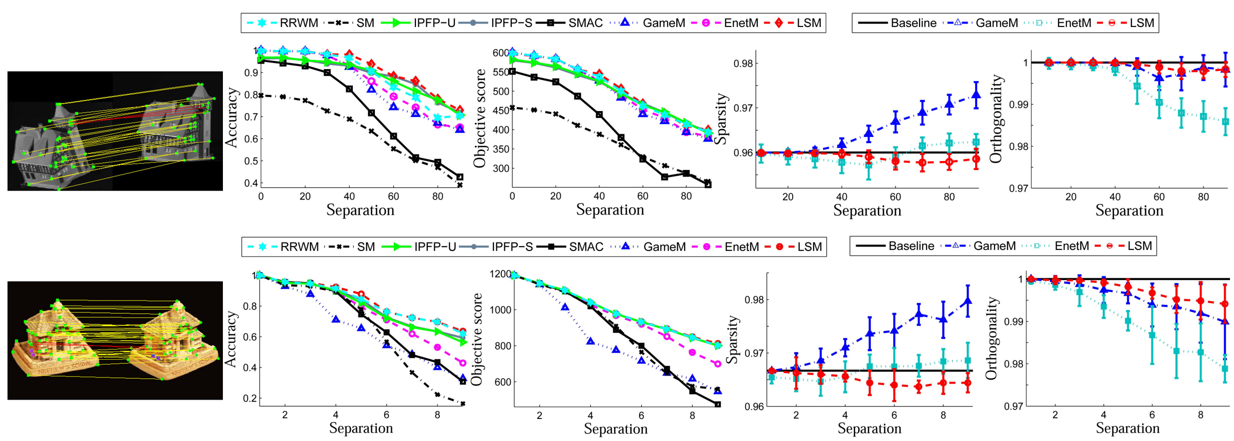

Figure 4: Matching results across image sequences (TOP: CMU sequence; BOTTOM: YORK sequence).

distance between two points. We have generated 200 random point sets for each noise level and then computed the average performances. Figure 2 shows the average and standard error of sparsity and orthogonality of the converged solution for different match selection methods. Here, we only compare our LSM with GameM and EnetM, because the methods, such as SM, RRWM and IFPF, generally return either continuous (non-sparse) or strictly discrete orthogonal solution. Here we note that (1) when the noise is large, the solution of GameM is generally too sparse to select desired number of matches, and its sparsity is usually changeable and uncontrolable. (2) Comparing with GameM, the sparsity of EnetM solution is more stable and controlable. However, the orthogonality of EnetM is lower than GameM, indicating that the one-to-one mapping constraint has been dropped more in EnetM. (3) Our LSM generally returns the most desirable sparse and (approximate) orthogonal solution for the problem. Figure 3 summarizes the comparison results on matching accuracy and objective score. Noted that: (1) comparing with SM, GameM and EnetM, LSM returns the highest objective score and accuracy, which clearly indicates that LSM can find the solution more optimal that these methods. (2) LSM also performs better than SMAC, RRWM and IPFP, demonstrating the effectiveness of LSM method.

Feature point matching across image sequence

In this section, we perform feature matching on CMU and YORK sequences (Zhou and la Torre 2012; Cho, Lee, and Lee 2010; Luo and Hancock 2001). For CMU sequence, there are 111 images of a toy house captured from moving viewpoints. For each image, 30 landmark points were manually marked with known correspondences and 5 outlier points were randomly generated. We have matched all images spaced by 10, 20, $\cdots$ 90 frames and computed the average performance per separation gap. YORK sequence contains 18 images and the adjacent images were obtained according to the rotation of 5 degree. For each image, 35 landmark points were manually marked with known correspondences. We have matched all images spaced by 1, 2 $\cdots$ 10 frames. The coordinates of their landmark points have been first normalized to the interval [0, 1], then the affinity matrix has been computed by $\mathbf{W}_{ij,kl} = \exp(-\|\mathbf{r}_{ik}^D - \mathbf{r}_{jl}^M\|_F^2/\sigma_r)$, where $\mathbf{r}_{ik}^D$ is the Euclidean distance between two points, σ_r was set to 0.025. Here, we can note that as the separation increases, the sparsity and orthogonality of LSM are more stable and desirable than GameM and EnetM. Also, LSM generally returns the best performance on matching accuracy and objective score, showing the optimality of LSM solution. These are consistent with the results on the synthetic data and further demonstrate the practicality and benefits of the proposed LSM method.

Real-world image matching

In this section, we first evaluate our method on the image pairs (30 pairs) selected from Caltech-101 and MSRC datasets (Cho, Lee, and Lee 2010). Following the experimental setting (Cho, Lee, and Lee 2010), the candidate correspondences have been generated using the MSER detector and SIFT feature descriptors. For each candidate match (i, j) and (k, l), the affinity $\mathbf{W}_{ij,kl}$ has been computed as $\mathbf{W}_{ij,kl} = \exp(-s_{ij,kl}/1500)$, where $s_{ij,kl}$ denotes the mutual projection error function used in (Cho, Lee, and Lee 2010). Our second evaluation has been performed on the image pairs (20 pairs) selected from Zurich Building Image Database (ZuBud) (Ng and Kingsbury 2010). Feature points and candidate correspondences have been detected and generated using the SIFT descriptor (Ng and Kingsbury 2010). The ground truths (correct matches) have been manually labeled for each image pair. The affinity matrix $\mathbf{W}_{ij,kl}$ has been computed as $\mathbf{W}_{ij,kl} = \exp(-|d_{ik} - d_{jl}|/1500)$, where d_{ik} is the Euclidean distance between the feature points i and k.For all image pairs in these two datasets, the average performances including true positive and false positive are computed. The comparison results are summarized in Fig. 5 (c), and some examples are shown in Fig. 5 (a-b) ((a) Caltech-101-MSRC image, (b) ZuBud image). Here, we can note that LSM returns higher true positive while maintains lower false positive value than other comparison methods, which further demonstrates the effectiveness of LSM method on conducting real-world image matching tasks.

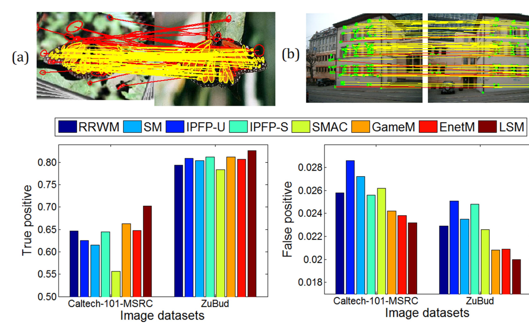

Figure 5: Matching results on two real-world image datasets.

Conclusions

In this work, we first formulate the matching problem to a match selection problem, and then develop a local sparse matching (LSM) method to the problem. The key point of LSM is that a $\ell_{1,2}$-norm constraint has been imposed and thus generates a desired local sparse solution for the matching problem. Experimentally, our LSM based solution incorporates more matching constraints in match selection process. A simple and effective algorithm has been derived. Promising results show the benefits of LSM method.

Acknowledgments

This work was supported in part by the National High Technology Research and Development Program (863 Program) of China (2014AA012204), in part by Sino-UK Higher Education Research Partnership for PhD Studies (CSC NO.201301310003), in part by the National Natural Science Foundation of China (61472002).

References

Albarelli, A.; Bulo, S. R.; Torsello, A.; and Pelillo, M. 2009. Matching as a non-coorperative game. In *ICCV*, 1319–1326.

Albarelli, A.; Rodolà, E.; and Torsello, A. 2012. Imposing semi-local geometric constraints for accurate correspondences selection in structure from motion: a game-theoretic perspective. *International Journal of Computer Vision* 97(1):36–53.

Cho, M.; Lee, J.; and Lee, K. M. 2010. Reweighted random walks for graph matching. In *ECCV*, 492–505.

Cour, M.; Srinivasan, P.; and Shi, J. 2006. Balanced graph matching. In *NIPS*, 313–320.

Ding, C.; Li, T.; and Jordan, M. I. 2010. Convex and semi-nonnegative matrix factorization. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 32(1):45–55.

Gold, S., and Rangarajan, A. 1996. A graduated assignment algorithm for graph matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 18(4):377–388.

Lee, D. D., and Seung, H. S. 2001. Algorithms for nonnegative matrix factorization. In *NIPS*, 556–562.

Leordeanu, M., and Hebert, M. 2005. A spectral technique for correspondence problem using pairwise constraints. In *ICCV*, 1482–1489.

Leordeanu, M.; Hebert, M.; and Sukthankar, R. 2009. An integer projected fixed point method for graph macthing and map inference. In *NIPS*, 1114–1122.

Liu, H., and Yan, S. 2010. Common visual pattern discovery via spatially coherent correspondences. In *CVPR*, 1609–1616.

Luo, B., and Hancock, E. 2001. Structural graph matching using the em algorithm and singlular value decomposition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 23(10):26–33.

Ng, E. S., and Kingsbury, N. G. 2010. Matching of interest point groups with pairwise spatial constraints. In *ICIP*, 2693–2696.

Pachauri, D.; Collins, M.; Kondor, R.; and Singh, V. 2012. Incoporating domain knowledge in matching problems via harmonic analysis. In *ICML*, 1271–1278.

Rodolà, E.; Bronstein, A. M.; Albarelli, A.; Bergamasco, F.; and Torsello, A. 2012. A game-theoretic approach to deformable shape matching. In *CVPR*, 182–189.

Rodolà, E.; Torsello, A.; Harada, T.; Kuniyoshi, Y.; and Cremers, D. 2013. Elastic net constraints for shape matching. In *ICCV*, 1169–1176.

Torresani, L.; Kolmogorov, V.; and Rother, C. 2008. Feature correspondence via graph matching: Models and global optimization. In *ECCV*, 596–609.

Zhou, F., and la Torre, F. D. 2012. Factorized graph matching. In *CVPR*, 127–134.