Robust Feature Point Matching With Sparse Model

Bo Jiang, Jin Tang, Bin Luo, *Senior Member, IEEE*, and Liang Lin

***Abstract*— Feature point matching that incorporates pairwise constraints can be cast as an integer quadratic programming (IQP) problem. Since it is NP-hard, approximate methods are required. The optimal solution for IQP matching problem is discrete, binary, and thus sparse in nature. This motivates us to use sparse model for feature point matching problem. The main advantage of the proposed sparse feature point matching (SPM) method is that it generates sparse solution and thus naturally imposes the discrete mapping constraints approximately in the optimization process. Therefore, it can optimize the IQP matching problem in an approximate discrete domain. In addition, an efficient algorithm can be derived to solve SPM problem. Promising experimental results on both synthetic points sets matching and real-world image feature sets matching tasks show the effectiveness of the proposed feature point matching method.**

***Index Terms*— Feature point matching, integer quadratic programming, nonnegative matrix factorization, sparse model.**

I. Introduction

MANY problems in computer vision can be formulated as a problem of finding consistent correspondences between two sets of features [1]–[4]. The goal of feature point matching is to find a mapping between two feature sets that preserves both unary features and binary relationships between feature points as much as possible [1], [4], [5]. Previous approaches have formulated feature point matching as an Integer Quadratic Programming (IQP) problem [5]–[9]. Since it is known to be NP-hard, feature point matching is usually solved approximately. Most of the recent literatures have developed approximate relaxations to the feature point matching problem [5], [6], [10]–[12], [15]. van Wyk et al. [12] proposed a matching method by iteratively projecting the approximate correspondence matrix onto the desired convex domain. Leordeanu and Hebert [7] proposed a simple approximate method (spectral matching, SM) to feature matching problem using a spectral relaxation technique. This method can find the global maximum solution of the relaxed continuous problem efficiently by computing the leading eigenvector of the symmetric nonnegative affinity matrix. However, SM ignores the mapping constraints in the relaxation step, and it obtains the final matching solution by using a greedy technique. Cour et al. [16] extended SM to spectral matching with affine constraint (SMAC) by incorporating the affine constraints into the spectral relaxation. Comparing with SM, it further encodes the one-to-one matching constraint, therefore it can approximate the original IQP problem more closely. Torresani et al. [9] represented the feature matching as an energy minimization problem which can be efficiently optimized by dual decomposition.

The above relaxation methods generally first develop a new continuous problem by relaxing the discrete mapping constraints and then aim to find the global optimum for this new continuous problem. At last, they obtain the final discretely binarized solution based on a post-optimization step (binarized step) using some discretization techniques such as Hungarian or greedy algorithm. The basic assumption behind these methods is that the continuous optimum of the relaxation problem is close to the discrete global optimum of the original IQP problem [17]. Recently, Leordeanu et al. [17] proposed an iterative matching method (IPFP) which optimized the IQP in the discrete domain and therefore satisfies the one-to-one mapping constraints strictly in the optimization process. It integrates the discretization step and objective function optimization simultaneously, and shows strong climbing and convergence properties. Based on IPFP, Leordeanu et al. [17] further show experimentally that, searching for a discrete solution instead of continuous one is generally more beneficial for finding the global optimum of IQP problem. In addition to optimization-based methods, probabilistic frameworks can also be used for interpreting and solving feature matching problems [13], [14], [18], [19]. Zass and Shashua [14] introduced a probabilistic model for soft Hypergraph matching between complex feature sets. Cho and Lee [13] interpreted graph matching problem using a random walk model and provided a robust matching algorithm by simulating random walks with re-weighting jumps enforcing the mapping constraints on the associated graph. Caetano et al. [18], [19] formulated feature point matching as a problem of finding a maximum probability configuration in a graphical model. In our work, we focus on optimization-based matching methods.

The optimal solution for feature matching problems (IQP) is discrete, binary and thus sparse in nature. To the best of our knowledge, most existing relaxation methods do not emphasize this sparse property. This motivates us to use sparse model for feature point matching problem. Following this way, we propose a new relaxation method for feature point matching problem by adapting a sparse model. There are

Manuscript received May 4, 2013; revised December 23, 2013 and June 6, 2014; accepted September 29, 2014. Date of publication October 9, 2014; date of current version October 24, 2014. This work was supported in part by the National High Technology Research and Development Program (863 Program) of China under Grant 2014AA012204, in part by the National Natural Science Foundation of China under Grant 61202228 and Grant 61472002, and in part by the Science and Technology Program of Guangzhou under Grant 2013J4500063. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Hassan Foroosh.

B. Jiang, J. Tang, and B. Luo are with the School of Computer Science and Technology, Anhui University, Hefei 230026, China (e-mail: zeyiabc@163.com; ahhftang@gmail.com; luobin@ahu.edu.cn).

L. Lin is with the School of Advanced Computing, Sun Yat-sen University, Guangzhou 510275, China (e-mail: linliang@ieee.org).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIP.2014.2362614

three main contributions. (1) A general quadratic programming problem with nonnegative and sparse constraint is proposed, followed by an efficient algorithm to solve it. (2) A new relaxation problem for feature point matching is proposed by using the proposed quadratic sparse model. (3) We show experimentally that our sparse feature matching method (SPM) can generate sparse solution and thus can approximately incorporate the discrete mapping constraints in the optimization process in nature. Comparing with many other relaxation methods, which aim to find the optimal solution in the continuous domain, our method can be regarded as optimizing the IQP problem in an approximate discrete domain and thus its optimal solution can be closer to the discrete optimum of original IQP problem.

The remainder of this paper is organized as follows. In Section II, we introduce the general formulation of feature matching problem as an IQP problem. In Section III, a general nonnegative quadratic programming sparse model is proposed, followed by an efficient algorithm to solve this model. In Section IV, we propose our sparse feature matching method (SPM) based on the proposed quadratic sparse model. In Section V, some benefits of the proposed model are demonstrated. In Section VI, we apply our method to some feature matching tasks on both synthetic point sets data and real-world image feature sets data.

II. Problem Formulation and Related Work

A. Problem Formulation

Given two sets of features M, containing n_M model features and D, containing n_D data features, a corresponding mapping is a set C of pairs (or assignments) (i, i'), where $i \in D$ and $i' \in M$. Some kinds of mapping constraints are usually imposed on C, such as: allowing one data feature from D to match at most one model feature from M and vice-versa (one-to-one mapping), or allowing one feature from one set to match more features from the other (one-to-many). If the features are discriminative, such as SIFT or shape context descriptors, then there is a score $S_u(a)$ for each assignment $a = (i, i')$ that measures how well the feature $i \in D$ matches the feature $i' \in M$. Also, for each pair of assignments (a, b), where $a = (i, i')$ and $b = (j, j')$, there is an affinity $S_r(a, b)$ that measures how compatible the feature pair (i, j) in data feature D are with the features (i', j') in model feature M. When the features are discriminative, then both score $S_u(a)$ and affinity $S_r(a, b)$ exist. However, when the features are non-discriminative, such as 2D or 3D points, we can also use shape context descriptor to discriminate the features. Therefore, the objective of feature point matching is to find the optimal mapping C^* between data features D and model features M that maximizes the following matching score,

$$C^* = \arg\max_C \sum_{a\in C} \sum_{b\in C} S_r(a, b) + \sum_{a\in C} S_u(a). \quad (1)$$

Note that if the features are non-discriminative, then $\mathrm{S}_u(a)$ can be set to 0 in this objective function. In this paper, the one-to-one mapping constraints are imposed on C. The mapping C can be represented by a permutation matrix $\mathbf{X}$ such that $\mathbf{X}_{ii'} = 1$ implies that feature i in D corresponds to feature i' in M, and $\mathbf{X}_{ii'} = 0$ otherwise. We denote $\mathbf{x} \in \{0, 1\}^{mn}$ as a row-wise vectorized replica of $\mathbf{X}$, i.e., $\mathbf{x} = (\mathbf{X}_{11} \ldots \mathbf{X}_{1n} \ldots \mathbf{X}_{m1} \ldots \mathbf{X}_{mn})^{\mathrm{T}}$, where $m = |D|$, $n = |M|$. The feature point matching problem (1) can be generally formulated as an Integer Quadratic Programming (IQP) problem, i.e., finding the indicator vector $\mathbf{x}^*$ that maximizes the score function as

$$\begin{aligned} &\mathbf{x}^* = \arg\max_{\mathbf{x}} \varepsilon_0(\mathbf{x}) = \mathbf{x}^{\mathrm{T}}\mathbf{W}\mathbf{x} + \mathbf{x}^{\mathrm{T}}\mathbf{S} \\ &\text{s.t.} \quad \mathbf{x}_{ii'} \in \{0, 1\}, \quad \forall i \sum_{i'=1}^{n} \mathbf{x}_{ii'} \le 1, \quad \forall j' \sum_{j=1}^{n} \mathbf{x}_{jj'} \le 1. \end{aligned} \quad (2)$$

$\mathbf{W}$ is a $mn \times mn$ affinity matrix with the non-diagonal element $\mathbf{W}_{a,a} = S_r(a, b)$ and the diagonal term $\mathbf{W}_{a,a} = 0$. $\mathbf{S}$ is $mn \times 1$ score matrix with the element $S_a = S_u(a)$. The two-way constraints in (2) guarantee the one-to-one matching constraints between D and M. If the features are non-discriminative, then $\mathbf{S}$ can be set to $\mathbf{0}$ and (2) becomes

$$\begin{aligned} &\mathbf{x}^* = \arg\max_{\mathbf{x}} \varepsilon_0(\mathbf{x}) = \mathbf{x}^{\mathrm{T}}\mathbf{W}\mathbf{x} \\ &\text{s.t.} \quad \mathbf{x}_{ii'} \in \{0, 1\}, \quad \forall i \sum_{i'=1}^{n} \mathbf{x}_{ii'} \le 1, \quad \forall j' \sum_{j=1}^{n} \mathbf{x}_{jj'} \le 1. \end{aligned} \quad (3)$$

B. Relaxation Algorithms

It is well known that, the problems (2, 3) are NP-hard and no efficient algorithm exists. Thus lots of approximate algorithms were proposed to find the approximate solutions [1], [6], [8], [16], [17], [37], [40]. In general, from the optimization perspective, a practical relaxation algorithm should have high *approximation ability*, i.e., it can approximate the original IQP problem as closely as possible. Specifically, it must satisfy the following two desired matching properties [1], [16], [17], [29]: (1) it maximizes the objective score as far as possible; (2) It satisfies the mapping constraints as closely as possible. In this paper, we call these properties as *objective property* and *constraint property*, respectively. These two properties have also been adopted in the work [1]. Usually, the approximate algorithms cannot guarantee that the solution satisfies the constraints strictly, and they obtain the final discrete correspondence solution using some discretization techniques which usually lead to weak local optimum for the original IQP problem. Leordeanu and Hebert [7] proposed a spectral technique (SM) to feature correspondence problems. This method can find the global maximum solution of the relaxed problem effectively and thus has strong objective property. However, the method does not satisfy the constraint property sufficiently because of the relaxation [1]. Leordeanu et al. [17] also recently proposed an iterative matching method (IPFP) which integrates the objective function optimization and discretization step simultaneously. This method optimizes the IQP in the discrete domain and thus satisfies the constraint property strictly. Cour et al. [16] extended SM to spectral matching with affine constraints (SMAC) by incorporating the matching constraints within the relaxation process. Comparing with SM, this method satisfies the constraint property more sufficiently and therefore can obtain more effective solution for the matching problem.

III. Nonnegative Quadratic Programming With Sparse Constraint

In this section, we study the general problem of nonnegative quadratic programming with sparse constraint, which is formulated as follows,

$$\max_{\mathbf{x}} \varepsilon_1(\mathbf{x}) = \mathbf{x}^{\mathrm{T}}\mathbf{W}\mathbf{x} + \mathbf{x}^{\mathrm{T}}\mathbf{S} \quad \text{s.t.} \ \ ||\mathbf{x}||_1 = 1, \quad \mathbf{x}_i \geq 0 \tag{4}$$

where $||\mathbf{x}||_1 = \sum_{i=1}^{mn} |\mathbf{x}_i|$ denotes the L_1 norm. This optimization can be explained as a problem subject to a L_1 norm constraint on the solution. One main feature for this kind of problems is that they can encourage sparse solutions [20]–[24], i.e., many components of the solution $\mathbf{x}$ are zero (close-to-zero). Note that, when the coefficients $\mathbf{W}_{ij}$ and $\mathbf{S}_i$ are all nonnegative, this problem has been widely studied in the works [34]–[36]. Here, we concern a more general case, i.e., with the presence of both positive and negative coefficients.

A. Computational Algorithm

Inspired by recent processes on multiplicative update based algorithms [25]–[28], in this paper we develop a more general algorithm to solve the problem in (4). There are two main aspects of the proposed algorithm: (i) it can naturally deal with the general polynomials with both positive and negative coefficients. (ii) It is extremely simple to implement and its convergence is guaranteed.

1) Update Algorithm: For any matrix $\mathbf{A}$ with both positive and negative elements, denote $\mathbf{A}^+ = (|\mathbf{A}| + \mathbf{A})/2, \mathbf{A}^- = (|\mathbf{A}| - \mathbf{A})/2$. Since $\mathbf{x}_i \geq 0$, problem (4) is equivalent to the following,

$$\max_{\mathbf{x}} \ \varepsilon_1(\mathbf{x}) = \mathbf{x}^{\mathrm{T}}\mathbf{W}\mathbf{x} + \mathbf{x}^{\mathrm{T}}\mathbf{S} \ \ \text{s.t.} \ \sum_{i=1}^{mn} \mathbf{x}_i = 1, \quad \mathbf{x}_i \geq 0. \tag{5}$$

This problem can be efficiently solved by an iterative algorithm. The algorithm iteratively updates a current solution $\mathbf{x}^{(t)}$ as,

$$\mathbf{x}_i^{(t+1)} = \mathbf{x}_i^{(t)} \sqrt{\frac{2(\mathbf{W}^+\mathbf{x}^{(t)})_i + \mathbf{S}_i^+ + 2(\mathbf{x}^{(t)})^{\mathrm{T}}\mathbf{W}^-\mathbf{x}^{(t)} + \mathbf{S}^{-\mathrm{T}}\mathbf{x}^{(t)}}{2(\mathbf{W}^-\mathbf{x}^{(t)})_i + \mathbf{S}_i^- + 2(\mathbf{x}^{(t)})^{\mathrm{T}}\mathbf{W}^+\mathbf{x}^{(t)} + \mathbf{S}^{+T}\mathbf{x}^{(t)}}}. \tag{6}$$

The iteration starts with an initial $\mathbf{x}^{(0)}$ and is repeated until convergence. Note that, when the coefficients are all positive, the update (6) degenerates to the following,

$$\mathbf{x}_i^{(t+1)} = \mathbf{x}_i^{(t)} \sqrt{\frac{2(\mathbf{W}\mathbf{x}^{(t)})_i + \mathbf{S}_i}{2(\mathbf{x}^{(t)})^{\mathrm{T}}\mathbf{W}\mathbf{x}^{(t)} + \mathbf{S}^{\mathrm{T}}\mathbf{x}^{(t)}}}. \tag{7}$$

Indeed, from the multiplicative perspective, we can also deduce the following update rule when the coefficients are all positive,

$$\mathbf{x}_i^{(t+1)} = \mathbf{x}_i^{(t)} \frac{2(\mathbf{W}\mathbf{x}^{(t)})_i + \mathbf{S}_i}{2(\mathbf{x}^{(t)})^{\mathrm{T}}\mathbf{W}\mathbf{x}^{(t)} + \mathbf{S}^{\mathrm{T}}\mathbf{x}^{(t)}}. \tag{8}$$

We will show in Experimental section that the update algorithm (7) generally performs slightly better than algorithm (8) on conducting feature matching tasks.

Note that, the update algorithm (8) can also be derived from the nonlinear growth transformations proposed in the works [36], [37], which are also known as Baum-Eagon theorem or Baum-Eagon inequality. This theorem provides an effective iterative method for maximizing the polynomial functions with positive coefficients. Leordeanu and Hebert [34] also proposed a general update framework for maximizing the polynomials functions with positive coefficients. To the best of our knowledge, these algorithms [34]–[37] cannot be directly used for the polynomials functions with the presence of negative coefficients, although it may be possible (under some special constraints) to transform these polynomials to the polynomials with positive coefficients which does not change the original global solutions [34], [37]. From the algorithm perspective, the proposed algorithm (6) can directly be used for the polynomials with the presence of negative coefficients, and thus can be regarded as an extension of the prior works. In the following, we will present a theoretical proof on the correctness and convergence of the proposed update algorithms (6) and (8) from multiplicative perspective. This proof can also be regarded as a new justification for those Baum-Eagon theorem inspired algorithms [34], [36], [37].

B. Analysis of the Algorithm

We show the correctness and convergence of the proposed algorithm. For correctness, we show that the update algorithm yields a correct solution at convergence. Since update (7) is a special case of update (6), here we only discuss the update (6) and (8). The correctness of these two algorithms is guaranteed by the following theorem.

Theorem 1: Both of update rule of (6) and (8) satisfy the first-order Karush-Kuhn-Tucker (KKT) optimality condition.

Proof: The Lagrangian function is

$$L(\mathbf{x}) = \mathbf{x}^{\mathrm{T}}\mathbf{W}\mathbf{x} + \mathbf{x}^{\mathrm{T}}\mathbf{S} - \lambda(\sum_{i=1}^{mn} \mathbf{x}_i - 1) \tag{9}$$

where Lagrangian multiplier λ enforces $\sum_{i=1}^{mn} \mathbf{x}_i = 1$. Then,

$$\frac{\partial L}{\partial \mathbf{x}_i} = 2(\mathbf{W}\mathbf{x})_i + \mathbf{S}_i - \lambda. \tag{10}$$

This leads to KKT complementary slackness condition,

$$\frac{\partial L}{\partial \mathbf{x}_i}\mathbf{x}_i = [2(\mathbf{W}\mathbf{x})_i + \mathbf{S}_i - \lambda]\,\mathbf{x}_i = 0. \tag{11}$$

Summing over index i, we obtain Lagrangian multiplier λ as,

$$\lambda = 2\mathbf{x}^{\mathrm{T}}\mathbf{W}\mathbf{x} + \mathbf{x}^{\mathrm{T}}\mathbf{S}. \tag{12}$$

Firstly, for update (6), at convergence,

$$\mathbf{x}_i^* = \mathbf{x}_i^* \sqrt{\frac{2(\mathbf{W}^+\mathbf{x}^*)_i + \mathbf{S}_i^+ + 2(\mathbf{x}^*)^{\mathrm{T}}\mathbf{W}^-\mathbf{x}^* + \mathbf{S}^{-\mathrm{T}}\mathbf{x}^*}{2(\mathbf{W}^-\mathbf{x}^*)_i + \mathbf{S}_i^- + 2(\mathbf{x}^*)^{\mathrm{T}}\mathbf{W}^+\mathbf{x}^* + \mathbf{S}^{+T}\mathbf{x}^*}} \tag{13}$$

It is equivalent to (note that $\mathbf{W} = \mathbf{W}^+ - \mathbf{W}^-$, $\mathbf{S} = \mathbf{S}^+ - \mathbf{S}^-$)

$$\Big(2(\mathbf{W}\mathbf{x}^*)_i + \mathbf{S}_i - 2(\mathbf{x}^*)^{\mathrm{T}}\mathbf{W}\mathbf{x}^* + \mathbf{S}^{\mathrm{T}}\mathbf{x}^*\Big)\,\mathbf{x}_i^{*2} = 0 \tag{14}$$

which is identical to

$$\Big(2(\mathbf{W}\mathbf{x}^*)_i + \mathbf{S}_i - 2(\mathbf{x}^*)^{\mathrm{T}}\mathbf{W}\mathbf{x}^* + \mathbf{S}^{\mathrm{T}}\mathbf{x}^*\Big)\, \mathbf{x}_i^* = 0 \qquad (15)$$

This is exactly the KKT condition (11).

Then, for update (8), at convergence

$$\mathbf{x}_i^* = \mathbf{x}_i^* \frac{2(\mathbf{W}\mathbf{x}^*)_i + \mathbf{S}_i}{2(\mathbf{x}^*)^{\mathrm{T}}\mathbf{W}\mathbf{x}^* + \mathbf{S}^{\mathrm{T}}\mathbf{x}^*} \qquad (16)$$

It is equivalent to $(2(\mathbf{W}\mathbf{x}^*)_i + \mathbf{S}_i - 2(\mathbf{x}^*)^{\mathrm{T}}\mathbf{W}\mathbf{x}^* + \mathbf{S}^{\mathrm{T}}\mathbf{x}^*)\mathbf{x}_i^* = 0$ which is the KKT condition (11). Thus, the update (8) also satisfies the KKT condition at convergence.

The convergence is guaranteed by the following Theorem 2.

Theorem 2: (1) The Lagrangian function $L(\mathbf{x})$ of (9) is monotonically increasing under the update rule in (6); (2) When the coefficients are all nonnegative, the Lagrangian $L(\mathbf{x})$ of (9) is monotonically increasing under the update rule in (8).

Proof: (1) We use the auxiliary function approach which has been widely used in many multiplicative inspired algorithms [20], [23], [26]–[28]. An auxiliary function $Z(\mathbf{x}, \tilde{\mathbf{x}})$ of function $L(\mathbf{x})$ satisfies $Z(\mathbf{x}, \mathbf{x}) = L(\mathbf{x})$ and $Z(\mathbf{x}, \tilde{\mathbf{x}}) \le L(\mathbf{x})$.

We define

$$\mathbf{x}^{(t+1)} = \arg\max_{\mathbf{x}} Z(\mathbf{x}, \mathbf{x}^{(t)}). \qquad (17)$$

Then by construction, we have

$$L(\mathbf{x}^{(t)}) = Z(\mathbf{x}^{(t)}, \mathbf{x}^{(t)}) \le Z(\mathbf{x}^{(t+1)}, \mathbf{x}^{(t)}) \le L(\mathbf{x}^{(t+1)}) \quad (18)$$

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

In the remainder of proof, we need: (1) find an appropriate auxiliary function; (2) Find the global maximum of the auxiliary function. We rewrite the Lagrangian function (9) as

$$\begin{aligned} L(\mathbf{x}) = {} & \sum\nolimits_{i,j=1}^{mn} \mathbf{W}_{ij}^{+}\mathbf{x}_i\mathbf{x}_j + \sum\nolimits_{i=1}^{mn} \mathbf{S}_i^{+}\mathbf{x}_i \\ & - \sum\nolimits_{i,j=1}^{mn} \mathbf{W}_{ij}^{-}\mathbf{x}_i\mathbf{x}_j - \sum\nolimits_{i=1}^{mn} \mathbf{S}_i^{-}\mathbf{x}_i \\ & - \lambda\Big(\sum\nolimits_{i=1}^{mn} \mathbf{x}_i - 1\Big) \end{aligned} \qquad (19)$$

We can show that one auxiliary function of $L\ (\mathbf{x})$ is

$$\begin{aligned} & Z(\mathbf{x}, \tilde{\mathbf{x}}) \\ & = \sum\nolimits_{i=1}^{mn}\sum\nolimits_{j=1}^{mn} \mathbf{W}_{ij}^{+}\tilde{\mathbf{x}}_i\tilde{\mathbf{x}}_j\left(1 + \log\frac{\mathbf{x}_i\mathbf{x}_j}{\tilde{\mathbf{x}}_i\tilde{\mathbf{x}}_j}\right) \\ & \quad - \sum\nolimits_{i=1}^{mn} \frac{(\mathbf{W}^{-}\tilde{\mathbf{x}})_i\mathbf{x}_i^2}{\tilde{\mathbf{x}}_i} + \sum\nolimits_{i=1}^{mn} \mathbf{S}_i^{+}\tilde{\mathbf{x}}_i\left(1 + \log\frac{\mathbf{x}_i}{\tilde{\mathbf{x}}_i}\right) \\ & \quad - \sum\nolimits_{i=1}^{mn} \frac{1}{2}\mathbf{S}_i^{-}\left(\frac{\mathbf{x}_i^2}{\tilde{\mathbf{x}}_i} + \tilde{\mathbf{x}}_i\right) - \lambda^{+}\left(\sum\nolimits_{i=1}^{mn} \frac{1}{2}\left(\frac{\mathbf{x}_i^2}{\tilde{\mathbf{x}}_i} + \tilde{\mathbf{x}}_i\right) - 1\right) \\ & \quad + \lambda^{-}\left(\sum\nolimits_{i=1}^{mn} \tilde{\mathbf{x}}_i\left(1 + \log\frac{\mathbf{x}_i}{\tilde{\mathbf{x}}_i}\right) - 1\right) \end{aligned} \qquad (20)$$

where $\lambda = \lambda^{+} - \lambda^{-}, \lambda^{+} \ge 0, \lambda^{-} \ge 0$.

Using the inequality $z \ge 1 + \log z$ and $a \le \frac{1}{2}\left(\frac{a^2}{b} + b\right)$, the first two terms in (20) is a lower bound of the first two terms in (19). Thus, $Z(\mathbf{x}, \tilde{\mathbf{x}})$ is an auxiliary function of $L(\mathbf{x})$.

According to (17), we need to find the global maximum of $Z(\mathbf{x}, \tilde{\mathbf{x}})$ for $\mathbf{x}$. The gradient is

$$\begin{aligned} \frac{\partial Z(\mathbf{x}, \tilde{\mathbf{x}})}{\partial \mathbf{x}_i} = {} & \left(2(\mathbf{W}^{+}\tilde{\mathbf{x}})_i + \mathbf{S}_i^{+} + \lambda^{-}\right)\frac{\tilde{\mathbf{x}}_i}{\mathbf{x}_i} \\ & - \left(2(\mathbf{W}^{-}\tilde{\mathbf{x}})_i + \mathbf{S}_i^{-} + \lambda^{+}\right)\frac{\mathbf{x}_i}{\tilde{\mathbf{x}}_i}. \end{aligned} \qquad (21)$$

The second derivative is

$$\begin{aligned} \frac{\partial^2 Z(\mathbf{x}, \tilde{\mathbf{x}})}{\partial \mathbf{x}_i \partial \mathbf{x}_j} = - \Bigg[& \left(2(\mathbf{W}^{+}\tilde{\mathbf{x}})_i + \mathbf{S}_i^{+} + \lambda^{-}\right)\frac{\tilde{\mathbf{x}}_i}{\mathbf{x}_i^2} \\ & + \left(2(\mathbf{W}^{-}\tilde{\mathbf{x}})_i + \mathbf{S}_i^{-} + \lambda^{+}\right)\frac{1}{\tilde{\mathbf{x}}_i}\Bigg]\delta_{ij} \end{aligned} \qquad (22)$$

where $\delta_{ij} = 1$ if $i = j$, otherwise $\delta_{ij} = 0$. Thus $Z(\mathbf{x}, \tilde{\mathbf{x}})$ is a concave function and has a unique global maximum, which is obtained by setting the first derivative (21) to zero, i.e.,

$$\mathbf{x}_i = \tilde{\mathbf{x}}_i\sqrt{\frac{2(\mathbf{W}^{+}\tilde{\mathbf{x}})_i + \mathbf{S}_i^{+} + \lambda^{-}}{2(\mathbf{W}^{-}\tilde{\mathbf{x}})_i + \mathbf{S}_i^{-} + \lambda^{+}}}. \qquad (23)$$

Thus, we obtain the update (6) by setting $\mathbf{x}^{(t+1)} = \mathbf{x}, \mathbf{x}^{(t)} = \tilde{\mathbf{x}}$.

(2) When the coefficients are all positive, we can rewrite the Lagrangian function (9) as,

$$L(\mathbf{x}) = \mathbf{x}^{\mathrm{T}}\mathbf{W}\mathbf{x} + \mathbf{x}^{\mathrm{T}}\mathbf{S} - \lambda\Big(\sum\nolimits_{i=1}^{mn} \mathbf{x}_i - 1\Big) \qquad (24)$$

Similarly, we derive the following auxiliary function,

$$\begin{aligned} Z(\mathbf{x}, \tilde{\mathbf{x}}) = {} & \sum\nolimits_{i=1}^{mn}\sum\nolimits_{j=1}^{mn} \mathbf{W}_{ij}\tilde{\mathbf{x}}_i\tilde{\mathbf{x}}_j\Big(1 + \log\frac{\mathbf{x}_i\mathbf{x}_j}{\tilde{\mathbf{x}}_i\tilde{\mathbf{x}}_j}\Big) \\ & + \sum\nolimits_{i=1}^{mn} \mathbf{S}_i\tilde{\mathbf{x}}_i\Big(1 + \log\frac{\mathbf{x}_i}{\tilde{\mathbf{x}}_i}\Big) - \lambda\Big(\sum\nolimits_{i=1}^{mn} \mathbf{x}_i - 1\Big). \end{aligned} \qquad (25)$$

We need to find the maximum of $Z(\mathbf{x}, \tilde{\mathbf{x}})$. The gradient is

$$\frac{\partial Z(\mathbf{x}, \tilde{\mathbf{x}})}{\partial \mathbf{x}_i} = 2(\mathbf{W}\tilde{\mathbf{x}})_i\frac{\tilde{\mathbf{x}}_i}{\mathbf{x}_i} - \lambda + \frac{\tilde{\mathbf{x}}_i\mathbf{S}_i}{\mathbf{x}_i}. \qquad (26)$$

The second derivative is

$$\frac{\partial^2 Z(\mathbf{x}, \tilde{\mathbf{x}})}{\partial \mathbf{x}_i \partial \mathbf{x}_j} = -\left[2\frac{(\mathbf{W}\tilde{\mathbf{x}})_i\tilde{\mathbf{x}}_i}{\mathbf{x}_i^2} + \frac{\tilde{\mathbf{x}}_i\mathbf{S}_i}{\mathbf{x}_i^2}\right]\delta_{ij}, \qquad (27)$$

Thus $Z(\mathbf{x}, \tilde{\mathbf{x}})$ is a concave function of $\mathbf{x}$ and has a unique global maximum, which is obtained as

$$\mathbf{x}_i = \frac{2(\mathbf{W}\tilde{\mathbf{x}})_i\tilde{\mathbf{x}}_i + \tilde{\mathbf{x}}_i\mathbf{S}_i}{\lambda} = \tilde{\mathbf{x}}_i\frac{2(\mathbf{W}\tilde{\mathbf{x}})_i + \mathbf{S}_i}{\lambda}. \qquad (28)$$

Thus, we obtain the update (8) by setting $\mathbf{x}^{(t+1)} = \mathbf{x}, \mathbf{x}^{(t)} = \tilde{\mathbf{x}}$.

IV. Nonnegative Quadratic Sparse Model for Feature Point Matching

The optimal solution for the original IQP matching problem is discrete, binary and thus sparse in nature, i.e., there exists small number of positive nonzero elements in the optimal solution. This motivates us to use sparse model for feature matching problem. Based on the above nonnegative quadratic sparse model, we present our sparse feature point matching (SPM) model in this section.

A. Sparse Feature Point Matching Model

By adding a L_1 norm on the constraint, our sparse feature point matching (SPM) model can be achieved as solving the following problem,

$$\max_{\mathbf{x}} \varepsilon_1(\mathbf{x}) = \sum_{i,j,i',j'=1}^{mn} \mathbf{W}_{ii',jj'}\mathbf{X}_{ii'}\mathbf{X}_{jj'} + \sum_{i,i'=1}^{mn} \mathbf{S}_{ii'}\mathbf{X}_{ii'} \\ = \mathbf{x}^{\mathrm{T}}\mathbf{W}\mathbf{x} + \mathbf{x}^{\mathrm{T}}\mathbf{S} \\ \text{s.t.} \quad ||\mathbf{x}||_1 = 1, \mathbf{x}_i \geq 0 \tag{29}$$

where $\mathbf{x} = (\mathbf{X}_{11} \ldots \mathbf{X}_{1n\ldots}\mathbf{X}_{m1} \ldots \mathbf{X}_{mn})^{\mathrm{T}}$. Here $\mathbf{W}$ is the affinity matrix with the non-diagonal elements containing pairwise affinity between two assignments and the diagonal elements are zero. $\mathbf{S}$ is the matching score matrix. In general, both $\mathbf{W}$ and $\mathbf{S}$ are nonnegative. However, in order to enforce one-to-one matching constraint in optimization process, the affinity $\mathbf{W}_{ii',ij'}$ between two conflict assignments $a = (i, i')$, $b = (i, j')(j \neq j')$ can be penalized to enforce $\mathbf{X}_{ii'}\mathbf{X}_{ij'}$ to be small, because at least one of assignment a and b must be incorrect. In most literature [7], [13], $\mathbf{W}_{ii',ij'}$ is set to zero. Although $\mathbf{W}$ is nonnegative in this case, $\mathbf{X}_{ii'}\mathbf{X}_{ij'}$ is not penalized indeed. Because if $\mathbf{W}_{ii',ij'} = 0$, then the objective term $\mathbf{W}_{ii',ij'}\mathbf{X}_{ii'}\mathbf{X}_{ij'}$ is ignored in the matching objective. Therefore, it is desirable to penalize the affinity $\mathbf{W}_{ii',ij'}$ with a negative value, i.e., $\mathbf{W}_{ii',ij'} = w$ $(w < 0)$. In this case, $\mathbf{W}$ contains both nonnegative and positive elements.

The above optimization can be explained as a problem subject to a L_1 norm constraint on the solution. The main feature for this kind of problems is that they can encourage sparse solutions [20]–[24], i.e., many components of the solution $\mathbf{x}$ are zero (close-to-zero). There are three main aspects for our SPM method. (1) SPM generates sparse solution and thus approximately incorporates the discrete mapping constraints in nature, i.e., it can return an approximate discrete solution and satisfies the constraint property sufficiently. (2) An efficient update algorithm can be derived to solve SPM optimization problem. (3) By enforcing the solution to be sparse in the optimization process, SPM optimizes the problem in an approximate discrete domain, and thus maximizes the objective score effectively. Therefore, SPM approximates the original IQP problem closely and thus leads to an effective solution for feature point matching problem. Note that, when $\mathbf{S} = \mathbf{0}$, this problem has also been adopted in the works [23], [38]–[41]. These works generally discuss the model from the game-theoretic perspective. Although the sparse property has been mentioned in the works [39], [41], it has not been explained and demonstrated in detail. Different from these works, in this paper we focus on a more general matching model (29) which can be regarded as an extension of the prior works [38]–[41], and derive a new efficient multiplicative update algorithm (update (6) or (8)) to solve this general problem. Moreover, we discuss this model from sparse model perspective, and will show experimentally the strong relationship between the sparsity of the solution and its effectiveness for the feature matching problem. Also, we conduct thorough experiments and compare SPM with some recent works on feature matching tasks on both synthetic and real world image data.

B. Matching Algorithm

As discussed in Section III, two update algorithms (update (6) and (8)) can be used for finding the optimal solution of the proposed SPM model. In the following, we denote these two algorithms as SPM-G and SPM-M, respectively. Also, when the affinity matrix $\mathbf{W}$ contains negative elements, we use the update (6) to find the optimal solution. We call it as SPM-P in the following. Experimentally, both SPM-G and SPM-M can generate similar optimal solutions and SPM-M usually performs slightly better than SPM-G on conducting feature matching tasks (see Experimental section in detail). Usually, problem (29) is not always convex and the final results depend on initializations. Here, we compute the initialization as $\mathbf{x}_i^{(0)} = \mathbf{v}_i(\sum_{i=1}^{mn} \mathbf{v}_i)^{-1}$, where $\mathbf{v}$ is the principal eigenvector of $\mathbf{W}$. Indeed, the eigenvector $\mathbf{v}$ of $\mathbf{W}$ is the global optimum of the SM method [7], which aims to solve the following problem,

$$\max_{\mathbf{x}} \quad \mathbf{x}^{\mathrm{T}}\mathbf{W}\mathbf{x} \quad s.t. \quad ||\mathbf{x}||_2 = 1, \tag{30}$$

where $||\mathbf{x}||_2 = (\sum_{i=1}^{mn} \mathbf{x}_i^2)^{1/2}$. By the Raleigh's ratio theorem, the optimal solution $\mathbf{x}^*$ of (30) is given by the principal eigenvector of $\mathbf{W}$ ($\mathbf{x}^* = \mathbf{v}$). If $\mathbf{W}$ has non-negative elements, by Perron-Frobenius theorem [7], the elements of $\mathbf{x}^*$ will be in the interval [0, 1]. This guaranteed the non-negativity of the initialization $\mathbf{x}^{(0)}$. Also, the optimal solution $\mathbf{x}^*$ of SM method [7] can be regarded as a global optimal solution for the matching problem (3) under the relaxed constraints ($||\mathbf{x}||_2 = 1$). Indeed, Baratchart et al. [35] proposed a general relaxation method for IQP problem by relaxing the integral constraint to the L_2 norm constraint, and show some theoretical results on the global optimality of this relaxed problem. In this paper, we focus on the L_1 norm instead of L_2 norm, because it leads to sparse solution which will be shown very important for conducting matching task although the global optimality property cannot be guaranteed in this case. We use the global optimal solution of the L_2 norm relaxed problem as the initialization for our SPM method. Therefore, the initialization vector $\mathbf{x}^{(0)}$ is close to the optimal solution of the original IQP matching problem and thus gives a good starting point for the proposed SPM. Similar idea has also been proposed in the work [34], which provided a two-stage optimization method for MAP problem. Indeed, this two-stage optimization can also be found in some other problems, such as spectral clustering involving both spectral embedding and k-means stages [42], [43]. Our SPM (SPM-G) matching method is summarized as follows. SPM-M and SPM-P are similarly obtained by replacing the update in Step 4 as the update (7) and (6).

V. Sparsity and Desirable Approximation Property

The main difference between SPM and many other relaxed methods is that a L1 norm constraint is imposed on the

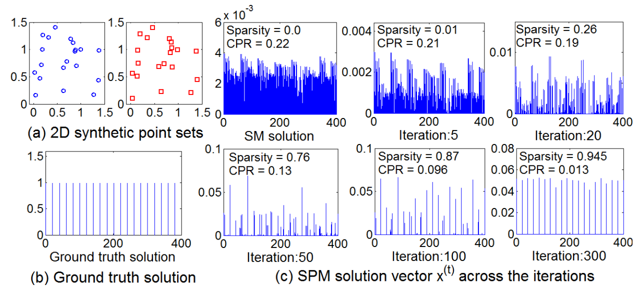


Fig. 1. Solution vector for SM and SPM (SPM-G); (a) Two 2D example point sets to be matched; (b) Ground truth solution vector; (c) Solution vector $\mathbf{x}^{(t)}$ of SPM across different iterations. Note that $\mathbf{x}^{(t)}$ becomes more and more sparse and approximates the ground truth solution more and more closely.

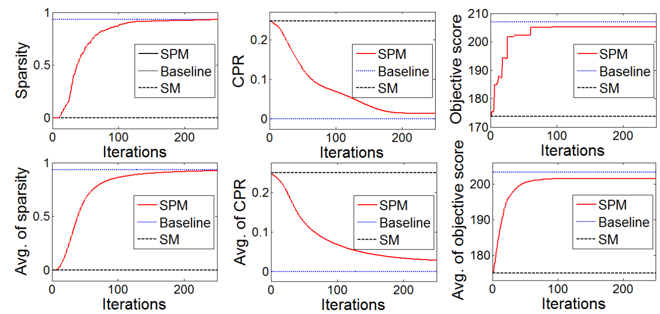


Fig. 2. Top row: sparsity, CPR and objective score of the solution vector $\mathbf{x}^{(t)}$ across the iterations on synthetic feature point matching. Bottom row: at each iteration, average results over 100 random experiments are displayed. Note that as the iteration increases, the solution of SPM approximates the optimum of original IQP problem (baseline) more closely than SM method.

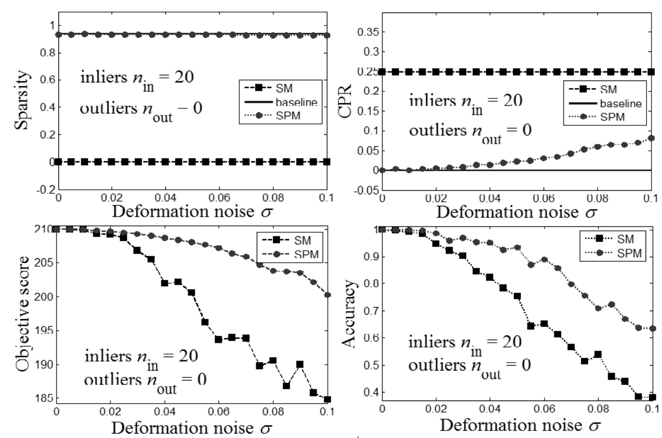


Fig. 3. Performance curves for our SPM vs. SM method on synthetic data matching; Note that SPM significantly outperforms SM in both objective score and matching accuracy.

solution $\mathbf{x}$ and thus encourages a sparse solution [20]–[24]. Here we show experimentally that, by enforcing the solution to be sparse in the optimization process, SPM optimizes the IQP matching problem in an approximate discrete domain and searches for an approximate discrete solution for the problem, i.e., it satisfies both constraint and objective properties approximately and strongly. Therefore, SPM approximates the IQP problem closely and thus attains effective solution for feature matching problem. For further illustration, the sparsity, objective score and constraint preserving residual are first defined. Define the function $f(\mathbf{x}_i)\,(i = 1, 2, \dots mn)$ for vector $\mathbf{x}$ as follows:

$$f(\mathbf{x}_i) = \begin{cases} 0 & \text{if } \mathbf{x}_i < 0.001 \times \text{mean}(\mathbf{x}) \\ \mathbf{x}_i & \text{otherwise} \end{cases} \tag{31}$$

where mean($\mathbf{x}$) denotes the mean value of elements in vector $\mathbf{x}$. Let $f(\mathbf{x})$ denote $f(\mathbf{x}) = (f(\mathbf{x}_1), f(\mathbf{x}_2) \dots f(\mathbf{x}_{mn}))$. Let $\mathbf{x}^*$ be the convergence solution of the relaxation algorithm. Let $\tilde{\mathbf{x}}^*$ be the final binarized discrete solution ($\tilde{\mathbf{x}}_i^* \in \{0, 1\}$) obtained by performing some discretization processes.

Definition 1 (Sparsity): The sparsity of $\mathbf{x}^*$ ($\mathbf{x}_i^* \geq 0$) measures the rate of zero (close-to-zero) elements in $\mathbf{x}^*$. It is calculated as

$$\text{Sparsity}(\mathbf{x}^*) = 1 - \tfrac{1}{N}||f(\mathbf{x}^*)||_0, \tag{32}$$

where $f(\cdot)$ is defined in (31), and $||\cdot||_0$ denotes the L_0 norm, which counts the number of nonzero entries in a vector. N is the number of entries in vector $\mathbf{x}^*$.

Definition 2 (Objective Score, OS): The objective score of $\mathbf{x}^*$ is defined as

$$\text{OS}(\mathbf{x}^*) = \tilde{\mathbf{x}}^{*\mathrm{T}}\mathbf{W}\tilde{\mathbf{x}}^*, \tag{33}$$

where $\tilde{\mathbf{x}}^*$ is the final binarized discrete solution ($\tilde{\mathbf{x}}_i^* \in \{0, 1\}$) obtained by performing some discretization processes.

Definition 3 (Constraint Preserving Residual, CPR): The constraint preserving residual of $\mathbf{x}^*$ is defined as

$$\text{CPR}(\mathbf{x}^*) = \min_{\beta} \frac{1}{||\tilde{\mathbf{x}}^*||_0}||\tilde{\mathbf{x}}^* - \beta\mathbf{x}^*|| \tag{34}$$

where β is a weighting parameter to compensate the loss of residual due to scaling.

As discussed in Section II, the above objective score and CPR can be regarded as measurements for the objective and constraint properties, respectively. The higher the objective score, the more optimal the solution of the original IQP problem can be. Meanwhile, the lower the CPR value, the more closely the solution satisfies the discrete mapping constraint. Especially, if CPR ($\mathbf{x}^*$) $= 0$, then the solution $\mathbf{x}^*$ satisfies the discrete mapping constraints strictly. In the following, we only show the sparse property of SPM solution under SPM-G algorithm. This property can also be obtained by SPM-M algorithm. Fig. 1 shows the solution $\mathbf{x}^{(t)}$ across different iterations on synthetic 2D point sets matching with deformation noise level $\sigma = 0.05$ (see the Experimental section in detail). Intuitively, as the iteration increases, the solution of SPM becomes more and more sparse and approximates the ground truth solution more and more closely. Fig. 2 shows the sparsity, CPR and objective score of the solution $\mathbf{x}^{(t)}$ across different iterations. The average values are computed by generating 100 random point sets under the deformation noise level $\sigma = 0.05$ (see the Experimental section). The baselines denote sparsity, CPR and objective score of ground truth solution. Here, we make the following observations. (1) As the iteration increases, $\mathbf{x}^{(t)}$ becomes more and more sparse in general and approximates the baseline more and more closely, which indicates the ability of the SPM algorithm (SPM-G) to maintain the sparse constraint. (2) The CPR curve

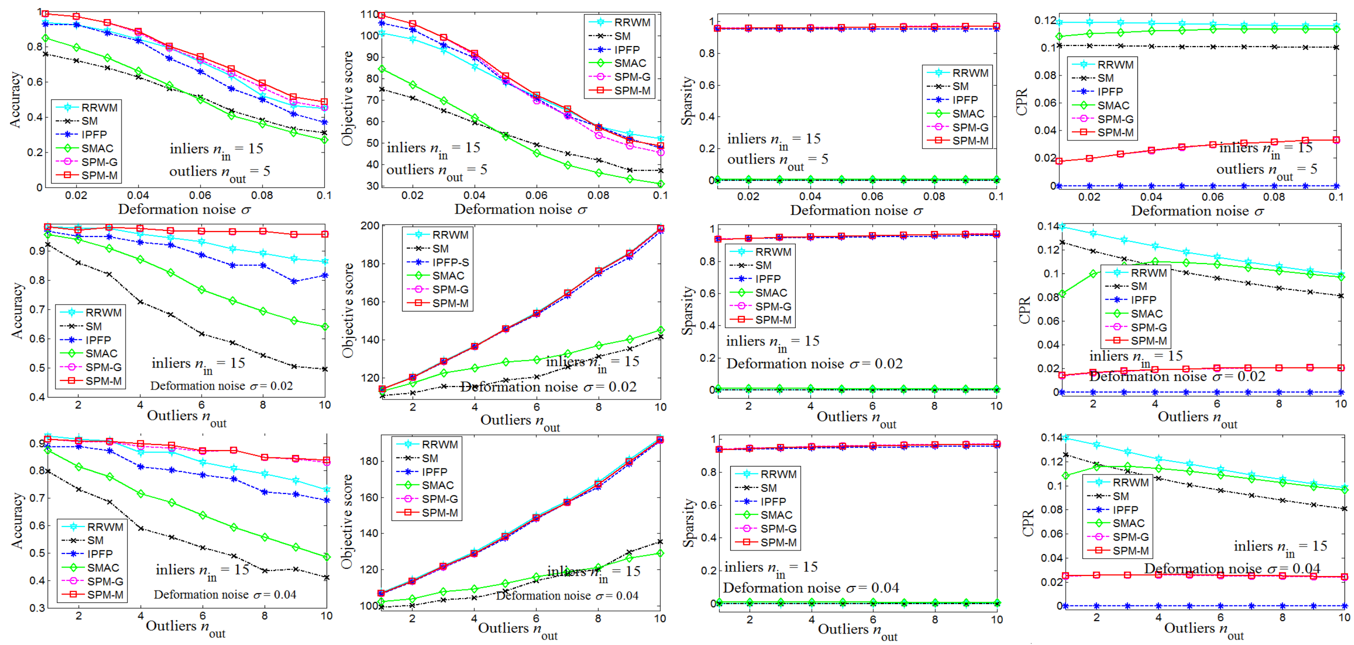


Fig. 4. Comparison results on synthetic point matching across deformation noise and outlier features in both feature sets.

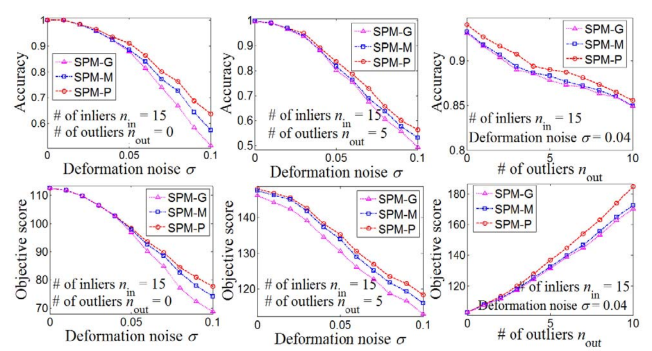


Fig. 5. Comparison results on synthetic point matching under SPM-G, SPM-M and SPM-P algorithm, respectively.

of $\mathbf{x}^{(t)}$ follows the same trend as the sparsity curve, indicating a strong relationship between CPR and sparsity. This clearly demonstrates that SPM incorporates the discrete mapping constraint more and more approximately in the optimization process, and thus can optimize the IQP matching problem in an approximate discrete domain, i.e., search for an approximate discrete solution for the feature matching problem. (3) As the iteration increases, the objective score of $\mathbf{x}^{(t)}$ increases in general and approximates the baseline more and more closely, suggesting that SPM can find the discrete solution for the original IQP problem more optimal than SM by further enforcing the solution to be sparse and optimizing the problem in an approximate discrete domain.

These observations suggest that SPM satisfies both constraint and objective properties sufficiently and thus approximates the original IQP matching problem more closely than SM, i.e., its optimum is closer to the discrete global optimum of the original IQP problem. These are consistent with the results shown in Fig. 1 and obviously demonstrate the benefits of SPM method. Fig. 3 shows the average performance for our SPM vs. SM method on synthetic data under different deformation noise levels. Here, we note that, SPM maintains sparse solution and retains lower CPR value than SM with different noise levels. Also, as noise level increases, SPM significantly outperforms SM in both objective score and matching accuracy, and shows larger performance gaps from SM method. Here, the accuracy is measured by the number of detected true matches divided by the total number of ground truths. These are consistent with the results shown in Fig. 2 and will be further quantified in the experiments.

VI. Experiment

In this section, we applied SPM some feature matching tasks. There are two aspects to this study [13], [29], [30]. We commence with a sensitivity study using the synthetic data. The aim is to evaluate the average performance of the method on the different levels of deformation noise and numbers of outlier feature points. The second part of the study evaluates the method on the real-world image data. Both in synthetic data and real-world image data experiments, we have used the one-to-one mapping constraints, i.e., one model feature can match at most one data feature and vice-versa. As discussed in Section IV, we implement our SPM using both the update rule (7) (SPM-M) and (8) (SPM-G), respectively. We compare our method with the other state-of-art methods including SM [7], IPFP [17], SMAC [16] and RRWM [13]. The parameters are set as follows: maximum iteration $\mathrm{T} = 200$, error $\delta_0 = 1\mathrm{e}-6$.

A. Synthetic Point Sets Matching

Our first experiment is based on synthetic 2D point sets data. Similar to the work [7], [13], we have randomly generated data set of n_{M} 2D model points for model features

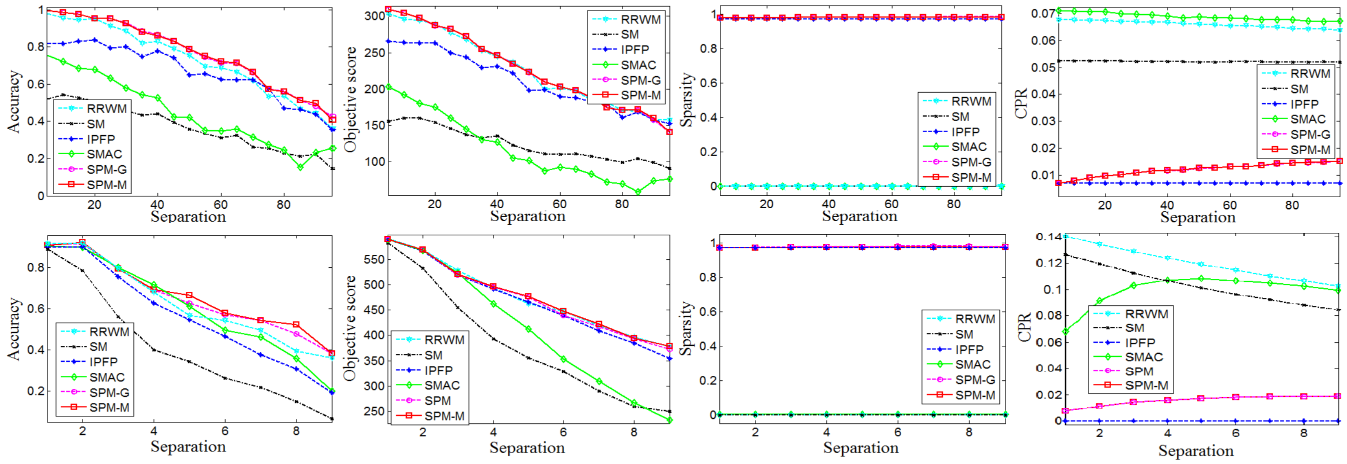


Fig. 6. Comparison results on feature points matching across CMU and YORK sequences, respectively.

M. The range of the x-y point coordinates is $\sqrt{n_M/10}$ to enforce an approximate constant density of 10 points over a 1.0×1.0 region. Then, we obtain the corresponding n_{D} data points in data features D by adding Gaussian noise $N(0, \sigma)$ to the n_{M} point positions from M and then transforming the whole data set with a random rotation and translation. The parameter σ controls the level of position deformation. In addition to the deformation noise, we have also evaluated the effect of the proposed method when outlier features exist in both feature sets. Here n_{out} outlier feature points have been added in both feature sets at random positions. For each assignment $a = (i, i')$, the matching score $\mathbf{S}_a$ is computed as $\mathbf{S}_a = 1 - \|\mathbf{a}_i^{\mathrm{D}} - \mathbf{a}_{i'}^{\mathrm{M}}\|_F^2 \left(\max_a \|\mathbf{a}_i^{\mathrm{D}} - \mathbf{a}_{i'}^{\mathrm{M}}\|_F^2\right)^{-1}$ where $\mathbf{a}_i^{\mathrm{D}}$, $\mathbf{a}_{i'}^{\mathrm{M}}$ are the shape context descriptors for the feature points i in D and i' in M, respectively. For each pair of assignments $(a,\ b)$, where $a = (i, i')$ and $b = (j, j')$, the affinity $\mathbf{W}_{a,b}$ is computed by $\mathbf{W}_{a,b} = e^{-\|d_{ij} - d_{i'j'}\|_F^2/\sigma_r}$, where d_{ij} is the Euclidean distance between the points i and j, and similar to $d_{i'j'}$. The affinity between two each pair of conflict assignments is set to zero, as discussed in Section IV. Scaling facto σ_r has been set to 0.03 in this experiment. For each noise level, we have generated 100 random point sets and then computed the average performances including accuracy, objective score, sparsity and CPR.

Fig. 4 shows the comparison results. Here we can observe that: (1) our SPM method (both SPM-G and SPM-M) can generate expected sparse solutions and return desirable lower CPR than the other methods, indicating that SPM satisfies the desirable constraint property more closely. Since IPFP puts the discretization step into its optimization process, it can satisfy the mapping constraints strictly (CPR = 0). However, our SPM method obviously returns the higher objective score and matching accuracy than IPFP method. (2) RRWM performs better than SM, SMAC, and IPFP, indicating the effectiveness of this probabilistic matching method. (3) SPM generally returns the higher matching accuracy and objective score than the other methods, demonstrating the effectiveness of the proposed sparse optimization based feature point matching method. (4) Compared with SPM-G, SPM-M generally performs slightly better in both matching accuracy and objective score, indicating the more effectiveness of the proposed multiplicative update algorithm (7) on conducting feature matching tasks. In order to evaluate the performance of the proposed SPM-P algorithm, here we penalize the affinity between two conflict assignments with a nonnegative value w ($w = -1$), as discussed in Section IV. In this case, the affinity matrix $\mathbf{W}$ contains negative elements and SPM-P can be directly used to find the optimal solution. Fig. 5 shows the comparison results. We can note that, SPM-P outperforms both SPM-M and SPM-G matching algorithms. This obviously demonstrates that SPM-P algorithm can return desired effective solution for feature matching problem with the presence of negative affinities, indicating the effectiveness and benefits of the proposed SPM-P algorithm.

B. Feature Point Matching Across Image Sequences

In this section, we perform feature point matching on CMU and YORK house sequences which have been widely used in previous works [13], [29], [30]. For CMU house sequence, there are 111 images of a toy house captured from moving viewpoints. For each image, we have first manually marked 25 landmark feature points with known correspondences and then added 10 outlier feature points at random positions. We have matched all images spaced by 5, 10, 15, 20, 95 and 100 frames and computed the average accuracy per separation gap. YORK house sequence contains 18 images and the adjacent images were obtained according to the rotation of 5 degree. For each image, 40 landmark points were manually marked with known correspondences. We have matched all images spaced by 1, 2, 3…9 and 9 frames and computed the average accuracy per separation gap. For each image pair in these two image sequences, the coordinates of their landmark points have been first normalized to the interval [0, 1]. Similar to the synthetic data, we computed the matching score for each assignment $a = (i, i')$ as $\mathbf{S}_a = 1 - \|\mathbf{a}_i^{\mathrm{D}} - \mathbf{a}_{i'}^{\mathrm{M}}\|_F^2/\max_a \|\mathbf{a}_i^{\mathrm{D}} - \mathbf{a}_{i'}^{\mathrm{M}}\|_F^2$, where $\mathbf{a}_i^{\mathrm{D}}, \mathbf{a}_{i'}^{\mathrm{M}}$ are the shape context descriptors for the feature point i in D and i' in M, respectively. The affinity $\mathbf{W}_{a,b}$ for each assignment pair (a, b) $(a = (i.i'),\ b = (j, j'))$ has been computed by $\mathbf{W}_{a,b} = e^{-\|d_{ij} - d_{i'j'}\|_F^2/\sigma_r}$, where d_{ij}

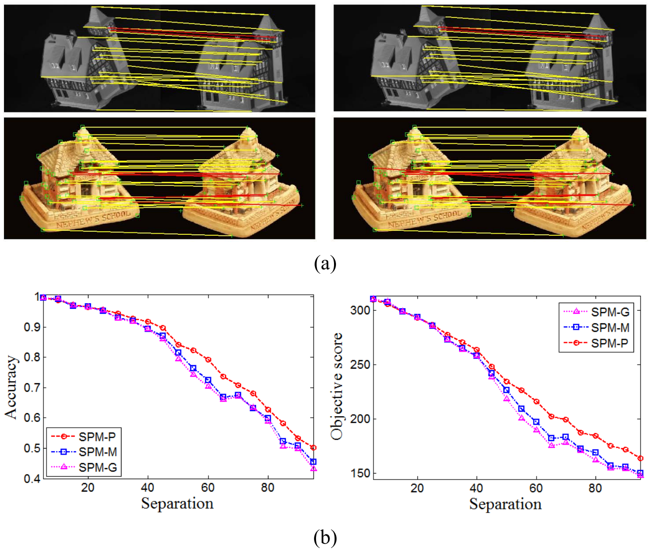


Fig. 7. Comparison results on feature points matching across CMU sequence under SPM-G, SPM-M and SPM-P algorithm, respectively. (a) Matching examples (left: SPM-G; right: SPM-M). (b) Comparison results under SPM-G, SPM-M and SPM-P algorithm.

is the Euclidean distance between two points i and j, and similar to $d_{i'j'}$. Scaling σ_r has been set to 0.03 in this experiment. Some matching examples are shown in Fig. 7(a). Fig. 6 shows the performance curves with respect to the separation gaps. Overall, it is possible to summarize the results as follows: (1) Comparing with SM, SMAC and RRWM, SPM (both SPM-G and SPM-M) possess the sparse property of the discrete IPFP method and thus satisfies the constraint property more closely. (2) SPM returns the best performances in both matching accuracy and objective score, indicating the effectiveness of SPM matching method. (3) SPM-M usually returns better performance than SPM-G, especially on YORK data set. These comparison results are generally consistent with the results in the synthetic data experiments, and further demonstrate the effectiveness of the proposed SPM model on achieving feature point matching tasks. Fig. 7(b) shows the comparison results of the SPM-M, SPM-G and SPM-P algorithms, respectively. We can note that SPM-P obviously outperforms SPM-M and SPM-G, which further demonstrates the benefits of the proposed SPM-P on conducting feature matching problem with the presence of negative affinities.

C. Real-World Image Matching

In this section, we tested our method on some real image feature matching problems. Firstly, we evaluate the robustness of the proposed method via experiments with various types of image pairs used in the works [31]. It contains images of various geometric and photometric transformations for different scene types. Fig. 8(a) shows the six images selected from this data set. Feature points and correct matches have been detected on each image pair [31]–[33]. For each image, we first generate 128-dim SIFT feature descriptors for the feature points. Then, using the distance of descriptor, all the possible candidate assignments were collected. Each model feature can match to the 4 closest data features in SIFT feature space, allowing multiple correspondences for each feature. Note that

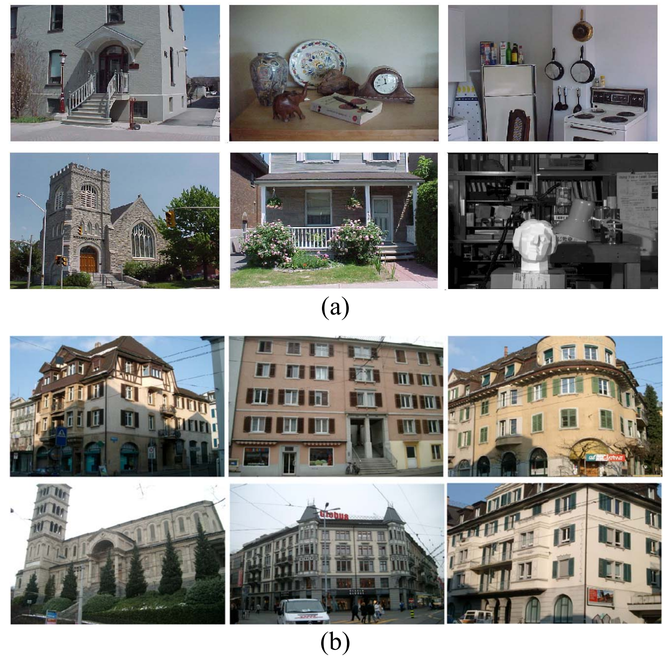


Fig. 8. Images used for evaluation in our experiments. (a) Images used in the work [31]. (b) Images selected from Zubud database.

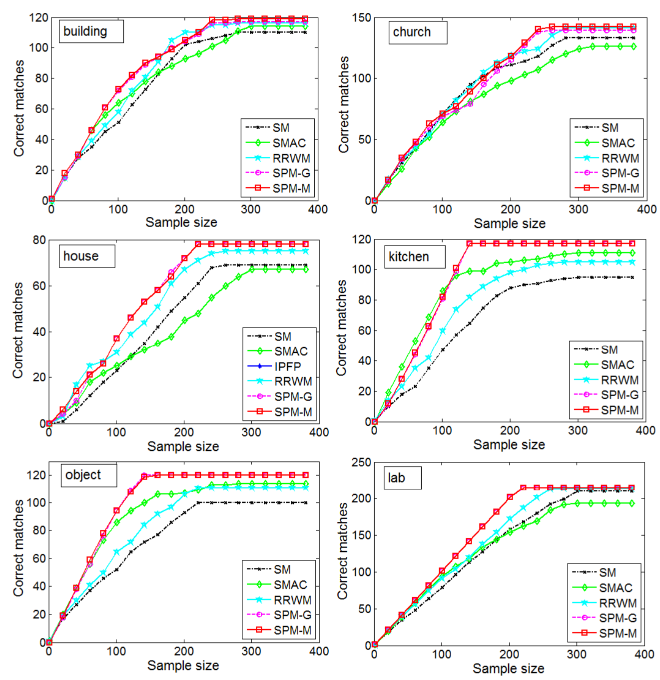


Fig. 9. Matching results between different images; the graph shows how many of the top k best correspondences are correct (k = sample size).

if $\mathbf{x}^*$ is the optimal solution of relaxation continuous problem for feature matching, then $\mathbf{x}^*(a)$ can be generally regarded as the confidence that a ($a = (i, i')$) is a correct assignment [7]. Thus, we can use a greedy algorithm to select top k best correspondences (assignments) from the possible candidate assignments [7]. We compare our SPM with SM, SMAC and RRWM, since IPFP generally returns binary solution ($\mathbf{x}^*(a) \in \{0, 1\}$) and its optimal solution $\mathbf{x}^*(a)$ cannot be

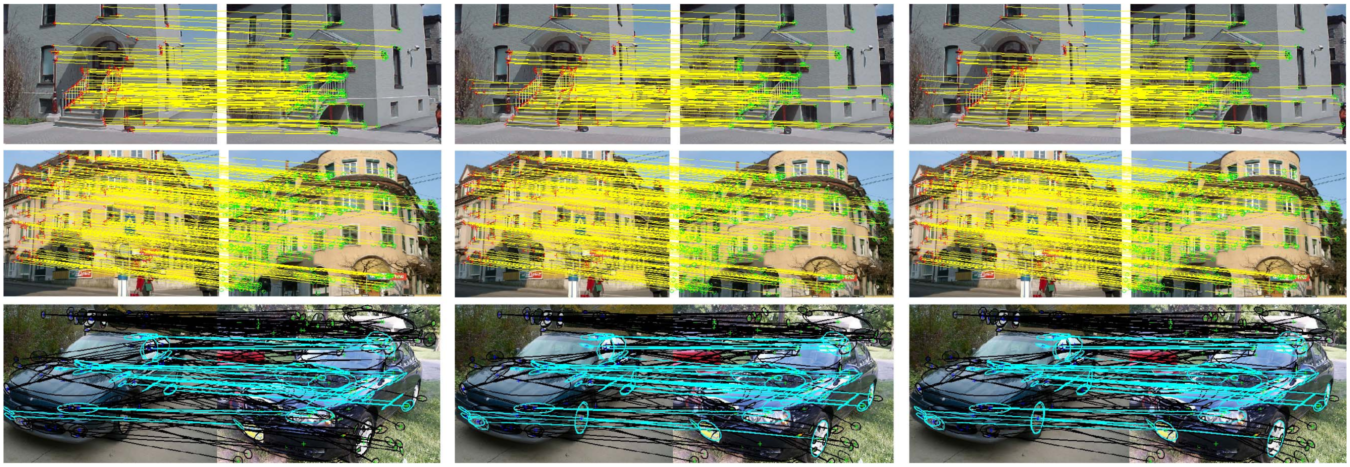


Fig. 10. Some matching examples between real-world images.

TABLE I
MATCHING RESULTS ON ZUBUD AND CALTECH-101-MSRC DATABASES

		SM	SMAC	RRWM	IPFP	SPM-G	SPM-M
ZuBud	Acc	95.11	91.76	92.16	96.10	99.95	99.95
	RS	98.20	96.25	98.05	98.70	100	100
Caltech -MSRC	Acc	62.61	52.38	68.15	63.09	71.82	72.01
	RS	90.01	75.47	99.87	91.00	89.64	90.69

regarded as the correct confidence and thus cannot be used directly here. The affinity between two candidate assignments $a = (i, i')$ and $b = (j, j')$ was computed as $\mathbf{W}_{a,b} = 1 - |d_{ij} - d_{i'j'}|/\max_{a,b}|d_{ij} - d_{i'j'}|$, where d_{ij} is the Euclidean distance between the feature points i and j, and similar to $d_{i'j'}$. The score for each assignment $a = (i, i')$ is computed as $\mathbf{S}_a = e^{-||\mathbf{a}_i^{\mathrm{D}} - \mathbf{a}_{i'}^{\mathrm{M}}||_F^2/\sigma_r}$ where $\mathbf{a}_i^{\mathrm{D}}, \mathbf{a}_{i'}^{\mathrm{M}}$ are the SIFT descriptors for the feature points i in D and i' in M, respectively. Scaling σ_r has been set to 150 in this experiment. Since the ground truths have been manually labeled for all the candidate correspondences of each image pair, the correct matches obtained by the matching algorithms can be accounted. Fig. 10 top row shows the top 150 best assignments for the object image pair using SM and SPM, respectively. Fig. 9 summarizes how many of the top k best correspondences are correct (k = sample size). It is noted that, RRWM generally works better than SM, SMAC for all image pairs. SPM outperforms RRWM and gives the best results. SPM-M slightly outperforms SPM-G on the building and church image pairs, and returns very similar performance with SPM-G on the other images. These obviously suggest that SPM can effectively find the correct matches from the candidate correspondences.

Secondly, we test our method on the image pairs selected from Zurich Building Image Database (ZuBud) [33]. Some image samples are shown in Fig. 8 (b). Similar to the first real image dataset, we generate the possible candidate assignments based on SIFT feature similarity (distance). The matching score $\mathbf{S}$ and affinity for each assignment pair $\mathbf{W}$ are computed as same as that in the prior real image dataset. 30 image pairs have been selected for evaluation in this experiment. We compare our method with SM, SMAC, RRWM and IPFP. Some matching examples are shown in middle row in Fig. 10. Table I summarizes the average matching performance. Here, the relative score (RS) for i-th method is computed as: Relative score$_i$ = Objective score$_i$/$\max_i$ {Objective score$_i$}. One can note that, IPFP generally performs better than SM, SMAC and RRWM in this dataset, indicating the effectiveness of IPFP method. SPM generally works better than IPFP and gives the best performances. This further suggests the effectiveness of SPM on conducting the real-world image matching tasks.

Thirdly, we evaluate our method on the image pairs selected from Caltech-101 and MSRC datasets [13]. Here, 30 image pairs containing various images have been selected, and the candidate correspondences have been generated using the SIFT feature descriptors and the MSER detector [32], [33]. The possible candidate correspondences have been firstly obtained if the feature pair has closer distance in SIFT feature space than a loose threshold $\delta = 0.6$. For each pair of assignments $a = (i, i')$ and $b = (j, j')$, the affinity $\mathbf{W}_{a,b}$ has been computed as $\mathbf{W}_{a,b} = 1 - d_{a,b}/\max_{a,b} d_{a,b}$, where $d_{a,b}$ is the dissimilarity between two candidate region correspondences a and b measured by adopting the mutual projection error function [13], [32], [33]. The matching score for each assignment $a = (i, i')$ has been computed as $\mathbf{S}_a = 1 - ||\mathbf{a}_i^{\mathrm{D}} - \mathbf{a}_{i'}^{\mathrm{M}}||_F^2/\max_a ||\mathbf{a}_i^{\mathrm{D}} - \mathbf{a}_{i'}^{\mathrm{M}}||_F^2$, where $\mathbf{a}_i^{\mathrm{D}}, \mathbf{a}_{i'}^{\mathrm{M}}$ are the SIFT descriptors for the features i in D and i' in M, respectively. The ground truths have been manually labeled for all candidate correspondences for each image pair, and the accuracy and relative objective score have been computed [13]. Some matching examples are shown in bottom row in Fig. 10, and the comparison results including average accuracy and relative objective score are summarized in Table I. We can note that SPM clearly outperforms the other methods in accuracy, demonstrating the effectiveness of our method on conducting real image matching.

D. Complexity Analysis

As summarized in Algorithm 1, our SPM can be efficiently computed by an iteration method. Assume $\mathbf{W}$ is a $n \times n$ matrix with full matching, then its computational

Algorithm 1 Sparse Feature Point Matching (SPM)

Input: Affinity matrix $\mathbf{W}$, score matrix $\mathbf{S}$, size of feature sets D, M: $m = |$ D $|$, $n = |$ M $|$, maximum iteration T, error δ_0

Output: Final discrete binary matching solution $\tilde{\mathbf{x}}^*$

1: Compute the principal eigenvector $\mathbf{v}$ of $\mathbf{W}$

2: Initialize: $\mathbf{x}_i^{(0)} = \mathbf{v}_i(\sum_{i=1}^{mn} \mathbf{v}_i)^{-1}$, $l_0 = \| \mathbf{x}^{(0)} \|_0$, $t = 0$

3: **While** t < T and $|\varepsilon_1(\mathbf{x}^{(t+1)}) - \varepsilon_1(\mathbf{x}^{(t)})| > \delta_0$ **do**

4: Update $\mathbf{x}_i^{(t)}$ $(i = 1, 2 \ldots mn)$ as follows

$$\mathbf{x}_i^{(t+1)} = \mathbf{x}_i^{(t)} \frac{2(\mathbf{W}\mathbf{x}^{(t)})_i + \mathbf{S}_i}{2(\mathbf{x}^{(t)})^{\mathrm{T}}\mathbf{W}\mathbf{x}^{(t)} + \mathbf{S}^{\mathrm{T}}\mathbf{x}^{(t)}}$$

5: $t = t + 1$

6: **End while**

7: $\mathbf{x}^* = \mathbf{x}^{(t+1)}$

8: Compute the binary matching solution $\tilde{\mathbf{x}}^*$ based on $\mathbf{x}^*$ using greedy algorithm [7]

9: **Return:** Final binary matching solution $\tilde{\mathbf{x}}^*$

complexity is less than $O(n^2)$ per iteration. Therefore the total complexity for SPM is less than $O(Mn^2)$, where M is the maximum iterations. Our experience is that Algorithm 1 converges quickly and the average iteration M is usually less than 300. Theoretically, SPM has the same complexity with IPFP and SMAC methods.

VII. Conclusion

A robust feature point matching method based on sparse model is proposed in this paper. Firstly, a general nonnegative sparse quadratic model has been proposed, followed by a general update algorithm to solve it. Then, we present our sparse feature matching method based on the proposed sparse quadratic model. We show that our SPM based solution is sparse and thus approximately imposes the discrete mapping constraints in the optimization process. Also, SPM can find better solution for original IQP feature matching problem than other methods. As important methods in computer vision and pattern recognition, sparse models have been drawing much attention from different communities. In this paper, it has been developed for feature point matching task and achieves promising results. We have shown the strong relationship between the sparsity of the solution and its effectiveness for feature point matching, which is the main contribution of this work.

References

[1] O. Choi and I. S. Kweon, "Robust feature point matching by preserving local geometric consistency," *Comput. Vis. Image Understand.*, vol. 113, no. 6, pp. 726–742, Jun. 2009.

[2] A. Egozi, Y. Keller, and H. Guterman, "Improving shape retrieval by spectral matching and meta similarity," *IEEE Trans. Image Process.*, vol. 19, no. 5, pp. 1319–1327, May 2010.

[3] L. Liang, L. Yongyi, P. Yan, and C. Xiaowu, "Integrating graph partitioning and matching for trajectory analysis in video surveillance," *IEEE Trans. Image Process.*, vol. 21, no. 12, pp. 4844–4857, Dec. 2012.

[4] D. H. Kim, I. D. Yun, and S. U. Lee, "Attributed relational graph matching based on the nested assignment structure," *Pattern Recognit.*, vol. 43, no. 3, pp. 914–928, Mar. 2010.

[5] D. Conte, P. Foggia, C. Sansone, and M. Vento, "Thirty years of graph matching in pattern recognition," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 18, no. 3, pp. 265–298, May 2004.

[6] S. Gold and A. Rangarajan, "A graduated assignment algorithm for graph matching," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, no. 4, pp. 377–388, Apr. 1996.

[7] M. Leordeanu and M. Hebert, "A spectral technique for correspondence problems using pairwise constraints," in *Proc. 10th IEEE Int. Conf. Comput. Vis.*, Oct. 2005, pp. 1482–1489.

[8] M. Zaslavskiy, F. Bach, and J.-P. Vert, "A path following algorithm for the graph matching problem," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 12, pp. 2227–2242, Dec. 2009.

[9] L. Torresani, V. Kolmogorov, and C. Rother, "Feature correspondence via graph matching: Models and global optimization," in *Proc. 10th Eur. Conf. Comput. Vis.*, 2008, pp. 596–609.

[10] J. Maciel and J. P. Costeira, "A global solution to sparse correspondence problems," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 2, pp. 187–199, Feb. 2003.

[11] F. Zhou and F. De la Torre, "Factorized graph matching," in *Proc. IEEE Conf. CVPR*, Jun. 2012, pp. 127–134.

[12] B. J. van Wyk and M. A. van Wyk, "A POCS-based graph matching algorithm," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 11, pp. 1526–1530, Nov. 2004.

[13] M. Cho, J. Lee, and K. M. Lee, "Reweighted random walks for graph matching," in *Proc. 11th ECCV*, 2010, pp. 492–505.

[14] R. Zass and A. Shashua, "Probabilistic graph and hypergraph matching," in *Proc. IEEE Conf. CVPR*, Jun. 2008, pp. 1–8.

[15] O. Enqvist, K. Josephon, and F. Kahl, "Optimal correspondences from pairwise constraints," in *Proc. IEEE 12th Int. Conf. Comput. Vis.*, Sep./Oct. 2009, pp. 1295–1302.

[16] T. Cour, P. Srinivasan, and J. Shi, "Balanced graph matching," in *Proc. Neural Inf. Process. Syst.*, 2006, pp. 313–320.

[17] M. Leordeanu, M. Hebert, and R. Sukthankar, "An integer projected fixed point method for graph matching and MAP inference," in *Proc. Neural Inf. Process. Syst.*, 2009, pp. 1114–1122.

[18] J. J. MuAuley and T. S. Caetano, "Fast matching of large point sets under occlusions," *Pattern Recognit.*, vol. 45, no. 1, pp. 563–569, Jan. 2012.

[19] T. S. Caetano, T. Caelli, D. Schuurmans, and D. A. C. Barone, "Graphical models and point pattern matching," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 10, pp. 1646–1663, Oct. 2006.

[20] D. Chris, T. Li, and M. I. Jordan, "Nonnegative matrix factorization for combinatorial optimization: Spectral clustering, graph matching, and clique finding," in *Proc. 8th IEEE Int. Conf. Data Mining*, Dec. 2008, pp. 183–192.

[21] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[22] J. Duchi, S. Shalev-Shwartz, Y. Singer, and T. Chandra, "Efficient projections onto the l_1-ball for learning in high dimensions," in *Proc. 25th Int. Conf. Mach. Learn.*, 2008, pp. 272–279.

[23] D. Luo, D. Chris, and H. Huang, "Parallelization with multiplicative algorithms for big data mining," in *Proc. IEEE Int. Conf. Data Mining*, Dec. 2012, pp. 489–498.

[24] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. Roy. Statist. Soc. Ser. B*, vol. 58, no. 1, pp. 267–288, 1994.

[25] D. Cai, X. F. He, J. W. Han, and T. S. Huang, "Graph regularized nonnegative matrix factorization for data representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 8, pp. 1548–1560, Aug. 2011.

[26] D. D. Lee and H. S. Seung, "Algorithms for nonnegative matrix factorization," in *Proc. Adv. Neural Inf. Process. Syst.*, 2000, pp. 556–562.

[27] F. Sha, L. K. Saul, and D. D. Lee, "Multiplicative updates for nonnegative quadratic programming in support vector machines," in *Proc. Adv. Neural Inf. Process. Syst.*, 2003, pp. 1041–1048.

[28] D. Chris, T. Li, and M. I. Jordan, "Convex and semi-nonnegative matrix factorization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 1, pp. 45–55, Jan. 2010.

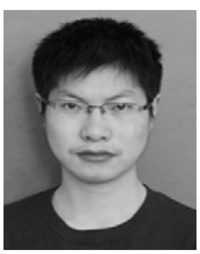
[29] B. Jiang, J. Tang, and B. Luo, "Graph matching with nonnegative sparse model," in *Graph-Based Representations in Pattern Recognition*. Berlin, Germany: Springer-Verlag, 2013, pp. 41–50.

[30] B. Luo, R. C. Wilson, and E. R. Hancock, "Spectral embedding of graphs," *Pattern Recognit.*, vol. 36, no. 10, pp. 2213–2230, 2003.

[31] E. Vincent and R. Laganiere, "Matching feature points in stereo pairs: A comparative study of some matching strategies," *Mach. Graph. Vis.*, vol. 10, no. 3, pp. 237–259, 2001.

[32] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proc. 7th IEEE Int. Conf. Comput. Vis.*, Sep. 1999, pp. 1150–1157.

[33] E. S. Ng and N. G. Kingsbury, "Matching of interest point groups with pairwise spatial constraints," in *Proc. 17th IEEE Int. Conf. Image Process.*, Sep. 2010, pp. 2693–2696.
[34] M. Leordeanu and M. Hebert, "Efficient MAP approximation for dense energy functions," in *Proc. 23rd Int. Conf. Mach. Learn.*, 2006, pp. 545–552.
[35] L. Baratchart, M. Berthod, and L. Pottier, "Optimization of positive generalized polynomials under l^p constraints," *J. Convex Anal.*, vol. 5, no. 2, pp. 353–379, 1998.
[36] L. E. Baum and G. R. Sell, "Growth transformations for functions on manifold," *Pacific J. Math.*, vol. 27, no. 2, pp. 211–227, 1968.
[37] S. R. Bulò and M. Pelillo, "A game-theoretic approach to hypergraph clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 6, pp. 1312–1327, Jun. 2013.
[38] P. Marcello, "Replicator equations, maximal cliques, and graph isomorphism," *Neural Comput.*, vol. 11, no. 8, pp. 1933–1955, 1999.
[39] E. Rodola, A. M. Bronstein, A. Albarelli, F. Bergamasco, and A. Torsello, "A game-theoretic approach to deformable shape matching," in *Proc. IEEE Conf. CVPR*, Jun. 2012, pp. 182–189.
[40] A. Albarelli, S. R. Bulò, A. Torsello, and M. Pelilo, "Matching as a non-cooperative game," in *Proc. IEEE Int. Conf. Comput. Vis.*, Sep./Oct. 2009, pp. 1319–1326.
[41] A. Albarelli, E. Rodola, and A. Torsello, "Imposing semi-local geometric constraints for accurate correspondences selection in structure from motion: A game-theoretic perspective," *Int. J. Comput. Vis.*, vol. 97, no. 1, pp. 36–53, 2012.
[42] C. Ding and X. He, "K-means clustering via principal component analysis," in *Proc. Int. Conf. Mach. Learn.*, 2004, pp. 225–232.
[43] A. Y. Ng, M. I. Jordan, and Y. Weiss, "On spectral clustering: Analysis and an algorithm," in *Proc. Adv. Neural Inf. Process. Syst.*, 2001, pp. 849–856.



Bo Jiang received the B.Eng. degree in computer science from Anhui University, Hefei, China, in 2012, where he is currently pursuing the Ph.D. degree in computer science. His current research interests include image and graph matching, image feature extraction, and pattern recognition.

Jin Tang received the B.Eng. degree in automation and the Ph.D. degree in computer science from Anhui University, Hefei, China, in 1999 and 2007, respectively, where he is currently a Professor with the School of Computer Science and Technology. His research interests include computer vision, pattern recognition, and machine learning.

Bin Luo received the B.Eng. degree in electronics and the M.Eng. degree in computer science from Anhui University, Hefei, China, in 1984 and 1991, respectively, and the Ph.D. degree in computer science from the University of York, York, U.K., in 2002. He has authored over 200 papers in journals, edited books, and refereed conferences. He is currently a Professor with Anhui University. He also chairs the IEEE Hefei Subsection. He served as a peer reviewer of international academic journals, such as the IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, *Pattern Recognition*, *Pattern Recognition Letters*, the *International Journal of Pattern Recognition and Artificial Intelligence*, *Knowledge and Information Systems*, and *Neurocomputing*. His current research interests include random graph-based pattern recognition, image and graph matching, graph spectral analysis, and video analysis.

Liang Lin is currently a Full Professor with the School of Advanced Computing, Sun Yat-sen University, Guangzhou, China. He received the B.S. and Ph.D. degrees from the Beijing Institute of Technology, Beijing, China, in 1999 and 2008, respectively, and the Ph.D. degree from the Department of Statistics, University of California at Los Angeles (UCLA), Los Angeles, CA, USA, in 2007. His Ph.D. dissertation received the China National Excellent Ph.D. Thesis Award Nomination in 2010. He was a Post-Doctoral Research Fellow with the Center for Vision, Cognition, Learning, and Art, UCLA. His research focuses on new models, algorithms and systems for intelligent processing, and understanding of visual data, such as images and videos. He has authored over 50 papers in top-tier academic journals and conferences, including the PROCEEDINGS OF THE IEEE, the IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, the IEEE TRANSACTIONS ON IMAGE PROCESSING, the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, the IEEE TRANSACTIONS ON MULTIMEDIA, *Pattern Recognition*, the Computer Vision and Pattern Recognition Conference, the International Conference on Computer Vision, the European Conference on Computer Vision, the ACM Multimedia Conference, and the Conference on Neural Information Processing Systems. He was supported by several promotive programs or funds for his works, such as the Program for New Century Excellent Talents of the Ministry of Education, China, in 2012, the Program of Guangzhou Zhujiang Star of Science and Technology in 2012, and the Guangdong Natural Science Funds for Distinguished Young Scholars in 2013. He was a recipient of the Best Paper Runners-Up Award in the ACM International Symposium on Non-Photorealistic Animation and Rendering in 2010, and the Google Faculty Award in 2012.