A Low-Rank Matrix Approximation Approach to Multiway Matching with Applications in Multi-Sensory Data Association

Spyridon Leonardos[1], Xiaowei Zhou[2] and Kostas Daniilidis[3]

***Abstract*— Consider the case of multiple visual sensors perceiving the same scene from different viewpoints. In order to achieve consistent visual perception, the problem of data association, in this case establishing correspondences between observed features, must be first solved. In this work, we consider multiway matching which is a specific instance of multi-sensory data association. Multiway matching refers to the problem of establishing correspondences among a set of images from noisy pairwise correspondences, typically by exploiting cycle-consistency. We propose a novel optimization-based formulation of multiway matching problem as a nonconvex low-rank matrix approximation problem. We propose two novel algorithms for numerically solving the problem at hand. The first one is an algorithm based on the Alternating Direction Method of Multipliers (ADMM). The second one is a Riemannian trust-region method on the multinomial manifold, the manifold of strictly positive stochastic matrices, equipped with the Fisher information metric. Experimental results demonstrate that the proposed methods have the state of the art performance in multiway matching while reducing the computational complexity compared to the state of the art.**

I. Introduction

Two-way matching or two-way data association has been long-standing problems in multi-sensory systems, in robotics and in computer vision. It refers to the problem of establishing correspondences between feature points in two overlapping views. The matching problem can be multiway if more than two views of an object or of a scene are available. Multiway matching refers to the problem of establishing correspondences among a collection of images from noisy pairwise correspondences. It is a more recent problem compared to two-view matching which has received increasing amount of attention during the last few years. Multiway matching has been successfully applied to computing consistent pointwise maps among a collection of shapes [22], [16], [14], [13] and to estimating consistent associations among a collection of images between either traditional feature descriptors [23], [10], [33], [31], [28] such as SIFT [20] and shape context features [4] or semantic descriptors [33], [25].

Several methods for multiway matching employ convex relaxation techniques over a Euclidean space or the cone of positive semidefinite matrices along with linear and nonnegativity constraints [14], [13], [10], [33]. We follow a different path: we formulate multiway matching as a nonconvex low-rank matrix approximation problem with significantly lower dimensionality. On the one hand, the ADMM algorithm can be applied to such a problem as it has been applied to nonconvex problems such as l_0-norm constrained problems and nonnegative matrix factorizations [8]. On the other hand, optimization methods on Riemannian manifolds [1] have been recently developed and have found numerous applications in structured nonconvex problems of similar nature to the one considered in this work, such as nonconvex matrix and tensor approximation and completion problems [15], [5], [6], [24]. In the context of this work, we consider the multinomial manifold, i.e. the manifold of discrete probability distributions, which has found applications in classification [18], image segmentation [3] and tensor factorization for clustering [24].

In this work, inspired from recent advances in ADMM-based algorithms and in optimization algorithms on Riemannian manifolds, we propose two novel algorithms, MatchADMM and MatchRTR, for the problem of multiway matching. In contrast to existing works [13], [10], [33] that optimize over the entire matrix of pairwise correspondences, we propose to optimize directly over the low-rank representation, thus significantly reducing the dimensionality of the problem. We also validate our methods using both synthetic and real datasets.

The remainder of the paper is structured as follows. Section II includes an overview of prior works on multiway matching. Notation and preliminaries are introduced in Section III. A formalization of multiway matching is the subject of Section IV-A followed by the proposed problem formulation in Section IV-B. Our first approach is presented in Section V. Our second approach is detailed in Section VI. Finally, experimental evaluation and comparison with existing approaches are included in Section VII.

II. Related Work

A necessary condition for good matching of multiple views is the cycle consistency, meaning that the composition of correspondences along a cycle of views should be equal to the identity. In practice, cycle consistency is not satisfied if pairs of views are matched separately, which is usually the case, due to the presence of erroneous pairwise correspondences. Nguyen *et al.*[22] and Zach *et al.*[32] were the first to propose the use the cycle consistency to identify the correctness of pairwise correspondences. Later, it was proposed [16], [14], [23] that finding cyclically consistent correspondences from noisy pairwise correspondences can be formulated as quadratic integer programming which in turn, can be relaxed into a generalized Rayleigh quotient

[1,3]Spyridon Leonardos and Kostas Daniilidis are with the Department of Computer and Information Science, University of Pennsylvania Philadelphia, PA 19104, USA {spyridon,kostas}@seas.upenn.edu

[2]Xiaowei Zhou is with the College of Computer Science of Zhejiang University, China. xzhou@cad.zju.edu.cn

problem whose solution is easily obtainable by the leading eigenvectors of the matrix of pairwise correspondences.

Although spectral relaxations are easily implementable and come with some theoretical guarantees, in practice, they lack robustness. This shortcoming was remedied by semidefinite programming relaxations proposed by Huang and Guibas [13] and Chen *et al.*[10]. These works provide theoretical guarantees under certain assumption on the noise distribution. However, the proposed semidefinite programs, even when solved with the Alternating Direction Method of Multipliers (ADMM), do not scale well. Zhou *et al.*[33] improved the scalability of the semidefinite programming relaxations by reformulating them as rank minimization and developed a more efficient algorithm by dropping the positive-definite constraint on the matrix of pairwise correspondences.

More recently, Tron *et al.*[25] proposed a generalization of the Quick Shift algorithm [26] for multiway matching and Maset *et al.*[21] suggested a practical modification to the spectral relation which improves performance but consistency is no longer satisfied. All of the aforementioned works address problem in a centralized setting, i.e. all pairwise measurements are available and optimized jointly. Distributed optimization approaches on the set of doubly stochastic matrices have been recently developed by Leonardos *et al.*[19], [**?**]. A different line of works that exploit cycle consistency has been pionered by Yan *et al.*[30], [29], [31], [28] for the problem of multiple graph matching. However, the problems of multiple graph matching and multiway matching are inherently different since the former requires affinities between feature points in the same view and the latter does not.

In the robotics community, several multi-robot SLAM algorithms have been recently developed raising the need for robust multi-sensory data association algorithms. Most of them address data association either in a pairwise fashion by directly generalizing data association techniques from the single robot case [11], [27], [12]. Closer to our approach is the work by Aragues et al. [2] which detects inconsistencies based on cycle detection and the spectral clustering method of Fathian et al. [**?**].

III. Preliminaries and Notation

First, we introduce some notions and notations regarding stochastic matrices and permutations that will be heavily used throughout the paper. A nonnegative matrix is stochastic if all its row sums are equal to 1 and doubly stochastic if both its row sums and column sums are equal to 1. A doubly stochastic matrix is a permutation matrix if its elements are either 0 or 1. The set of $n \times n$ permutation matrices is denoted by $\mathcal{P}_n$. An $m \times n$ stochastic matrix, with $m \leq n$, is a partial permutation matrix if its elements are either 0 or 1 and it has orthogonal rows. The set of $m \times n$ partial permutation matrices is denoted by $\mathcal{P}_{m,n}$. Let $[n] \doteq \{1, 2, \ldots, n\}$ for some positive integer n. A mapping $\pi : [n] \to [n]$ is a permutation of $[n]$ if it is bijective. The set of all permutations of $[n]$ forms a group under composition $\circ$, termed the symmetric group $\mathfrak{S}_n$. A permutation $\pi \in \mathfrak{S}_n$ is represented by an $n \times n$ permutation matrix $X \in \mathcal{P}_n$ such that

$$(X)_{ij} = \begin{cases} 1, & \text{if } \pi(j) = i \\ 0, & \text{otherwise,} \end{cases} \tag{1}$$

where $(X)_{ij}$ denotes the element of matrix X in the ith row and jth column. A mapping $\pi : [m] \to [n]$, $m \leq n$, is a partial permutation if it is injective. A partial permutation matrix is represented by an $m \times n$ partial permutation matrix.

IV. Consistent Multiway Matching

A. Consistency

In this section, we formalize the problem of consistent multiway matching. We assume there are m collections (views), each containing n features. The pairwise association $\pi_{ij} \in \mathfrak{S}_n$ is defined as follows: we have that $\pi_{ij}(l) = k$ if the lth feature in collection j corresponds to the kth feature in collection i. We denote by $\widetilde{\pi}_{ij} \in \mathfrak{S}_n$ the, possibly erroneous, estimated pairwise association between collections i and j, which is the output of some pairwise matching algorithm. Let $\widetilde{X}_{ij}$ denote the corresponding matrix representation of $\widetilde{\pi}_{ij}$. A set of pairwise correspondences $\{\widetilde{\pi}_{ij}\}_{i,j=1}^m$ is *consistent* [13] if

$$\widetilde{\pi}_{ij} \circ \widetilde{\pi}_{jk} = \widetilde{\pi}_{ik} \qquad (\text{resp. } \widetilde{X}_{ij}\widetilde{X}_{jk} = \widetilde{X}_{ik}), \tag{2}$$

for all valid indices i, j, k. It has been shown by Huang and Guibas [13] that the consistency constraint (2) is equivalent to the existence of $\pi_1, \pi_2, \ldots, \pi_m \in \mathfrak{S}_n$ (resp. $X_1, X_2, \ldots, X_m \in \mathcal{P}_n$), such that for all valid i, j we have

$$\widetilde{\pi}_{ij} = \pi_i \circ \pi_j^{-1} \qquad (\text{resp. } \widetilde{X}_{ij} = X_i X_j^T). \tag{3}$$

More generally, assume that collection i has n_i features and let $N = \sum_{i=1}^m n_i$. Let K denote the total number of distinct features, termed size of the "universe of features" [10], [33], with $n_i \leq K$ for all $i = 1, \ldots, m$. We have that $(\widetilde{X}_{ij})_{kl} = 1$ if the lth feature in collection j corresponds to the kth features in collection i. In the case where all pairwise association are available, the consistency condition can be equivalently reformulated as follows: define the block matrix $\widetilde{\mathbf{X}} \in \mathbb{R}^{N \times N}$ by

$$\widetilde{\mathbf{X}} = \begin{bmatrix} I & \widetilde{X}_{12} & \cdots & \widetilde{X}_{1m} \\ \widetilde{X}_{12}^T & I & \cdots & \widetilde{X}_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \widetilde{X}_{1m}^T & \widetilde{X}_{2m}^T & \cdots & I \end{bmatrix} \tag{4}$$

Then, the set of pairwise correspondences $\{\widetilde{X}_{ij}\}_{i,j=1}^m$ is *consistent* if and only if $\widetilde{\mathbf{X}}$ has a low-rank approximation of the form

$$\widetilde{\mathbf{X}} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix} \cdot \begin{bmatrix} X_1^T & X_2^T & \cdots & X_n^T \end{bmatrix} \tag{5}$$

for some $X_1, X_2, \ldots, X_m$ such that each X_i is a partial permutation matrix, that is $X_i \in \mathcal{P}_{n_i,K}$ for all $i = 1, 2, \ldots, m$.

Intuitively, $X_i \in \mathcal{P}_{n_i,K}$ is the matrix representation of a (partial) permutation map $\pi_i : [n_i] \to [K]$ from the labels of collection i to some global labels, termed the "universe of features" [10], [33].

B. Proposed Formulation

Based on the preceding analysis, we formulate consistent multiway matching as the following (combinatorial) symmetric low-rank matrix approximation problem:

$$\begin{aligned} \underset{X}{\text{minimize}} \quad & \phi(X) \doteq (1/4)\|\widetilde{\mathbf{X}} - XX^T\|_F^2 \\ \text{subject to} \quad & X_i \in \mathcal{P}_{n_i,K}, \qquad i = 1, 2, \ldots, m \end{aligned} \tag{6}$$

Existing works [13], [10], [33] for multiway matching attempt to find a positive definite matrix $\mathbf{X} \in \mathbb{R}^{N\times N}$ that satisfies the consistency constraints and is close to the measured $\widetilde{\mathbf{X}}$. However, these approaches do not scale well with the number of views. In contrast to these approaches that optimize over the pairwise correspondences, we propose to solve directly for the labels $X = [X_1^T \cdots X_m^T]^T \in \mathbb{R}^{N\times K}$. Optimizing directly in the label space significantly reduces the dimension of the domain of the optimization problem from N^2 to $NK \ll N^2$. For instance, when $n_1 = n_2 = \cdots = n_m = K$, the dimension of the domain of the optimization problem reduces from m^2K^2 i.e. quadratic in m, to mK^2, i.e. linear in m.

There are two main reasons as to why problem (6) is challenging. The first challenge is the combinatorial hardness due to the permutation constraints along with the nonlinearity of the objective. This is customarily remedied by relaxing the domain of the problem from permutations to doubly stochastic or stochastic matrices. This is, indeed, the path we follow in one of the two proposed approaches, which is presented in Section VI. The second challenge is that even if the domain of the problem was convex, the objective would still be nonconvex. This is the price we pay for significantly reducing the dimension of the domain of the proposed optimization problem compared to prior art as discussed earlier. Nevertheless, we present experimental evidence pointing to the fact that both of the proposed methods for solving (6) are at least as accurate as the state of the art.

V. ADMM Algorithm for Multiway Matching

A. ADMM

The Alternating Direction Method of Multipliers (ADMM) [8] is a widely used optimization method especially suitable for large scale problems. With ADMM, one can solve optimization problems with separable objective and linear equality constraints, that is, optimization problems of the following so-called standard form:

$$\begin{aligned} \underset{x,z}{\text{minimize}} \quad & f(x) + g(z) \\ \text{subject to} \quad & Ax + Bz = c, \end{aligned} \tag{7}$$

where f, g are real-valued functions. For problem (7), the augmented Lagrangian is given by

$$\begin{aligned} L_\rho(x, z, y) = f(x) + g(z) + y^T(Ax + Bz - c) \\ + (\rho/2)\|Ax + Bz - c\|_2^2, \end{aligned} \tag{8}$$

where $\rho > 0$ is the penalty parameter. Then, the ADMM iterations are as follows

$$x^{k+1} := \underset{x}{\operatorname{argmin}}\, L_\rho(x, z^k, y^k), \tag{9a}$$

$$z^{k+1} := \underset{z}{\operatorname{argmin}}\, L_\rho(x^k, z, y^k), \tag{9b}$$

$$y^{k+1} := y^k + \rho(x^{k+1} - z^{k+1}). \tag{9c}$$

Intuitively, the ADMM algorithm is an approximate dual ascent method. Instead of minimizing the Lagrangian jointly over x and z to compute the gradient of the dual function, the augmented Lagrangian is minimized in an alternate fashion. Although the standard form (7) may seem restrictive at first sight, it is in fact as general as one can wish for. Any optimization problem can be written in form (7) as follows. Assume that the problem at hand is minimizing a real-valued objective $f(x)$ over some set $\mathcal{C}$. Let $I_\mathcal{C}$ denotes the indicator function of the set $\mathcal{C}$, which takes the value 0 in $\mathcal{C}$ and the value $+\infty$ outside of $\mathcal{C}$. Then, we can equivalently write it in standard form as

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & f(x) + I_\mathcal{C}(z) \\ \text{subject to} \quad & x - z = 0, \end{aligned} \tag{10}$$

and the ADMM iterations are as follows

$$x^{k+1} := \underset{x}{\operatorname{argmin}} \left\{ f(x) + (\rho/2)\|x - z^k + u^k\|_F^2 \right\}, \tag{11a}$$

$$z_i^{k+1} := \Pi_\mathcal{C}\left(x^{k+1} + u^k\right), \tag{11b}$$

$$u^{k+1} := u^k + x^{k+1} - z^{k+1}, \tag{11c}$$

where $u^k = (1/\rho)y^k$ is the scaled dual vector. For the above iterations to make sense in terms of computational tractability, the projection operation $\Pi_\mathcal{C}(\cdot)$ onto the set $\mathcal{C}$ has to be efficiently computable. Altough ADMM is commonly used for convex optimization problems for which convergence and optimality guarantees exist under mild assumptions, it has been applied to nonconvex problems as well such as nonnegative matrix approximation, minimization with respect to cardinality (l_0 norm) or rank constraints. The method we propose in the following subsection falls into this category of problems.

B. The MatchADMM Algorithm

In this subsection, we derive an ADMM algorithm for solving problem (6). Based on the analysis of the previous subsection, we modify problem (6) into an equivalent problem in ADMM standard form and then, we derive the ADMM iterations.

Specifically, for each $n_i \times K$ matrix X_i, we introduce a new matrix Z_i of the same dimensions and we use the shorthand notation $Z = [Z_1^T \cdots Z_m^T]^T$. Let $I_{\mathcal{P}_{n_i\times K}}$ denote the indicator function of the set of $n_i \times K$ partial permutations, that is

$$I_{\mathcal{P}_{n_i\times K}}(Z_i) = \begin{cases} 0, & Z_i \in \mathcal{P}_{n_i,K} \\ +\infty, & Z_i \notin \mathcal{P}_{n_i,K}. \end{cases} \tag{12}$$

Then, problem (6) is equivalent to the following optimization problem in ADMM standard form

$$\begin{aligned} \underset{X,Z}{\text{minimize}} \quad & \phi(X) + \sum_{i=1}^{m} I_{\mathcal{P}_{n_i \times K}}(Z_i) \\ \text{subject to} \quad & X_i - Z_i = 0, \quad i = 1, 2, \ldots, m, \end{aligned} \tag{13}$$

which we solve by the proposed MatchADMM algorithm, outlined in Algorithm 1. Note that $U = [U_1^T \cdots U_m^T]^T$ is the scaled dual matrix of the same dimensions as X and Z. As it is customarily done when ADMM is used for nonconvex problems, we increase the penalty parameter ρ^k at every iteration by a factor of α to improve the convergence of the the algorithm. We use the default value $\alpha = 1.2$. As noted in [8], when a varying penalty parameter is used in the scaled form of ADMM, the scaled dual matrix must also be rescaled accordingly.

To check whether the MatchADMM has converged, we look at the primal and dual residuals, which in our case take the following simple form

$$R^k = X^k - Z^k, \qquad S^k = -\rho^k (Z^k - Z^{k-1}). \tag{14}$$

In practice, we observe that there is a nonnegative integer k_0 such that $Z^k = Z^{k_0}$ for all $k \geq k_0$ and as $k \to \infty$, $(X^k, U^k) \to (Z^{k_0}, \mathbf{0})$. The above observation suggests that we can obtain the limit as $k \to \infty$ of the sequence (Z^k, X^k, U^k) generated by Algorithm 1 in a finite number of steps by simply detecting the first iteration index k_0 dual residual becomes equal to zero. In general, however, convergence properties of ADMM are hard to analyze for nonconvex problems [8] and the obtained solution might depend on initialization and the value of the penalty parameter. We initialize the algorithm with a random stochastic matrix X^0, $U^0 = \mathbf{0}$ and $\rho^0 = 1$.

The first step of every iteration of MatchADMM, the X-update, involves solving an optimization problem, for which we use the trust-region solver of Manopt [7]. The required gradient and Hessian that need to be provided to the solver are computed in the Appendix. Unfortunately, the aforementioned optimization problem is not convex, in general, since $\phi(X)$ is not convex. The Z-update step involves the projection onto the set of partial permutations is done using the Hungarian algorithm [17].

Algorithm 1 MatchADMM

Input: Correspondences $\{\widetilde{X}_{ij}\}_{i,j=1}^m$, universe size K, parameters $\alpha > 1$, ρ^0
Ouput: Permutations $\{X_i\}_{i=1}^m$, consistent correspondences $\{X_{ij} \doteq X_i X_j^T\}_{i,j=1}^m$
repeat
$X^{k+1} := \underset{X \in \mathbb{R}^{N \times K}}{\operatorname{argmin}} \phi(X) + (\rho^k/2)\|X - Z^k + U^k\|_F^2$
$Z_i^{k+1} := \Pi_{\mathcal{P}_{n_i,K}}\left(X_i^{k+1} + U_i^k\right)$
$U^{k+1} := (1/\alpha)(U^k + X^{k+1} - Z^{k+1})$
$\rho^{k+1} := \alpha \rho^k$
until Convergence

VI. Multiway Matching on the Multinomial Manifold

In this section, we present a second distinct approach to solving (6). We relax the domain of the problem from partial permutations to stochastic matrices and propose a regularized optimization problem on this set. This is summarized in Section VI-A. Next, we present the MatchRTR, see Algorithm 2, a Riemannian trust-region method [1] for efficiently solving the proposed relaxation. Trust-region methods globally converge to critical points of the cost function while rendering saddle points and local maxima unstable and achieving a local superlinear rate of convergence.

A. Proposed Relaxation

To deal with the computational intractability of problem (6), we relax the domain of the problem from partial permutation matrices to stochastic matrices. Intuitively, for each observed feature, we maintain a row vector containing the probabilities of correspondences between that feature and the idealized features of the universe. A tighter relaxation is possible by including column sums constraints, however we observed that adding these constraints does not offer any improvement in matching accuracy. Since partial permutation matrices lie in the intersection of stochastic and row orthonormal matrices, we propose the following regularizer

$$\psi(X) = (1/4)\|I - XX^T\|_F^2, \tag{15}$$

to obtain the final formulation as follows:

$$\begin{aligned} \underset{X}{\text{minimize}} \quad & \phi_\lambda(X) \doteq \phi(X) + \lambda \sum_{i=1}^{m} \psi(X_i) \\ \text{subject to} \quad & X_i \mathbf{1} = \mathbf{1}, \; X_i \geq 0, \quad i = 1, 2, \ldots, m. \end{aligned} \tag{16}$$

B. The MatchRTR Algorithm

The multinomial manifold $\mathcal{M}_K^N$ [24] is defined as the set of $N \times K$ strictly positive stochastic matrices. We consider the structure of the multinomial manifold $\mathcal{M}_K^N$ as an embedded submanifold of $\mathbb{R}^{N \times K}$ equipped with the Fisher information metric. The Fisher information metric gives the multinomial manifold a differentiable structure. With this metric, points on its boundary are now infinitely far away. Table I contains all the ingredients required by the state of the art trust-region solver of Manopt [7]. These include the definition of Fisher information metric $g_X(U, V)$ for two elements of U, V of the tangent space of $\mathcal{M}_K^N$ at a point $X \in \mathcal{M}_K^N$, denoted by $T_X \mathcal{M}_K^N$, the orthogonal projection onto $T_X \mathcal{M}_K^N$, a first-order retraction $R_X(U)$ and the corresponding vector transport $\mathcal{T}_{tU} V$ along $R_X(tU)$. Furthermore, Table I includes the formulas for computing the Riemannian gradient, $\operatorname{grad} f(X)$, and the Riemannian Hessian $\operatorname{Hess} f(X)[U]$ of a real-valued function f defined on the multinomial manifold from their Euclidean counterparts $\operatorname{Grad} f(X)$ and its differential $D \operatorname{Grad} f(X)[U]$.

Let $\widehat{m}_{X^k}(U)$ denote the quadratic model of the objective around the current estimate X^k at iteration k, which is given

by

$$\widehat{m}_{X^k}(U) = \phi_\lambda(X^k) + g_{X^k}(\operatorname{grad}\phi_\lambda(X^k), U) + (1/2)g_{X^k}(\operatorname{Hess}\phi_\lambda(X^k)[U], U). \quad (17)$$

The basic trust-region subproblem at iteration k consists of approximately minimizing the aforementioned quadratic model of the objective around the current estimate X^k, that is

$$\underset{U \in T_{X^k}\mathcal{M}_K^N : \|U\|^2 \leq r^2}{\text{minimize}} \quad \widehat{m}_{X^k}(U), \quad (18)$$

where r is the trust-region radius, using the truncated conjugate gradient method [1] which is terminated after a fixed number of iterations. After the trust-region subproblem has been approximately solved and the optimal solution U^k has been obtained, the new estimate $X^+ = R_{X^k}(U^k)$ is either accepted or rejected based on the following measure of accuracy of the quadratic model:

$$\tau^k \doteq \frac{\phi_\lambda(X^k) - \phi_\lambda(R_{X^k}(U^k))}{\widehat{m}_{X^k}(0) - \widehat{m}_{X^k}(U^k)}. \quad (19)$$

This procedure is outlined in Algorithm 2. We pick an initial point at random and we use the default value $\tau' = 0.1$ while $r_{\max}$ depends on the dimension of the problem. We refer the reader to [1] for more details about trust-region methods. The only missing ingredient is the computation of the Euclidean gradient and Hessian of the objective ϕ_λ which can be found in the Appendix.

Algorithm 2 MatchRTR

Input: $\{\widetilde{X}_{ij}\}_{i,j=1}^m$, $X^0 \in \mathcal{M}_K^N$, universe size K, parameters τ', $r_{\max}$
Output: Permutations $\{X_i\}_{i=1}^m$, consistent $\{X_{ij} \doteq X_i X_j^T\}_{i,j=1}^m$
repeat
  Compute U^k by solving (18) using truncated conjugate gradient
  if $\tau^k > \tau'$ **then**
    $X^{k+1} = R_{X^k}(U^k)$
  else
    $X^{k+1} = X^k$
  end if
  if $\tau^k < 1/4$ **then**
    $r = r/4$
  else if $\tau^k > 3/4$ and $\|U^k\| = r$ **then**
    $r = \min\{2r, r_{\max}\}$
  end if
until Convergence
Discretize with Hungarian algorithm.

C. Computational Complexity

To analyze the complexity of MatchRTR, we use the simple fact [9] that the cost of matrix multiplication of an $m \times n$ with an $n \times p$ matrix is $2mnp = O(mnp)$ flops (floating point operations). All manifold operations of Table I require $O(NK)$ flops So, the main computational burden consists of the evaluation of the objective ϕ_λ, its Euclidean gradient $\operatorname{Grad}\phi_\lambda$ and its Hessian $D\operatorname{Grad}\phi_\lambda$. Assuming that $\widetilde{\mathbf{X}}$ is relatively sparse, specifically that each row has $O(m)$ nonzero elements, these three quantities can be computed in $O(NK^2 + mNK)$ flops. Each iteration of the competing methods MatchLift [10] and MatchALS [33], has complexity of $O(N^3)$ and $O(N^2K)$, respectively. Since $m \ll N$ and $K \ll N$ this is a significant improvement. MatchRTR requires fewer iterations, on average, since second-order information about the objective is used.

VII. Experimental Results

Synthetic data. First, we evaluate the performance of the proposed methods using synthetic data. We adopt a synthetic experimental setup similar to the one used in [10], [33]. We compare against the spectral method [23], the semidefinite programming approach MatchLift [10] and the rank-minimization approach MatchALS [33]. We fix the number of features in each collection as $n = 20$. We vary the number of collections from 5 to 50 and the pairwise matching error rate from 0.1 to 0.9. To quantify the performance of each method, we use the F-score, i.e. the harmonic mean of precision p and recall r, given by $F = 2pr/(p+r)$. The output error we plot in Fig. 1 is then $1 - F$.

Multiimage feature matching. Next, we compare the same methods in a multiimage matching scenario. First, we use the CMU Hotel[1] and House[2] datasets, Although simple, these datasets have been used in all prior works. The House sequence contains 111 images and the Hotel sequence contains 101. We also include results on the first halfs of the two sequences. In addition, we evaluate using the Affine Covariant Regions Datasets [3] (Wall, UBC, Bikes, Leuven, Trees, Graffiti, Bark) which consist of sequences of 6 images with significant overlap but with viewpoint, scale and image quality variability. To obtain pairwise correspondences, we extract SIFT descriptors [20] and use SIFT matching.

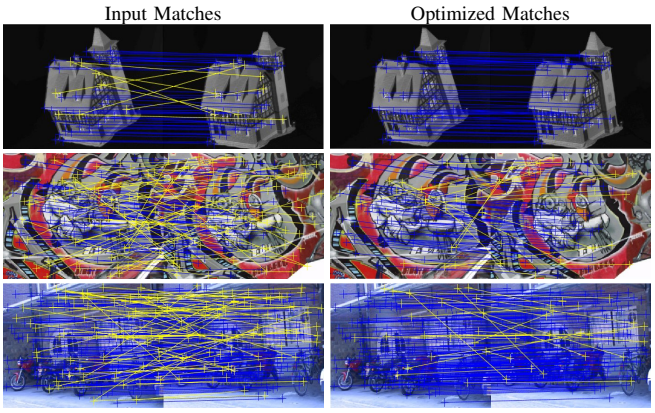

Fig. 2: Qualitative results on the CMU and Affine Covariant Regions datasets obtained by the proposed methods. Blue and yellow lines correspond to inliers and outliers, respectively.

[1] http://vasc.ri.cmu.edu//idb/html/motion/hotel/
[2] http://vasc.ri.cmu.edu//idb/html/motion/house
[3] http://www.robots.ox.ac.uk/~vgg/data/data-aff.html

Definition	$\mathcal{M}_K^N = \{X \in \mathbb{R}^{N\times K} : X > 0, X\mathbf{1} = \mathbf{1}\}$
Tangent Space	$T_X\mathcal{M}_K^N = \{U \in \mathbb{R}^{N\times K} : U\mathbf{1} = \mathbf{0}\}$
Riemannian Metric	$g_X(U,V) = \sum_{i,j}(U)_{ij}(V)_{ij}/(X)_{ij}$
Projection	$\Pi_X(U) = U - (U\mathbf{1}\mathbf{1}^T) \odot X$
Retraction	$R_X(U) = \big(X \odot \exp(U \oslash X)\big) \oslash \big((X \odot \exp(U \oslash X))\mathbf{1}\mathbf{1}^T\big)$
Vector Transport	$\mathcal{T}_{tU}V = \Pi_{R_X(tU)}(V)$
Riemannian Gradient	$\operatorname{grad} f(X) = \Pi_X(\operatorname{Grad} f(X) \odot X)$
Riemannian Hessian	$\operatorname{Hess} f(X)[U] = \Pi_X\left(D\operatorname{Grad} f(X)[U] \odot X + (1/2)\left(\operatorname{Grad} f(X) - (\operatorname{Grad} f(X) \odot X)11^T\right) \odot U\right)$

TABLE I: Necessary ingredients for optimization on the multinomial manifold. The Hadamard product is denoted by $\odot$, the elementwise division and elementwise exponentiation are denoted by $\oslash$ and $\exp(\cdot)$, respectively.

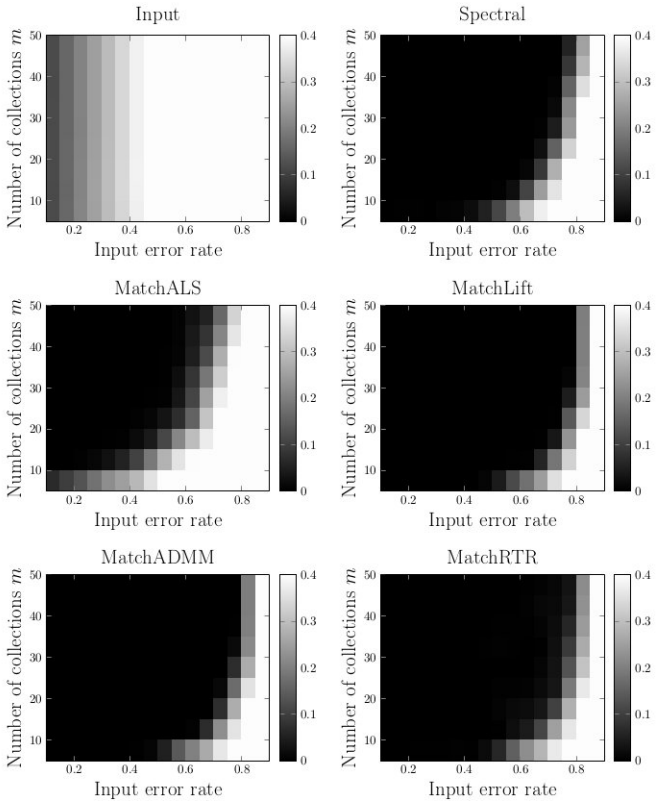

Fig. 1: Comparison of the proposed approach with the spectral method [23], MatchLift [10] and MatchALS [33], under various input error rates and number of collections m. The darker the area the smaller the output error rate. It can be seen that our approach clearly outperforms the spectral method and MatchALS, and has similar accuracy with MatchLift which has, though, significantly higher computational complexity.

VIII. Conclusions

In this work, we proposed two approximate low-rank factorization approaches for the problem of multiway matching subject to cycle consistency. Both demonstrated state of the art results in a multiimage feature matching setting and robustness with respect to the percentage of erroneous pairwise correspondences.

Dataset	MatchRTR	MatchADMM	MatchALS	MatchLift	Spectral	Input
Hotel101	0.899	**0.946**	0.933	0.747	0.693	0.726
House111	0.935	**0.971**	0.960	0.822	0.824	0.793
Hotel51	0.950	**0.965**	0.906	0.929	0.843	0.852
House56	0.992	**1**	0.994	0.986	0.970	0.922
Wall	**0.578**	0.573	0.543	0.558	0.467	0.519
UBC	**0.886**	0.868	0.873	0.688	0.747	0.837
Bikes	**0.893**	0.870	0.862	0.677	0.706	0.836
Leuven	**0.883**	0.878	0.853	0.789	0.712	0.827
Trees	**0.753**	0.726	0.688	0.718	0.541	0.648
Graffiti	**0.502**	0.496	0.465	0.464	0.375	0.456
Bark	**0.471**	0.421	0.361	0.420	0.342	0.376

TABLE II: Comparison of the competing methods on the CMU and on Affine Covariant Regions datasets. We report the F-score for each method. The proposed methods outperform the state of the art in these datasets.

Appendix

Gradient and Hessian for X-Update of MatchADMM. Let $\phi_\rho^k(X) \doteq \phi(X) + (\rho^k/2)\|X - Z^k + U^k\|_F^2$. The gradient and Hessian of ϕ_ρ^k can be computed by

$$\operatorname{grad}\phi_\rho^k(X) = XX^TX - \widetilde{\mathbf{X}}X + \rho^k(X - Z^k + U^k), \tag{20}$$

$$\operatorname{Hess}\phi_\rho^k(X)[V] = VX^TX + 2X\operatorname{sym}(X^TV) - \widetilde{\mathbf{X}}V + \rho^kV. \tag{21}$$

Euclidean Gradient and Hessian for MatchRTR objective. The Euclidean gradient and Hessian of ϕ_λ can be computed by

$$\operatorname{Grad}\phi_\lambda(X) = -(\widetilde{\mathbf{X}} + \lambda I - XX^T)X + \lambda\sum_{i=1}^m J_iX_iX_i^TX_i, \tag{22}$$

and the corresponding Hessian can be computed by

$$D\operatorname{Grad}\phi_\lambda(X)[U] = -(\widetilde{\mathbf{X}} + \lambda I)U + UX^TX + \tag{23}$$

$$2X\operatorname{sym}(X^TU) + \lambda\sum_{i=1}^m J_i\left(U_iX_i^TX_i + 2X_i\operatorname{sym}(X_i^TU_i)\right), \tag{24}$$

where $\operatorname{sym}(\cdot)$ denotes the symmetric part of a matrix and J_i is a properly defined $N \times n_i$ matrix that injects $X_iX_i^TX_i$ to the ith block of the gradient.

REFERENCES

[1] P.-A. Absil, R. Mahony, and R. Sepulchre. *Optimization algorithms on matrix manifolds*. Princeton University Press, 2009.

[2] R. Aragues, E. Montijano, and C. Sagues. Consistent data association in multi-robot systems with limited communications. In *Robotics: Science and Systems*, pages 97–104, 2011.

[3] F. Åström, S. Petra, B. Schmitzer, and C. Schnörr. Image labeling by assignment. *Journal of Mathematical Imaging and Vision*, pages 1–28, 2016.

[4] S. Belongie, J. Malik, and J. Puzicha. Shape context: A new descriptor for shape matching and object recognition. In *Advances in neural information processing systems*, pages 831–837, 2001.

[5] N. Boumal and P. Absil. Rtrmc: A riemannian trust-region method for low-rank matrix completion. *Advances in Neural Information Processing Systems*, 24:406–414, 2011.

[6] N. Boumal and P.-A. Absil. Low-rank matrix completion via preconditioned optimization on the grassmann manifold. *Linear Algebra and its Applications*, 475:200–239, 2015.

[7] N. Boumal, B. Mishra, P.-A. Absil, R. Sepulchre, et al. Manopt, a matlab toolbox for optimization on manifolds. *Journal of Machine Learning Research*, 15(1):1455–1459, 2014.

[8] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Found. Trends Mach. Learn.*, 3(1):1–122, Jan. 2011.

[9] S. Boyd and L. Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

[10] Y. Chen, L. Guibas, and Q. Huang. Near-optimal joint object matching via convex relaxation. In *International Conference on Machine Learning*, 2014.

[11] A. Cunningham, K. M. Wurm, W. Burgard, and F. Dellaert. Fully distributed scalable smoothing and mapping with robust multi-robot data association. In *Proceedings of the IEEE International Conference on Robotics and Automation*, pages 1093–1100, 2012.

[12] D. Fox, J. Ko, K. Konolige, B. Limketkai, D. Schulz, and B. Stewart. Distributed multirobot exploration and mapping. *Proceedings of the IEEE*, 94(7):1325–1339, 2006.

[13] Q.-X. Huang and L. Guibas. Consistent shape maps via semidefinite programming. *Computer Graphics Forum*, 32(5):177–186, 2013.

[14] Q.-X. Huang, G.-X. Zhang, L. Gao, S.-M. Hu, A. Butscher, and L. Guibas. An optimization approach for extracting and encoding consistent maps in a shape collection. *ACM Transactions on Graphics*, 31(6):167, 2012.

[15] M. Journée, F. Bach, P.-A. Absil, and R. Sepulchre. Low-rank optimization on the cone of positive semidefinite matrices. *SIAM Journal on Optimization*, 20(5):2327–2351, 2010.

[16] V. G. Kim, W. Li, N. J. Mitra, S. DiVerdi, and T. A. Funkhouser. Exploring collections of 3d models using fuzzy correspondences. *ACM Transactions on Graphics*, 31(4):54, 2012.

[17] H. W. Kuhn. The hungarian method for the assignment problem. *Naval research logistics quarterly*, 2(1-2):83–97, 1955.

[18] G. Lebanon and J. Lafferty. Hyperplane margin classifiers on the multinomial manifold. In *International Conference on Machine Learning*.

[19] S. Leonardos, X. Zhou, and K. Daniilidis. Distributed consistent data association via permutation synchronization. In *Robotics and Automation (ICRA), 2017 IEEE International Conference on*, pages 2645–2652. IEEE, 2017.

[20] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.

[21] E. Maset, F. Arrigoni, and A. Fusiello. Practical and efficient multi-view matching. In *Proceedings of IEEE International Conference on Computer Vision*, volume 2, page 5, 2017.

[22] A. Nguyen, M. Ben-Chen, K. Welnicka, Y. Ye, and L. Guibas. An optimization approach to improving collections of shape maps. *Computer Graphics Forum*, 30(5):1481–1491, 2011.

[23] D. Pachauri, R. Kondor, and V. Singh. Solving the multi-way matching problem by permutation synchronization. In *Advances in Neural Information Processing Systems*, pages 1860–1868, 2013.

[24] Y. Sun, J. Gao, X. Hong, B. Mishra, and B. Yin. Heterogeneous tensor decomposition for clustering via manifold optimization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(3):476–489, 2016.

[25] R. Tron, X. Zhou, C. Esteves, and K. Daniilidis. Fast multi-image matching via density-based clustering. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4057–4066, 2017.

[26] A. Vedaldi and S. Soatto. Quick shift and kernel methods for mode seeking. *Computer vision–ECCV 2008*, pages 705–718, 2008.

[27] S. B. Williams, G. Dissanayake, and H. Durrant-Whyte. Towards multi-vehicle simultaneous localisation and mapping. In *Proceedings of the IEEE International Conference on Robotics and Automation*, volume 3, pages 2743–2748, 2002.

[28] J. Yan, M. Cho, H. Zha, X. Yang, and S. M. Chu. Multi-graph matching via affinity optimization with graduated consistency regularization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(6):1228–1242, 2016.

[29] J. Yan, Y. Li, W. Liu, H. Zha, X. Yang, and S. M. Chu. Graduated consistency regularized optimization for multi graph matching. In *European Conference on Computer Vision*, pages 407–422. Springer, 2014.

[30] J. Yan, Y. Tian, H. Zha, X. Yang, Y. Zhang, and S. M. Chu. Joint optimization for consistent multiple graph matching. In *Computer Vision (ICCV), 2013 IEEE International Conference on*, pages 1649–1656. IEEE, 2013.

[31] J. Yan, J. Wang, H. Zha, X. Yang, and S. Chu. Consistency-driven alternating optimization for multigraph matching: A unified approach. *IEEE Transactions on Image Processing*, 24(3):994–1009, 2015.

[32] C. Zach, M. Klopschitz, and M. Pollefeys. Disambiguating visual relations using loop constraints. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2010.

[33] X. Zhou, M. Zhu, and K. Daniilidis. Multi-image matching via fast alternating minimization. In *IEEE International Conference on Computer Vision*, pages 4032–4040, 2015.