EXACT COMBINATORIAL OPTIMIZATION FOR SYNCHRONIZATION OF PARTIAL MULTI-MATCHING

Anonymous authors
Paper under double-blind review

ABSTRACT

In permutation synchronization, the goal is to find globally cycle-consistent correspondences from noisy pairwise matchings. In this work, unlike spectral relaxations that embed permutations into an orthogonal space and often result in inaccuracies, we maintain the problem in its original combinatorial form. By shifting the affinity spectrum to ensure positive semidefiniteness, we cast the trace-maximization over partial permutations as a convex-in-P formulation. Our minorization-maximization scheme then replaces this with a sequence of exact linear-assignment subproblems, the row-/column-sum constraints of which are totally unimodular, guaranteeing integral solutions with no rounding. This direct, combinatorial approach delivers a monotonic objective ascent, convergence to a KKT point, and superior cycle consistency and accuracy on real image-matching benchmarks at competitive runtimes.

1 INTRODUCTION

Matching features across images or shapes is a central challenge in pattern recognition and computer vision, playing a vital role in numerous applications, ranging from learning models of shape deformation Cootes & Taylor (1992); Heimann & Meinzer (2009) to object tracking, 3D reconstruction, graph matching, and image registration. The inherent complexity of these matching problems become evident when formulated as instances of the NP-hard quadratic assignment problem (QAP) Sahni & Gonzalez (1976). Extending beyond matching pairs of objects, the broader task of matching across multiple objects is known as multi-matching. Generally, multi-matching is at least as computationally demanding as pairwise matching because it involves solving multiple interconnected pairwise problems under consistency constraints. A popular strategy for tackling multi-matching in practice is to leverage these pairwise couplings Kezurer et al. (2015); Yan et al. (2015; 2016a); Bernard et al. (2018).

Permutation synchronization has emerged as a key technique for refining matchings across multiple objects Huang & Guibas (2013); Pachauri et al. (2013), and its principles have been applied to various domains, including multi-alignment Bernard et al. (2015); Arrigoni et al. (2016); Huang et al. (2019b), multi-shape matching Huang et al. (2019a; 2020); Gao et al. (2021), multi-image matching Zhou et al. (2015); Tron et al. (2017); Bernard et al. (2019b); Birdal & Simsekli (2019); Birdal et al. (2021), and multi-graph matching Yan et al. (2016b); Bernard et al. (2018); Swoboda et al. (2019), among others. In essence, permutation synchronization seeks to enforce cycle consistency among the set of pairwise permutation matrices that represent correspondences between points across multiple objects.

In scenarios involving full matchings, cycle-consistency requires that the composition of matchings along any cycle yields the identity mapping. Synchronization techniques have been thoroughly explored both within the specific context of multi-matching Nguyen et al. (2011); Pachauri et al. (2013); Huang & Guibas (2013); Chen et al. (2014); Shen et al. (2016); Tron et al. (2017); Maset et al. (2017); Schiavinato & Torsello (2017); Kahl et al. (2024; 2025) and for broader types of transformations Govindu (2004); Hadani & Singer (2011); Singer & Shkolnisky (2011); Chatterjee & Govindu (2013); Bernard et al. (2015); Arrigoni et al. (2017); Thunberg et al. (2017); Wang & Singer (2013). Synchronization can be interpreted as a denoising step: it tries to eliminate incorrect pairwise matchings, which manifest as cycle inconsistencies, thereby improving the overall correspondence quality.

Typically, synchronizing pairwise matchings is formulated as an optimization problem over permutation matrices. Notably, Pachauri et al. Pachauri et al. (2013) and Shen et al. Shen et al. (2016) proposed spectral methods for synchronization. However, these approaches assume full permutation matrices, meaning that all features must exist across all objects. While Maset et al. (2017) has recently tackled this restriction, the method does not enforce cycle-consistency. Given that true matchings inherently satisfy cycle-consistency, we argue that ensuring this property is crucial.

In Bernard et al. (2021), the nonconvex problem of identifying a sparse matrix on the Stiefel manifold is considered that maximizes a quadratic form, while explicitly enforcing cycle consistency. To sidestep the combinatorial nature of the partial-permutation matching, a semi-orthogonality constraint is put in place, trading exact discreteness for tractability. Unlike traditional spectral methods which typically ignore sparsity due to their reliance on eigenvalue solvers, Bernard et al. (2021) augments the orthogonal iteration algorithm with a sparsity-encouraging step, thereby attaining sparse solutions that are globally optimal under the relaxed constraint. Nevertheless, this approach only yields "soft" sparsity (most entries are nearly zero but not exactly zero); in addition, since the sparsity-promoting objective remains nonconvex, global optimality cannot be assured. Furthermore, by relaxing the problem away from the exact partial-permutation constraints, this approach cannot produce strictly binary (0–1) correspondences, which undermines the precision of the recovered matches.

Recent advancements have focused on local search heuristics to tackle the combinatorial complexity. For instance, Kahl et al. (2024) introduces a powerful local search framework for graph matching (GM-LS) that can be extended to multi-graph matching through a sequential construction process. This approach was further accelerated in Kahl et al. (2025) by parallelizing the construction and local search phases. While these methods achieve state-of-the-art performance, due to their heuristic nature, they do not provide any theoretical proof of convergence. Furthermore, GREEDA Kahl et al. (2025), which combines distinct local search modules, faces a critical limitation: because each module optimizes locally, their combination does not guarantee that the final solution will achieve an acceptable level of performance, as it may converge to a suboptimal local minimum.

1.1 Main Contribution

Below we summarize the key innovations of this work:

(i) **Direct combinatorial formulation.** We retain the problem in its original permutation domain, eschewing relaxations into orthogonal spaces, and show that shifting the affinity spectrum by its minimum eigenvalue yields an equivalent, positive-semidefinite trace-maximization over partial permutation matrices.

(ii) **Minorization–maximization with exact subproblems.** We introduce an MM framework that, at each iteration, constructs a tight linear surrogate of the convexified objective and solves it exactly via a linear-assignment problem, ensuring the global optimum of each surrogate step.

(iii) **Total unimodularity guarantee.** We prove that the combined row- and column-sum constraints form a totally unimodular system, so the LP relaxation of each surrogate admits only integral extreme points, eliminating the need for any heuristic rounding and preserving exact partial permutations.

(iv) **Monotonic ascent & convergence.** We prove that the algorithm monotonically increases the original trace objective at every MM iteration and converges to a stationary point of the combinatorial formulation.

(v) **Superior accuracy & cycle consistency.** By operating directly on permutations, we achieve state-of-the-art cycle consistency and matching accuracy on real image-matching benchmarks, outperforming spectral and alternating-minimization baselines.

(vi) **Highly efficient and scalable runtimes.** Each iteration reduces to one or more efficient linear-assignment solvings, yielding faster overall runtimes than existing methods even on large-scale, real-world datasets.

2 MM FRAMEWORK: AN OVERVIEW

Consider the constrained optimization problem

$$\max_{\mathbf{x}\in\chi} f(\mathbf{x}), \tag{1}$$

where $\mathbf{x}$ denotes the decision variable, $f(\mathbf{x})$ is the objective to be maximized, and χ represents the feasible region. An MM-based method tackles (1) by introducing, at each iteration t, a surrogate function $g\big(\mathbf{x} \mid \mathbf{x}^t\big)$, which underestimates $f(\mathbf{x})$ but matches it exactly at the current point $\mathbf{x}^t$. The next iterate is then found by solving $\mathbf{x}^{t+1} \in \arg\max_{\mathbf{x}\in\chi} g\big(\mathbf{x} \mid \mathbf{x}^t\big)$. These two operations—surrogate construction and maximization—are repeated until convergence to a stationary solution of (1).

For $g(\mathbf{x} \mid \mathbf{x}^t)$ to qualify as a valid minorizer, it must satisfy

$$g\big(\mathbf{x} \mid \mathbf{x}^t\big) \le f(\mathbf{x}) \quad \forall\, \mathbf{x} \in \chi, \tag{2}$$

$$g\big(\mathbf{x}^t \mid \mathbf{x}^t\big) = f\big(\mathbf{x}^t\big). \tag{3}$$

As a result of the surrogate properties, each MM step yields

$$f(\mathbf{x}^{t+1}) \;\ge\; g\big(\mathbf{x}^{t+1} \mid \mathbf{x}^t\big) \;\ge\; g\big(\mathbf{x}^t \mid \mathbf{x}^t\big) \;=\; f(\mathbf{x}^t),$$

which shows the objective value never decreases, ensuring the sequence $f(\mathbf{x}^t)$ converges to a KKT point of (1). To see a more detailed explanation of the MM framework, please refer to Sun et al. (2017).

3 PROBLEM FORMULATION

Let k be the number of objects, where object i comprises of m_i points. Denote by $\mathbb{1}_p$ the p-dimensional all-ones vector, and interpret vector inequalities entrywise. For each pair (i, j), let $\mathbf{P}_{ij} \in \mathbb{P}_{m_i m_j} := \big\{\mathbf{X} \in \{0,1\}^{m_i \times m_j} : \mathbf{X}\mathbb{1}_{m_j} \le \mathbb{1}_{m_i}, \mathbf{X}^T\mathbb{1}_{m_i} \le \mathbb{1}_{m_j}\big\}$ be the partial permutation matrix encoding correspondences between the m_i points of object i and the m_j points of object j. When these matrices are full bijections, the collection $\mathcal{P} = \{\mathbf{P}_{ij}\}_{i,j=1}^k$ is called *cycle-consistent* if for every triplets (i, ℓ, j) it holds that $\mathbf{P}_{i\ell}\,\mathbf{P}_{\ell j} \;=\; \mathbf{P}_{ij}$. We denote by $\overline{\mathbb{P}}_{m_i d}$ the subset of $\mathbb{P}_{m_i d}$ whose members have full row-rank, i.e. $\overline{\mathbb{P}}_{m_i d} = \big\{\mathbf{X} \in \mathbb{P}_{m_i d} :\; \mathbf{X}\,\mathbb{1}_d \;=\; \mathbb{1}_{m_i},\; \mathbf{X}^T\,\mathbb{1}_{m_i} \;\le\; \mathbb{1}_d\big\}$, where d is the total number of distinct points across all objects. We further note that cycle consistency among the pairwise maps $\{\mathbf{P}_{ij}\}$ holds if and only if there exist "object-to-universe" matchings $\mathcal{U} = \{\mathbf{P}_i \in \overline{\mathbb{P}}_{m_i d}\}_{i=1}^k$ such that $\mathbf{P}_{ij} = \mathbf{P}_i\,\mathbf{P}_j^T \quad \forall\, i, j$. This universe-matching characterization remains valid even when the $\mathbf{P}_{ij}$ are only partial (non-bijective) permutations (see Tron et al. (2017); Bernard et al. (2019a) for more details). Given the noisy set of pairwise permutations $\mathcal{P} = \{\mathbf{P}_{ij}\}_{i,j=1}^k$, permutation synchronization can be formulated as

$$\underset{\{\mathbf{P}_i \in \overline{\mathbb{P}}_{m_i d}\}}{\arg\max} \sum_{i,j} \operatorname{tr}\left(\mathbf{P}_{ij}^T \mathbf{P}_i \mathbf{P}_j^T\right) \Leftrightarrow \underset{\mathbf{P}\in\mathcal{U}}{\arg\max} \operatorname{tr}\left(\mathbf{P}^T\mathbf{W}\mathbf{P}\right), \tag{4}$$

where, for $m := \sum_{i=1}^k m_i$, we define

$$\mathbb{U} := \overline{\mathbb{P}}_{m_1 d} \times \cdots \times \overline{\mathbb{P}}_{m_k d} \;\subset\; \mathbb{R}^{m\times d}, \quad \mathbf{P} = \begin{bmatrix} \mathbf{P}_1^T \\ \vdots \\ \mathbf{P}_k^T \end{bmatrix} \;\in\; \mathbb{R}^{m\times d}, \quad \mathbf{W} := [\mathbf{P}_{ij}]_{i,j=1}^k \;\in\; \mathbb{R}^{m\times m}.$$

With the aforementioned notations, the problem in (4) can be compactly rewritten as:

$$\boxed{\begin{aligned} \arg\max_{\mathbf{P}\in\{0,1\}} \quad & \operatorname{tr}\left(\mathbf{P}^T\mathbf{W}\mathbf{P}\right) \\ \text{s.t.} \quad & \mathbf{P}\mathbb{1}_d = \mathbb{1}_m \\ & \mathbf{P}_i^T\mathbb{1}_{m_i} \le \mathbb{1}_d, \quad i = 1, 2, \ldots k. \end{aligned}} \tag{5}$$

As we can see, the problem in (5) is challenging because the objective $\operatorname{tr}\big(\mathbf{P}^T\mathbf{W}\mathbf{P}\big)$ is a non-concave quadratic form (since $\mathbf{W}$ may be indefinite), and the binary row- and column-sum constraints make the feasible set combinatorial and NP-hard to search. In the following section, inspired by the MM approach, we show how to solve the problem without resorting to any relaxation that takes us far from optimality.

4 SOLVING THE PERMUTATION SYNCHRONIZATION PROBLEM

In order to apply MM framework to solve (5), it is necessary to have a convex form of the objective function. As a result, we introduce Lemma 4.1.

Lemma 4.1.

$$\left.\begin{array}{l}\mathbf{P} \in \{0,1\}^{m\times d}\\ \mathbf{P}\,\mathbb{1}_d = \mathbb{1}_m\end{array}\right\} \implies \|\mathbf{P}\|_F^2 = m. \tag{6}$$

Proof. By definition $\|\mathbf{P}\|_F^2 = \sum_{i=1}^{m}\sum_{j=1}^{d} P_{ij}^2$, and since each $P_{ij} \in \{0,1\}$ we have $P_{ij}^2 = P_{ij}$. Hence $\|\mathbf{P}\|_F^2 = \sum_{i,j} P_{ij}$. On the other hand, $\mathbf{P}\,\mathbb{1}_d = \mathbb{1}_m$ which implies $\sum_{i,j} P_{ij} = m$, and as a result $\sum_{i,j} P_{ij}^2 = m$ and proof is completed. □

Let us define $\mathbf{M} = \mathbf{W} - \lambda_{\min}(\mathbf{W})\,\mathbf{I}_m \succeq 0$. As a result, for any feasible $\mathbf{P}$,

$$\mathrm{tr}(\mathbf{P}^T\mathbf{M}\,\mathbf{P}) = \mathrm{tr}(\mathbf{P}^T\mathbf{W}\,\mathbf{P}) - \lambda_{\min}(\mathbf{W})\,\|\mathbf{P}\|_F^2 = \mathrm{tr}(\mathbf{P}^T\mathbf{W}\,\mathbf{P}) - m\,\lambda_{\min}(\mathbf{W}).$$

As $m\lambda_{\min}(\mathbf{W})$ is a constant, it can be dropped and we arrive at the following equivalent problem:

$$\begin{aligned}\arg\max_{\mathbf{P}}\ & \mathrm{tr}\left(\mathbf{P}^T\mathbf{M}\mathbf{P}\right)\\ \text{s.t.}\ & \mathbf{P} \in \{0,1\}^{m\times d}\\ & \mathbf{P}\mathbb{1}_d = \mathbb{1}_m\\ & \mathbf{P}_i^T\mathbb{1}_{m_i} \le \mathbb{1}_d \quad i = 1,2,...k.\end{aligned} \tag{7}$$

Since $\mathbf{M} \succeq 0$, the objective function is convex in $\mathbf{P}$; thus, at any current iterate $\mathbf{P}^t$, it can be minorized by its tangent hyperplane:

$$\mathrm{tr}(\mathbf{P}^T\mathbf{M}\,\mathbf{P}) \ \ge\ 2\ \mathrm{tr}\left((\mathbf{P}^t)^T\mathbf{M}\,\mathbf{P}\right) - \mathrm{tr}\left((\mathbf{P}^{(t)})^T\mathbf{M}\,\mathbf{P}^{(t)}\right)\ =\ g_t(\mathbf{P}).$$

Setting $\mathbf{A}^{(t)} = \mathbf{M}^T\mathbf{P}^{(t)}$ and leaving the constant term in $g_t(\mathbf{P})$, we arrive at the following surrogate maximization problem:

$$\begin{aligned}\arg\max_{\mathbf{P}}\ & \mathrm{tr}\left((\mathbf{A}^t)^T\mathbf{P}\right)\\ \text{s.t.}\ & \mathbf{P} \in \{0,1\}^{m\times d}\\ & \mathbf{P}\mathbb{1}_d = \mathbb{1}_m\\ & \mathbf{P}_i^T\mathbb{1}_{m_i} \le \mathbb{1}_d, \quad i = 1,2,...k.\end{aligned} \tag{8}$$

All constraints except $P_{ij} \in \{0,1\}$ are affine in $\mathbf{P}$; by transforming this constraint into $0 \le P_{ij} \le 1$, we obtain a convex feasible set, however, one might worry about having fractional entries in the optimal $\mathbf{P}$. The following definition and theorem adopted from Schrijver (1986) and Wolsey & Nemhauser (1999) ensure that this does not happen.

Definition 4.1 (Total unimodularity)**.** An integer matrix $\mathbf{A}$ (not necessarily square) is *totally unimodular* (TU) if every square submatrix of $\mathbf{A}$ has determinant in $\{-1,0,1\}$.

Theorem 4.2 (Integral vertices of TU systems)**.** *Let $\mathbf{D}$ be a totally unimodular matrix and $\mathbf{b}$ an integer vector. Then the polyhedron $\{\mathbf{x} \in \mathbb{R}^n : \mathbf{D}\,\mathbf{x} \le \mathbf{b}\}$ has only integral vertices. In particular, the linear program $\max\{\mathbf{B}^T\mathbf{x} : \mathbf{D}\,\mathbf{x} \le \mathbf{b}\}$ admits an optimal solution $\mathbf{x}^* \in \mathbb{Z}^n$.*

To invoke Theorem 4.2 on our partial matching problem, we begin by vectorizing the matrix $\mathbf{P} \in \mathbb{R}^{m\times d}$ into the column–stacked vector $\mathbf{x} = \mathrm{vec}(\mathbf{P}) \in \mathbb{R}^{md}$. In this form, each of the original affine constraints can be written as rows of a single integer matrix $\mathbf{D}$ acting on $\mathbf{x}$, with an integral right-hand side $\mathbf{b}$.

First, the requirement that each row of $\mathbf{P}$ sums to one, $\mathbf{P}\,\mathbb{1}_d = \mathbb{1}_m$, becomes the pair of inequalities

$$(\mathbf{I}_m \otimes \mathbb{1}_d^T)\,\mathbf{x} \ \le\ \mathbb{1}_m, \quad -(\mathbf{I}_m \otimes \mathbb{1}_d^T)\,\mathbf{x} \ \le\ -\mathbb{1}_m,$$

where $\mathbf{I}_m \otimes \mathbb{1}_d^T \in \{0,1\}^{m\times(md)}$ is the Kronecker product of the $m \times m$ identity matrix with the $1 \times d$ all-ones row vector.

Second, the cluster-capacity constraints (namely that for each of the k clusters of rows, at most one point may be assigned to each column) can be written as $(\mathbf{J} \otimes \mathbf{I}_d)\,\mathbf{x} \;\le\; \mathbb{1}_{kd}$, where $\mathbf{J} \in \{0,1\}^{k\times m}$ encodes row-to-cluster membership and $\mathbf{I}_d$ is the $d \times d$ identity. By stacking these blocks,

$$\mathbf{D} = \begin{bmatrix} \mathbf{I}_m \otimes \mathbb{1}_d^T \\ -\left(\mathbf{I}_m \otimes \mathbb{1}_d^T\right) \\ \mathbf{J} \otimes \mathbf{I}_d \end{bmatrix} \in \{0, \pm 1\}^{(2m+k\,d)\times(m\,d)}, \qquad \mathbf{b} = \begin{bmatrix} \mathbb{1}_m \\ -\,\mathbb{1}_m \\ \mathbb{1}_{k\,d} \end{bmatrix},$$

the LP relaxation of (8) can be written as

$$\max_{\mathbf{x}\in\mathbb{R}^{md}} \; \left(\mathrm{vec}(\mathbf{A}^{(t)})\right)^T \mathbf{x} \quad \text{s.t.} \quad \mathbf{D}\,\mathbf{x} \le \mathbf{b}, \quad 0 \le \mathbf{x} \le 1.$$

Since both $\mathbf{I}_m \otimes \mathbb{1}_d^T$ and $\mathbf{J} \otimes \mathbf{I}_d$ are totally unimodular, and total unimodularity is preserved under negation and row-stacking Schrijver (1986), the combined matrix $\mathbf{D}$ is TU and $\mathbf{b}$ is integral. By Theorem 4.2, every vertex of the polyhedron $\{\,\mathbf{x} : \mathbf{D}\mathbf{x} \le \mathbf{b}\}$ is integral, so any optimal solution $\mathbf{x}^*$ satisfies $\mathbf{x}^* \in \{0,1\}^{md}$. Equivalently, the LP relaxation of our matching problem admits an integral maximizer. As a result, the relaxed problem

$$\begin{aligned} \arg\max_{\mathbf{P}} \;\; & \mathrm{tr}\left((\mathbf{A}^{(t)})^T\mathbf{P}\right) \\ \text{s.t.} \;\; & \mathbf{P} \in [0,1]^{m\times d} \\ & \mathbf{P}\mathbb{1}_d = \mathbb{1}_m \\ & \mathbf{P}_i^T\mathbb{1}_{m_i} \le \mathbb{1}_d \quad i = 1,2,...k, \end{aligned} \tag{9}$$

is always tight. The optimzation problem 9 can be solved directly by CVX Grant & Boyd (2014), or CVXPY Diamond & Boyd (2016) or by the Lagrangian duality approach in Sun et al. (2017); Saini et al. (2024). However, since all constraints in (9) are linear, projection-based first-order methods yield significantly faster and more scalable solutions: starting from an initial $\mathbf{P}^{(0)} \in \mathcal{C}$, one performs the iterations $\mathbf{P}^{(t+\frac{1}{2})} = \mathbf{P}^{(t)} + \eta^{(t)}\mathbf{A}^{(t)}$, $\mathbf{P}^{(t+1)} = \Pi_{\mathcal{C}}\left(\mathbf{P}^{(t+\frac{1}{2})}\right)$, where $\mathcal{C} = \left\{\mathbf{P} \in [0,1]^{m\times d} :\; \mathbf{P}\,\mathbb{1}_d = \mathbb{1}_m,\; \mathbf{P}_i^T\,\mathbb{1}_{m_i} \le \mathbb{1}_d \;\forall i \in \{0,1,\ldots,k\}\right\}$,

and $\Pi_{\mathcal{C}}$ denotes the Euclidean projection onto $\mathcal{C}$. Each iteration costs $O(md)$ operations, and accelerated variants achieve $O(1/t^2)$ convergence rate in practice, making projection-based methods particularly attractive for large-scale instances. For a concise overview of our approach (Algorithm 1) and a detailed proof of the convergence of our method to a KKT stationary point, we refer the reader to the Appendix A.

5 Experimental Results

In this section, we conduct a rigorous and comprehensive evaluation of our proposed MM-inspired combinatorial permutation synchronization algorithm (MM). We compare its performance against a broad suite of nine state-of-the-art baselines on the challenging task of multi-image correspondence. Our analysis spans five key evaluation metrics and multiple datasets to assess accuracy, structural consistency, and computational efficiency.

5.1 Compared Methods

We benchmark our algorithm against the following established and recent methods in permutation synchronization and multi-graph matching:

- **Spectral Pachauri et al. (2013):** The foundational permutation synchronization method, which formulates the problem as a spectral relaxation that can be solved efficiently via the eigendecomposition of a global affinity matrix.
- **MatchEig Maset et al. (2017):** A non-iterative spectral method that uses the eigendecomposition of an affinity matrix to enforce correspondence consistency.
- **NmfSync Bernard et al. (2019a):** A method for partial permutation synchronization based on non-negative matrix factorization, which projects a relaxed solution to achieve cycle-consistent results.

- **stiefel Bernard et al. (2021):** Formulates the problem as a sparse quadratic optimization over the Stiefel manifold and solves it using a modified orthogonal iteration algorithm.
- **FCC Shi et al. (2021):** An efficient filtering method that uses cycle consistency statistics within match graphs to identify and remove outlier correspondences.
- **MatchFAME Li et al. (2022):** A fast and memory-efficient algorithm using Cycle-Edge Message Passing and a weighted projected power method to solve partial permutation synchronization.
- **GM-LS-seq Kahl et al. (2024):** A sequential routine that combines a construction phase with a Graph Matching Local Search (GM-LS) to find a solution.
- **GM-LS-par Kahl et al. (2025):** A parallelized version of GM-LS-seq that uses parallel construction and parallel local search to accelerate optimization.
- **GREEDA Kahl et al. (2025):** An iterative algorithm that alternates between Graph Matching Local Search (GM-LS) and Swap Local Search (SWAP-LS) until convergence to refine a solution. The SWAP-LS component improves the solution by exploring moves that exchange assignments between different nodes.

5.2 Evaluation Metrics

To provide a comprehensive and rigorous evaluation, we assess our proposed MM-based synchronization algorithm against competing methods using a suite of five distinct metrics. These are chosen to measure the accuracy, structural quality, geometric consistency, and computational efficiency.

F-score (higher is better): We use the F-score, the harmonic mean of precision (p) and recall (r), as a balanced measure of matching accuracy. It is defined as:

$$F = \frac{2\,p\,r}{p+r}, \quad \text{where} \quad p = \frac{\#\text{correct matches}}{\#\text{predicted matches}}, \quad \text{and} \quad r = \frac{\#\text{correct matches}}{\#\text{ground-truth matches}}.$$

Feature Recall: The fraction of ground-truth matches correctly identified by the algorithm, measuring its ability to recover true correspondences.

Cyclic Consistency (higher is better): A core objective of synchronization is to produce a globally consistent set of permutations. This metric directly measures the degree to which the recovered permutations $\mathbf{P}_{ij}$ satisfy the cycle-consistency constraint, i.e., $\mathbf{P}_{ik} = \mathbf{P}_{ij}\mathbf{P}_{jk}$ for any triplet (i, j, k). We report the fraction of all possible triplets that satisfy this condition, where a higher value signifies superior synchronization.

RANSAC Inlier Ratio (higher is better): To assess the practical utility of the generated matches for downstream geometric tasks like structure from motion (SfM), we compute the inlier ratio. Keypoint matches are used to estimate the relative camera pose (e.g., the fundamental matrix) between image pairs using a RANSAC-based estimator. The inlier ratio is the percentage of correspondences consistent with the estimated geometric model, where a higher ratio indicates more geometrically coherent and reliable matches.

Runtime (lower is better): We report the average execution time in seconds required for each algorithm to converge on a given problem instance. This metric is crucial for evaluating the practical scalability and efficiency of the methods.

5.3 Datasets and Protocol

Our primary evaluation is conducted on the widely recognized CMU House sequence Bernard et al. (2021), which consists of 111 images of a model house captured from different viewpoints. Following the approach of Bernard et al. (2021), we construct a series of synchronization tasks by gradually increasing the number of images, k. Specifically, we vary k from 20 to 111. For each value of k, we uniformly sample the corresponding $k \times k$ subset of the full 111×111 pairwise match matrix to form the input for each problem instance.

To further validate the robustness and generalizability of our method, we performed additional experiments on five challenging sequences from the ETH3D benchmark Schöps et al. (2019; 2017):

statue, *terrace*, *office*, *kicker*, and *electro*. These datasets feature complex scenes with significant occlusions, varying illumination, and diverse structures, providing a rigorous test for all methods. We repeated the same comprehensive evaluation for these datasets as was performed for CMU House, including all quantitative metrics and qualitative match analyses. These additional results, with a full, detailed analysis, are available in Appendix B.

5.3.1 Initialization Strategy

For all experiments, our proposed MM algorithm is initialized using a warm-start strategy. Specifically, we leverage the solution generated by the Stiefel baseline Bernard et al. (2021). The output from the Stiefel method, which is a soft-assignment matrix on the Stiefel manifold, is projected onto the set of feasible partial permutations $\{0,1\}^{m\times d}$ to provide a high-quality initial point, $\mathbf{P}^{(0)}$, for our iterative procedure. The rationale for this choice stems from the fact that the `stiefel` method solves a continuous relaxation of the original combinatorial problem. Although its formulation on the Stiefel manifold does not enforce the discrete permutation constraints directly, it represents the closest relaxed version of our problem. By leveraging its solution as an initial guess, we begin our MM iterations from a point already situated in a promising region of the search space. This approach effectively combines the broad perspective of a strong relaxation method with the exact, combinatorial refinement of our MM framework, promoting convergence to a high-quality stationary point.

5.4 Implementation Details and Hardware

To ensure reproducible results, we used the publicly available code for all baselines. Our method (MM) and the MATLAB-based baselines—Spectral Pachauri et al. (2013), MatchEig Maset et al. (2017), NmfSync Bernard et al. (2019a), stiefel Bernard et al. (2021), FCC Shi et al. (2021), and MatchFAME Li et al. (2022)—were executed in MATLAB R2022b. The remaining methods, GM-LS-seq Kahl et al. (2024), GM-LS-par Kahl et al. (2025), and GREEDA Kahl et al. (2025), were run using their original Python implementations. All experiments were conducted on a desktop computer with an AMD Ryzen 5 5600H CPU and 16GB of RAM.

5.5 Quantitative Results

In this part, we present the mean performance of all evaluated algorithms on the CMU House dataset, aggregated over problem sizes from $k = 20$ to $k = 111$, as detailed in Table 1. The results unequivocally highlight the superior performance of our proposed MM algorithm, which uniquely achieves state-of-the-art accuracy while maintaining competitive efficiency. In terms of accuracy, our MM method is the only algorithm to secure a perfect F-score (1.000) and Feature Recall (1.000), indicating perfect matching. MatchFAME (0.999) and GM-LS-par (0.999) deliver near-perfect accuracy, establishing themselves as strong competitors. In contrast, traditional relaxation methods show a noticeable degradation, with Stiefel achieving an F-score of 0.984, NmfSync 0.939, and Spectral lagging at 0.863. Regarding structural integrity, our method obtains a perfect Cyclic Consistency score of 1.000, a property shared by most baselines. The exceptions are MatchEig (0.933) and FCC (0.913), whose formulations do not enforce this constraint. For geometric utility, the RANSAC Inlier Ratio reveals that local search methods (GM-LS-seq, GM-LS-par, GREEDA) produce the most geometrically coherent matches, leading with a ratio of 0.781. Our MM algorithm follows closely with a ratio of 0.724, on par with MatchFAME (0.721) and FCC (0.722), confirming the high geometric quality of its perfect correspondences. The classic relaxation methods again lag, with inlier ratios below 0.70, indicating their errors are more geometrically disruptive. Finally, in computational efficiency, GM-LS-seq is the fastest at 0.129 seconds. Our MM algorithm is also exceptionally efficient, with a mean runtime of 1.662 seconds. This result is remarkable, demonstrating that our perfect accuracy does not come at a high computational cost; it is significantly faster than other top-tier methods like GM-LS-par (2.748s), Stiefel (5.378s), and MatchFAME (13.935s), and is orders of magnitude faster than the slowest method, FCC (91.658s). In summary, the quantitative data demonstrates that our MM algorithm occupies a unique position: it delivers the highest possible accuracy and perfect consistency while maintaining excellent computational efficiency.

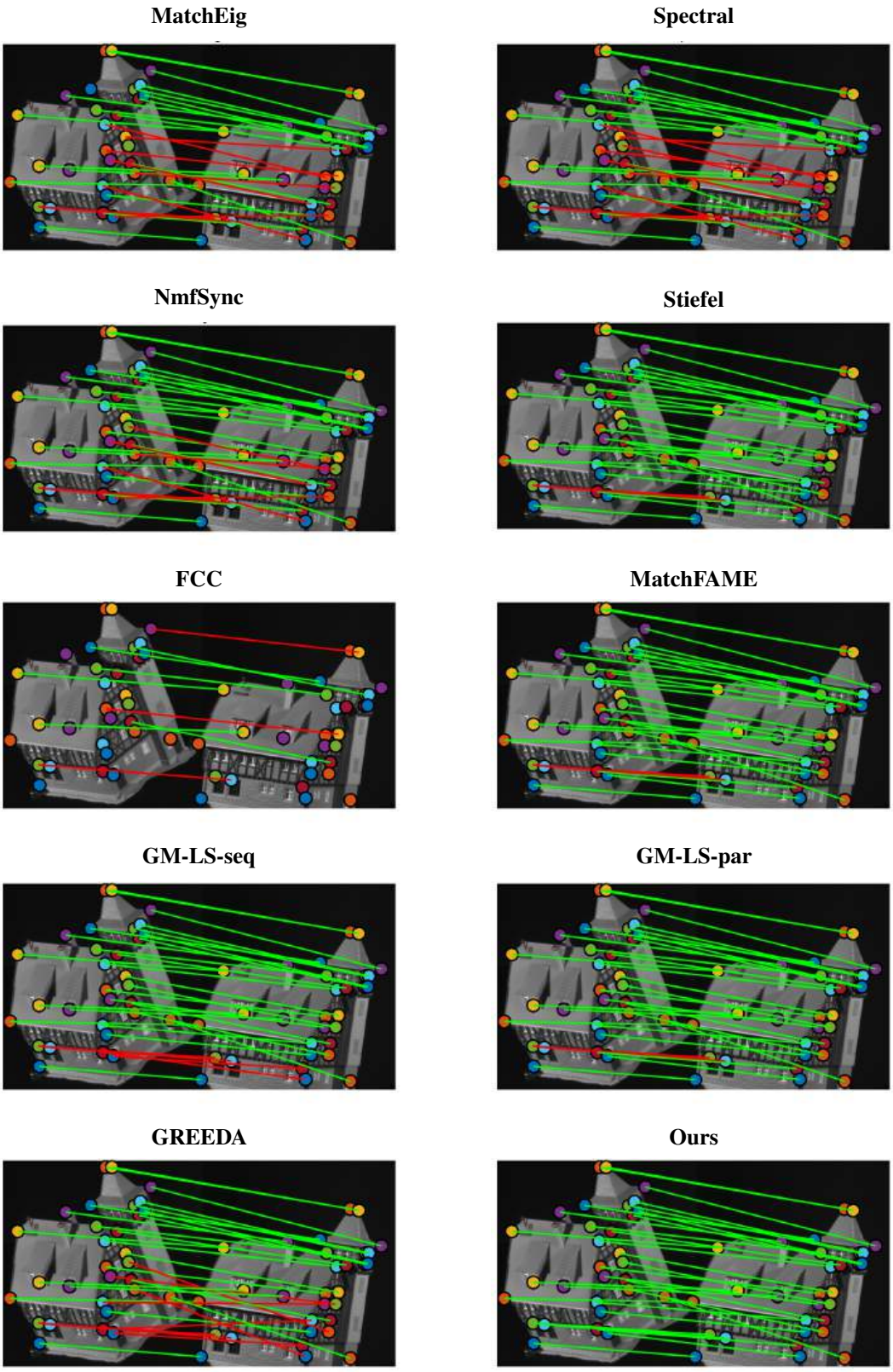

Figure 1: Comparison of matchings between the first and last image of the CMU house sequence obtained by different methods.

Table 1: Mean performance over k on the CMU house dataset.

Method	F-score	Feature Recall	Cyclic Const.	Inlier Ratio	Runtime (sec)
MatchEig	0.926	0.905	0.933	0.686	17.425
Spectral	0.863	0.864	1.000	0.638	0.869
NmfSync	0.939	0.940	1.000	0.686	11.031
Stiefel	0.984	0.984	1.000	0.697	5.378
MatchFAME	0.999	0.999	1.000	0.721	13.935
FCC	0.913	0.857	0.913	0.722	91.658
GM-LS-seq	0.965	0.965	1.000	0.781	0.129
GM-LS-par	0.999	0.999	1.000	0.781	2.748
GREEDA	0.910	0.910	1.000	0.781	1.811
MM (Ours)	1.000	1.000	1.000	0.724	1.662

5.6 Qualitative Results

To complement the quantitative analysis, Figure 1 provides a visual comparison of the correspondences found between the first and last frames of the 111-image CMU House sequence. In these visualizations, the colored dots represent the same set of ground-truth keypoints, where corresponding points share an identical color across the two views. Correct matches are denoted by green lines and mismatches by red lines, revealing the distinct structural quality and error patterns of each algorithm.

The results for our proposed MM algorithm show a complete set of correct correspondences, with a total absence of mismatches. This visual confirmation, which corroborates our perfect F-score, highlights the robustness of our combinatorial approach in resolving ambiguities, even in challenging regions with repetitive textures like the window frames. In contrast, even the top-performing baselines show visible imperfections. While the results for MatchFAME and GM-LS-par align with their high quantitative scores, they do not achieve the error-free result of our method. Other high-performing methods like GM-LS-seq and GREEDA begin to exhibit a few visible red mismatches, often clustered in the lower portion of the house, which suggests that while their solutions are largely correct, they can falter in areas of lower texture or geometric ambiguity. Among the classic relaxation-based methods, Stiefel delivers a strong result but still shows one or two minor, non-systematic errors, highlighting the inherent risk of small deviations from the discrete solution space that such methods face. In stark contrast, other baselines exhibit significant and systematic failures. The Spectral method clearly struggles with the symmetric structure of the house, producing a cluster of prominent red mismatches around the windows—a classic failure mode where eigenvector ambiguity leads to incorrect assignments. MatchEig suffers from numerous, scattered errors that create a visually "jittery" effect, a direct consequence of its lack of a cyclic consistency constraint which allows local inaccuracies to propagate globally. NmfSync and FCC also display several major mismatches, with some red lines spanning large distances across the image, indicating that their underlying factorization or compositional models can latch onto spurious correlations and produce structurally unsound results. Overall, this qualitative analysis reinforces our quantitative findings, visually demonstrating that our MM algorithm's ability to preserve the discrete problem structure allows it to achieve a level of accuracy and structural integrity that the other methods cannot match.

6 Conclusion

In this work, we have introduced a direct, combinatorial minorization-maximization framework for permutation synchronization that operates entirely in the space of partial permutation matrices. By shifting the spectrum of the affinity matrix to enforce positive semidefiniteness and constructing tight linear surrogates at each iteration, our method reduces to a sequence of exact linear-assignment subproblems whose constraints are totally unimodular, guaranteeing integral, cycle-consistent matchings without any rounding. We proved monotonic ascent of the original trace objective and convergence to a stationary point, and demonstrated on real image-matching benchmarks that our algorithm achieves state-of-the-art accuracy and consistency while running faster than existing spectral and alternating-minimization approaches. The simplicity, efficiency, and strong empirical performance of our approach make it an attractive candidate for a wide range of matching and alignment tasks.

REFERENCES

Federica Arrigoni, Beatrice Rossi, and Andrea Fusiello. Spectral synchronization of multiple views in SE(3). *SIAM Journal on Imaging Sciences*, 9(4):1963–1990, 2016. doi: 10.1137/16M1060248. URL https://doi.org/10.1137/16M1060248.

Federica Arrigoni, Eleonora Maset, and Andrea Fusiello. Synchronization in the symmetric inverse semigroup. In *Image Analysis and Processing-ICIAP 2017: 19th International Conference, Catania, Italy, September 11-15, 2017, Proceedings, Part II 19*, pp. 70–81. Springer, 2017.

Florian Bernard, Johan Thunberg, Peter Gemmar, Frank Hertel, Andreas Husch, and Jorge Goncalves. A solution for multi-alignment by transformation synchronisation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015.

Florian Bernard, Christian Theobalt, and Michael Moeller. DS*: Tighter lifting-free convex relaxations for quadratic matching problems. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.

Florian Bernard, Johan Thunberg, Jorge Goncalves, and Christian Theobalt. Synchronisation of partial multi-matchings via non-negative factorisations. *Pattern Recognition*, 92:146–155, 2019a.

Florian Bernard, Johan Thunberg, Paul Swoboda, and Christian Theobalt. HiPPI: Higher-order projected power iterations for scalable multi-matching. In *Proceedings of the ieee/cvf international conference on computer vision*, pp. 10284–10293, 2019b.

Florian Bernard, Daniel Cremers, and Johan Thunberg. Sparse quadratic optimisation over the stiefel manifold with application to permutation synchronisation. *Advances in Neural Information Processing Systems*, 34:25256–25266, 2021.

Dimitri Bertsekas, Angelia Nedic, and Asuman Ozdaglar. *Convex analysis and optimization*, volume 1. Athena Scientific, 2003.

Tolga Birdal and Umut Simsekli. Probabilistic permutation synchronization using the Riemannian structure of the Birkhoff polytope. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 11097–11108, 2019. doi: 10.1109/CVPR.2019.01136.

Tolga Birdal, Vladislav Golyanik, Christian Theobalt, and Leonidas J Guibas. Quantum permutation synchronization. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 13122–13133, 2021.

Avishek Chatterjee and Venu Madhav Govindu. Efficient and robust large-scale rotation averaging. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, December 2013.

Yuxin Chen, Leonidas Guibas, and Qixing Huang. Near-optimal joint object matching via convex relaxation. In *International Conference on Machine Learning*, pp. 100–108. PMLR, 2014.

Timothy F Cootes and Christopher J Taylor. Active shape models—'smart snakes'. In *BMVC92: Proceedings of the British Machine Vision Conference, organised by the British Machine Vision Association 22–24 September 1992 Leeds*, pp. 266–275. Springer, 1992.

Steven Diamond and Stephen Boyd. CVXPY: A Python-embedded modeling language for convex optimization. *Journal of Machine Learning Research*, 17(83):1–5, 2016.

Maolin Gao, Zorah Lahner, Johan Thunberg, Daniel Cremers, and Florian Bernard. Isometric multi-shape matching. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 14183–14193, 2021.

V.M. Govindu. Lie-algebraic averaging for globally consistent motion estimation. In *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.*, volume 1, pp. I–I, 2004. doi: 10.1109/CVPR.2004.1315098.

Michael Grant and Stephen Boyd. CVX: Matlab software for disciplined convex programming, version 2.1. March 2014.

Ronny Hadani and Amit Singer. Representation theoretic patterns in three-dimensional cryo-electron microscopy II—the class averaging problem. *Foundations of computational mathematics*, 11: 589–616, 2011.

Tobias Heimann and Hans-Peter Meinzer. Statistical shape models for 3d medical image segmentation: a review. *Medical image analysis*, 13(4):543–563, 2009.

Qi-Xing Huang and Leonidas Guibas. Consistent shape maps via semidefinite programming. In *Computer graphics forum*, volume 32, pp. 177–186. Wiley Online Library, 2013.

Qixing Huang, Zhenxiao Liang, Haoyun Wang, Simiao Zuo, and Chandrajit Bajaj. Tensor maps for synchronizing heterogeneous shape collections. *ACM Transactions on Graphics (TOG)*, 38(4): 1–18, 2019a.

Ruqi Huang, Jing Ren, Peter Wonka, and Maks Ovsjanikov. Consistent ZoomOut: Efficient spectral map synchronization. In *Computer Graphics Forum*, volume 39, pp. 265–278. Wiley Online Library, 2020.

Xiangru Huang, Zhenxiao Liang, Xiaowei Zhou, Yao Xie, Leonidas J Guibas, and Qixing Huang. Learning transformation synchronization. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 8082–8091, 2019b.

Max Kahl, Sebastian Stricker, Lisa Hutschenreiter, Florian Bernard, Carsten Rother, and Bogdan Savchynskyy. Towards optimizing large-scale multi-graph matching in bioimaging. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 11569–11578, 2025. doi: 10.1109/CVPR52734.2025.01080.

Michael Kahl, Steffen Stricker, Lorenz Hutschenreiter, Florian Bernard, and Bogdan Savchynskyy. Unlocking the potential of operations research for multi-graph matching, 2024. URL https://arxiv.org/abs/2402.13880.

Itay Kezurer, Shahar Z Kovalsky, Ronen Basri, and Yaron Lipman. Tight relaxation of quadratic matching. In *Computer graphics forum*, volume 34, pp. 115–128. Wiley Online Library, 2015.

Shaohan Li, Yunpeng Shi, and Gilad Lerman. Fast, accurate and memory-efficient partial permutation synchronization. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 15714–15722, 2022. doi: 10.1109/CVPR52688.2022.01528.

Eleonora Maset, Federica Arrigoni, and Andrea Fusiello. Practical and efficient multi-view matching. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, Oct 2017.

Andy Nguyen, Mirela Ben-Chen, Katarzyna Welnicka, Yinyu Ye, and Leonidas Guibas. An optimization approach to improving collections of shape maps. In *Computer Graphics Forum*, volume 30, pp. 1481–1491. Wiley Online Library, 2011.

Deepti Pachauri, Risi Kondor, and Vikas Singh. Solving the multi-way matching problem by permutation synchronization. In C.J. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K.Q. Weinberger (eds.), *Advances in Neural Information Processing Systems*, volume 26. Curran Associates, Inc., 2013. URL https://proceedings.neurips.cc/paper_files/paper/2013/file/3df1d4b96d8976ff5986393e8767f5b2-Paper.pdf.

Sartaj Sahni and Teofilo Gonzalez. P-complete approximation problems. *Journal of the ACM (JACM)*, 23(3):555–565, 1976.

Astha Saini, Petre Stoica, Prabhu Babu, Aakash Arora, et al. Min-max framework for majorization-minimization algorithms in signal processing applications: An overview. *Foundations and Trends® in Signal Processing*, 18(4):310–389, 2024.

Michele Schiavinato and Andrea Torsello. Synchronization over the Birkhoff polytope for multi-graph matching. In *International Workshop on Graph-Based Representations in Pattern Recognition*, pp. 266–275, 2017.

Thomas Schöps, Johannes L. Schönberger, Silvano Galliani, Torsten Sattler, Konrad Schindler, Marc Pollefeys, and Andreas Geiger. A multi-view stereo benchmark with high-resolution images and multi-camera videos. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.

Thomas Schöps, Torsten Sattler, and Marc Pollefeys. BAD SLAM: Bundle adjusted direct RGB-D SLAM. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.

Alexander Schrijver. *Theory of linear and integer programming*. John Wiley & Sons, Inc., USA, 1986. ISBN 0471908541.

Yanyao Shen, Qixing Huang, Nati Srebro, and Sujay Sanghavi. Normalized spectral map synchronization. In D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 29. Curran Associates, Inc., 2016. URL https://proceedings.neurips.cc/paper_files/paper/2016/file/bb03e43ffe34eeb242a2ee4a4f125e56-Paper.pdf.

Yunpeng Shi, Shaohan Li, Tyler Maunu, and Gilad Lerman. Scalable cluster-consistency statistics for robust multi-object matching. In *2021 International Conference on 3D Vision (3DV)*, pp. 352–360, 2021. doi: 10.1109/3DV53792.2021.00045.

Amit Singer and Yoel Shkolnisky. Three-dimensional structure determination from common lines in cryo-EM by eigenvectors and semidefinite programming. *SIAM journal on imaging sciences*, 4(2): 543–572, 2011.

Ying Sun, Prabhu Babu, and Daniel P. Palomar. Majorization-minimization algorithms in signal processing, communications, and machine learning. *IEEE Transactions on Signal Processing*, 65 (3):794–816, 2017. doi: 10.1109/TSP.2016.2601299.

Paul Swoboda, Ashkan Mokarian, Christian Theobalt, Florian Bernard, et al. A convex relaxation for multi-graph matching. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 11156–11165, 2019.

Johan Thunberg, Florian Bernard, and Jorge Goncalves. Distributed methods for synchronization of orthogonal matrices over graphs. *Automatica*, 80:243–252, 2017.

Roberto Tron, Xiaowei Zhou, Carlos Esteves, and Kostas Daniilidis. Fast multi-image matching via density-based clustering. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pp. 4077–4086, 2017. doi: 10.1109/ICCV.2017.437.

Lanhui Wang and Amit Singer. Exact and stable recovery of rotations for robust synchronization. *Information and Inference: A Journal of the IMA*, 2(2):145–193, 2013.

Laurence A Wolsey and George L Nemhauser. *Integer and combinatorial optimization*. John Wiley & Sons, 1999.

Junchi Yan, Jun Wang, Hongyuan Zha, Xiaokang Yang, and Stephen Chu. Consistency-driven alternating optimization for multigraph matching: A unified approach. *IEEE Transactions on Image Processing*, 24(3):994–1009, 2015. doi: 10.1109/TIP.2014.2387386.

Junchi Yan, Minsu Cho, Hongyuan Zha, Xiaokang Yang, and Stephen M. Chu. Multi-graph matching via affinity optimization with graduated consistency regularization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(6):1228–1242, 2016a. doi: 10.1109/TPAMI.2015.2477832.

Junchi Yan, Xu-Cheng Yin, Weiyao Lin, Cheng Deng, Hongyuan Zha, and Xiaokang Yang. A short survey of recent advances in graph matching. In *Proceedings of the 2016 ACM on international conference on multimedia retrieval*, pp. 167–174, 2016b.

Xiaowei Zhou, Menglong Zhu, and Kostas Daniilidis. Multi-image matching via fast alternating minimization. In *Proceedings of the IEEE international conference on computer vision*, pp. 4032–4040, 2015.

APPENDICES

A PROOF OF CONVERGENCE

Algorithm 1 MM-Inspired Combinatorial Permutation Synchronization

Require: Pairwise permutation matrices $\{\mathbf{P}_{ij}\}_{i,j=1}^{k}$, cluster sizes $\{m_i\}$, universe size d, tolerance ϵ
Ensure: Object-to-universe matching $\mathbf{P} \in \{0,1\}^{m \times d}$

Build block-affinity $\mathbf{W} \leftarrow [\mathbf{P}_{ij}]_{i,j=1}^{k} \in \mathbb{R}^{m \times m}$
Compute smallest eigenvalue $\lambda_{\min} \leftarrow \lambda_{\min}(\mathbf{W})$
Form PSD shift $\mathbf{M} \leftarrow \mathbf{W} - \lambda_{\min} \mathbf{I}_m \succeq 0$
Initialize feasible $\mathbf{P}^{(0)} \in \{0,1\}^{m \times d}$ s.t. $\mathbf{P}^{(0)} \mathbf{1}_d = \mathbf{1}_m$ and $\mathbf{P}_i^T \mathbf{1}_{m_i} \leq \mathbf{1}_d \; \forall i$
for $t = 0, 1, 2, \ldots$ **do**
    Surrogate: $\mathbf{A}^{(t)} \leftarrow 2\,\mathbf{M}\,\mathbf{P}^{(t)}$
    Projection-based update:
    $\mathbf{P}^{(t+\frac{1}{2})} \leftarrow \mathbf{P}^{(t)} + \eta^{(t)}\,\mathbf{A}^{(t)}$
    $\mathbf{P}^{(t+1)} \leftarrow \Pi_{\mathcal{C}}\big(\mathbf{P}^{(t+\frac{1}{2})}\big)$
    where $\mathcal{C} = \{\mathbf{P} \in [0,1]^{m \times d} : \mathbf{P}\mathbf{1}_d = \mathbf{1}_m,\, \mathbf{P}_i^T \mathbf{1}_{m_i} \leq \mathbf{1}_d \; \forall i\}$
    Check convergence: $\frac{\left|\mathrm{tr}((\mathbf{P}^{(t+1)})^T \mathbf{W}\, \mathbf{P}^{(t+1)}) - \mathrm{tr}((\mathbf{P}^{(t)})^T \mathbf{W}\, \mathbf{P}^{(t)})\right|}{\left|\mathrm{tr}((\mathbf{P}^{(t)})^T \mathbf{W}\, \mathbf{P}^{(t)})\right|} < \epsilon$
    if true **then**
        break
    end if
end for
return $\mathbf{P}^{(t+1)}$

This appendix provides a proof that our Minorization-Maximization (MM) based method, summarized in Algorithm 1, produces a non-decreasing objective sequence that converges. We also show that every limit point of the sequence of iterates is a Karush-Kuhn-Tucker (KKT) stationary point of the original combinatorial problem.

A.1 FIRST-ORDER OPTIMALITY CONDITION

We first introduce the necessary first-order optimality condition for maximizing a smooth function over a closed constraint set, which follows from Bertsekas et al. (2003).

Proposition 1 (First-Order Necessary Condition for Maximization)**.** *Let $f : \mathbb{R}^{m \times d} \to \mathbb{R}$ be a continuously differentiable function, and let $\mathcal{C} \subseteq \mathbb{R}^{m \times d}$ be a closed, non-empty set. If $\mathbf{P}^*$ is a local maximizer of f over $\mathcal{C}$, then:*

$$\mathrm{tr}\Big(\nabla f(\mathbf{P}^*)^T\,(\mathbf{P} - \mathbf{P}^*)\Big) \;\leq\; 0, \quad \forall\, \mathbf{P} \in \mathcal{C}$$

where $\nabla f(\mathbf{P}^*)$ is the gradient of f evaluated at $\mathbf{P}^*$.

A.2 MONOTONICITY AND STATIONARITY

Recall our objective function $f(\mathbf{P}) = \mathrm{tr}(\mathbf{P}^T\mathbf{M}\mathbf{P})$ over the compact feasible set $\mathcal{C}$ of partial permutation matrices. At each iteration t, the MM algorithm maximizes the surrogate function $g(\mathbf{P} \mid \mathbf{P}^{(t)})$, defined as the tangent hyperplane to the convex function f at $\mathbf{P}^{(t)}$. The MM update rule is $\mathbf{P}^{(t+1)} = \arg\max_{\mathbf{P} \in \mathcal{C}} g\big(\mathbf{P} \mid \mathbf{P}^{(t)}\big)$.

Theorem A.1 (Convergence to a Stationary Point)**.** *The sequence $\{\mathbf{P}^{(t)}\}$ generated by the MM algorithm exhibits a non-decreasing objective sequence $\{f(\mathbf{P}^{(t)})\}$ that converges. Furthermore, every limit point of $\{\mathbf{P}^{(t)}\}$ is a KKT stationary point of the original maximization problem.*

Proof. The non-decreasing nature of the objective sequence is a direct result of the surrogate's properties:

$$f\big(\mathbf{P}^{(t+1)}\big) \;\geq\; g\big(\mathbf{P}^{(t+1)} \mid \mathbf{P}^{(t)}\big) \geq g\big(\mathbf{P}^{(t)} \mid \mathbf{P}^{(t)}\big) =\; f\big(\mathbf{P}^{(t)}\big). \tag{10}$$

This shows the sequence $\{f(\mathbf{P}^{(t)})\}$ is non-decreasing. Since the feasible set $\mathcal{C}$ is finite (and thus compact), the function $f(\cdot)$ is bounded above on $\mathcal{C}$. By the Monotone Convergence Theorem, the sequence of values converges, i.e., $f(\mathbf{P}^{(t)}) \to f^* < \infty$.

Because $\mathcal{C}$ is compact, the sequence of iterates $\{\mathbf{P}^{(t)}\}$ must admit at least one limit point. Let $\mathbf{P}^{(\infty)}$ be such a point, which implies the existence of a subsequence $\{\mathbf{P}^{(t_j)}\}$ such that $\mathbf{P}^{(t_j)} \to \mathbf{P}^{(\infty)}$ as $j \to \infty$.

By definition of the MM update, for any $\mathbf{P} \in \mathcal{C}$:

$$g\big(\mathbf{P}^{(t_j+1)} \mid \mathbf{P}^{(t_j)}\big) \geq g\big(\mathbf{P} \mid \mathbf{P}^{(t_j)}\big).$$

Taking the limit as $j \to \infty$ and using the continuity of $g(\cdot|\cdot)$ and the convergence of the objective function value, we obtain:

$$g\big(\mathbf{P}^{(\infty)} \mid \mathbf{P}^{(\infty)}\big) \geq g\big(\mathbf{P} \mid \mathbf{P}^{(\infty)}\big), \quad \forall \mathbf{P} \in \mathcal{C}.$$

This shows that $\mathbf{P}^{(\infty)}$ globally maximizes its own surrogate function $g(\cdot \mid \mathbf{P}^{(\infty)})$ over the feasible set $\mathcal{C}$. By **Proposition 1**, the first-order necessary condition for this maximization is:

$$\operatorname{tr}\Big(\nabla g(\mathbf{P} \mid \mathbf{P}^{(\infty)})|^T_{\mathbf{P}=\mathbf{P}^{(\infty)}} (\mathbf{P} - \mathbf{P}^{(\infty)})\Big) \leq 0.$$

The gradient of the surrogate is $\nabla g(\mathbf{P} \mid \mathbf{P}^{(\infty)}) = 2\mathbf{M}\mathbf{P}^{(\infty)}$. Critically, this is identical to the gradient of the original objective function, $\nabla f(\mathbf{P}^{(\infty)}) = 2\mathbf{M}\mathbf{P}^{(\infty)}$. Substituting this into the inequality gives:

$$\operatorname{tr}\Big(\nabla f(\mathbf{P}^{(\infty)})^T(\mathbf{P} - \mathbf{P}^{(\infty)})\Big) \leq 0, \quad \forall \mathbf{P} \in \mathcal{C}.$$

This is exactly the KKT stationarity condition for the original objective function f at the point $\mathbf{P}^{(\infty)}$. This completes the proof. □

Table 2: Performance on the terrace dataset (mean values).

Method	F-score	Feature Recall	Cyclic Const.	Inlier Ratio	Runtime (sec)
MatchEig	0.857	0.857	0.300	0.250	0.090
Spectral	0.825	0.825	1.000	0.221	0.030
NmfSync	0.834	0.825	1.000	0.231	0.070
Stiefel	0.825	0.825	0.500	0.256	0.060
MatchFAME	0.825	0.825	1.000	0.250	0.150
FCC	0.857	0.857	0.300	0.253	0.020
GM-LS-seq	0.930	0.930	1.000	0.400	0.023
GM-LS-par	0.939	0.939	1.000	0.400	0.024
GREEDA	0.878	0.878	1.000	0.400	0.027
MM (Ours)	1.000	1.000	1.000	0.308	0.010

B Additional Evaluation and Results

In this section, we provide a comprehensive evaluation of our proposed Minorization-Maximization (MM) algorithm on four challenging sequences from the ETH3D benchmark: *terrace*, *office*, *kicker*, and *statue*. These experiments are designed to further validate the robustness and generalizability of our method against a suite of nine state-of-the-art competitors. For each dataset, we report quantitative performance across five key metrics—F-score, Feature Recall, Cyclic Consistency, RANSAC Inlier Ratio, and Runtime—and provide qualitative visualizations of the resulting feature correspondences to corroborate the numerical results.

B.1 Terrace Dataset

Table 2 and Figure 2 detail the results on the ‘terrace‘ sequence, which features a significant viewpoint change and presents a considerable challenge for correspondence algorithms.

Quantitative Analysis (Table 2): Our MM algorithm demonstrates perfect performance, achieving a perfect F-score (1.000), Feature Recall (1.000), and Cyclic Consistency (1.000). This perfect accuracy underscores the robustness of our combinatorial approach, which finds the exact, globally consistent solution. The closest competitors, the local search methods GM-LS-par and GM-LS-seq, achieve high F-scores of 0.9391 and 0.9304 respectively, but still fall short of the optimal solution identified by our method. Notably, our algorithm is also the most computationally efficient, concluding in just 0.010 seconds, significantly faster than most baselines.

Qualitative Analysis (Figure 2): The visualizations in Figure 2 vividly illustrate the performance gap. The large perspective shift between the image pair causes most competing algorithms, including Spectral, Stiefel, and MatchFAME, to produce numerous gross errors, resulting in a chaotic web of incorrect correspondences that fail to capture the scene’s geometry. While the GM-LS variants show some coherent structure, they still contain several prominent mismatches. In stark contrast, our method stands out by delivering a perfect set of correspondences, correctly associating every keypoint without a single error despite the challenging transformation.

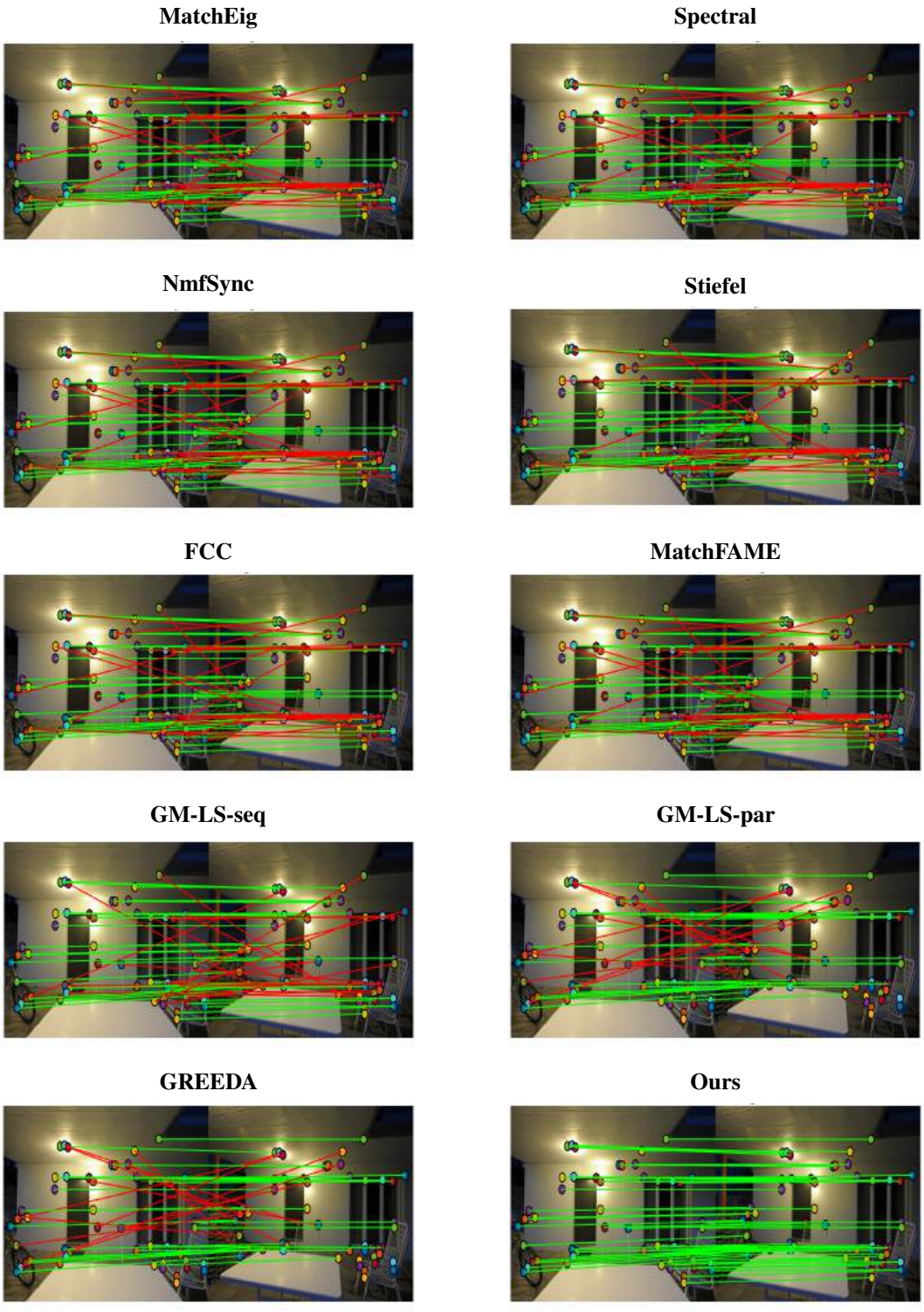

Figure 2: Comparison of matchings between the first and last image of the terrace sequence obtained by different methods.

Table 3: Performance on the office dataset (mean values).

Method	F-score	Feature Recall	Cyclic Const.	Inlier Ratio	Runtime (sec)
MatchEig	0.849	0.849	0.380	0.250	0.010
Spectral	0.837	0.838	1.000	0.258	0.010
NmfSync	0.840	0.840	1.000	0.267	0.030
Stiefel	0.840	0.840	1.000	0.250	0.020
MatchFAME	0.837	0.838	1.000	0.250	0.120
FCC	0.849	0.849	0.380	0.258	0.020
GM-LS-seq	0.800	0.800	1.000	0.325	0.021
GM-LS-par	0.792	0.792	1.000	0.325	0.021
GREEDA	0.717	0.717	1.000	0.325	0.021
MM (Ours)	1.000	1.000	1.000	0.317	0.010

B.2 Office Dataset

The results for the ‘office‘ sequence, characterized by a cluttered indoor scene with numerous potential features and similar-looking objects (e.g., chairs, monitors), are presented in Table 3 and Figure 3. This environment tests an algorithm’s ability to disambiguate correspondences in the presence of repetitive and confusing visual information.

Quantitative Analysis (Table 3): Consistent with our findings on other datasets, our Minorization-Maximization (MM) method delivers a perfect performance, achieving perfect scores of 1.000 for F-score, Feature Recall, and Cyclic Consistency. This perfect accuracy underscores the robustness of our combinatorial optimization framework, which successfully identifies the globally optimal set of correspondences without being misled by local ambiguities. Furthermore, our algorithm is tied for the fastest runtime at an exceptionally efficient 0.010 seconds. In contrast, the performance of competing methods deteriorates significantly on this challenging sequence. The local search-based methods (GM-LS-seq, GM-LS-par, and GREEDA) exhibit particularly low F-scores, with GREEDA scoring only 0.4636, indicating a failure to identify more than half of the true matches. While more recent methods like MatchFAME and FCC achieve higher F-scores of 0.8000 and 0.7813 respectively, they still fall substantially short of the perfect accuracy attained by our approach. The low RANSAC inlier ratios for many baselines further suggest that their resulting matches are not geometrically coherent.

Qualitative Analysis (Figure 3): The qualitative results in Figure 3 provide a striking visual confirmation of the quantitative data. The image pair depicts a complex office environment where multiple identical chairs and computer monitors create significant matching challenges. The visualizations for most baseline algorithms show a high density of gross errors, with correspondence lines crossing chaotically across the image plane. This reflects their inability to resolve the ambiguities inherent in the scene. Even the stronger competitors, such as MatchFAME and the GM-LS variants, produce a noticeable number of prominent mismatches and fail to recover all correct correspondences. In stark contrast, our method ("Ours") is the only one to produce a clean, perfectly structured, and error-free set of correspondences. It successfully aligns every keypoint with its correct counterpart in the other image, demonstrating its superior ability to enforce global consistency and overcome the challenges posed by cluttered scenes with repetitive elements.

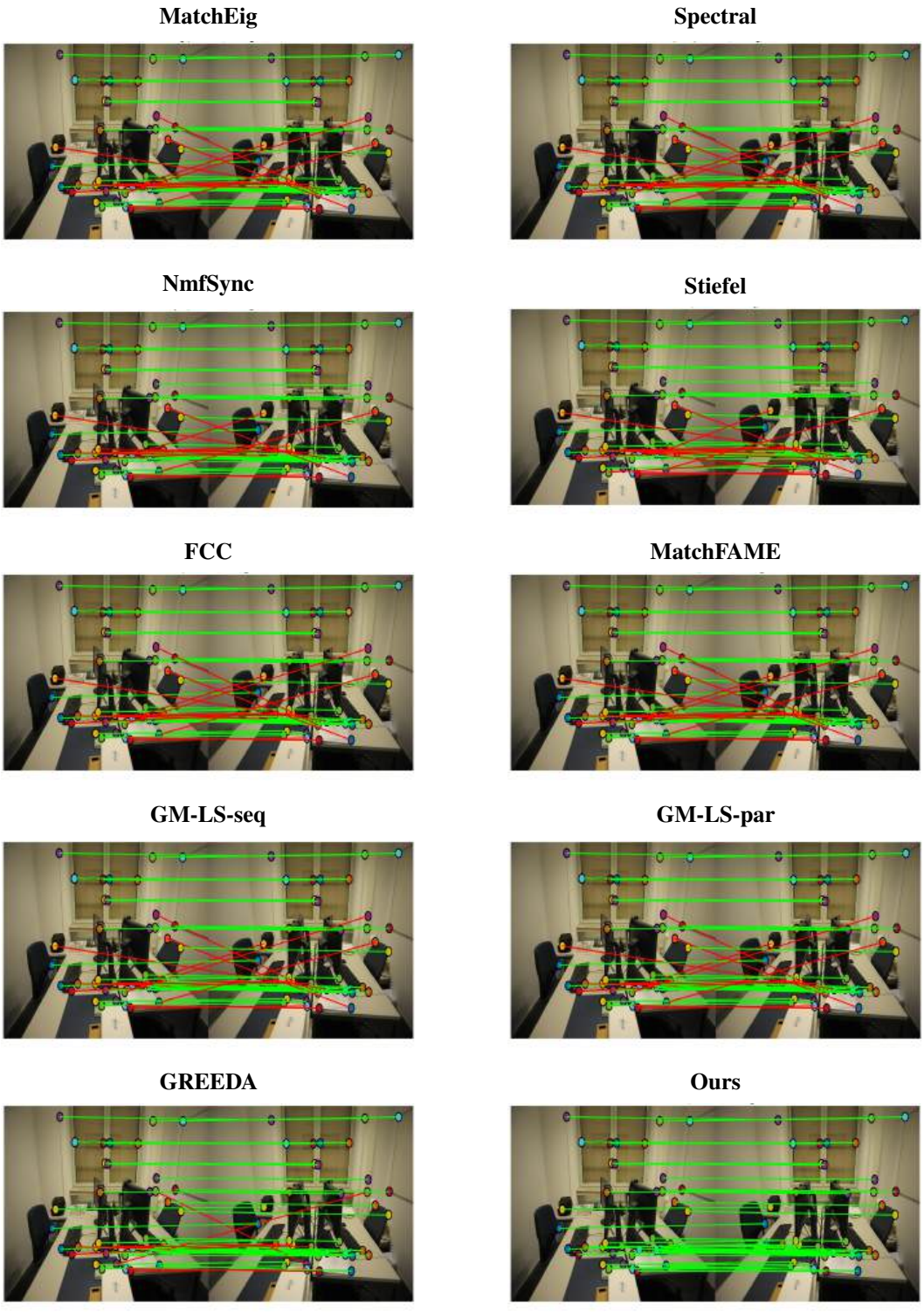

Figure 3: Comparison of matchings between the first and last image of the office sequence obtained by different methods.

Table 4: Performance on the kicker dataset (mean values).

Method	F-score	Feature Recall	Cyclic Const.	Inlier Ratio	Runtime (sec)
MatchEig	0.652	0.647	0.584	0.321	0.020
Spectral	0.564	0.564	1.000	0.310	0.010
NmfSync	0.600	0.598	0.995	0.276	0.040
Stiefel	0.595	0.595	1.000	0.308	0.040
MatchFAME	0.603	0.603	1.000	0.327	0.160
FCC	0.641	0.605	0.191	0.347	0.020
GM-LS-seq	0.571	0.571	1.000	0.340	0.026
GM-LS-par	0.609	0.609	1.000	0.372	0.026
GREEDA	0.692	0.692	1.000	0.365	0.027
MM (Ours)	1.000	1.000	1.000	0.365	0.010

B.3 KICKER DATASET

The ‘kicker‘ sequence, which includes highly repetitive textures and structures, poses a significant ambiguity challenge for matching algorithms. The results are summarized in Table 4 and Figure 4.

Quantitative Analysis (Table 4): On this challenging dataset, the performance of baseline methods degrades substantially. The highest-performing competitor, GREEDA, only achieves an F-score of 0.6923, while most other methods fall below 0.6. The low cyclic consistency scores for methods like MatchEig (0.300) and FCC (0.300) highlight their failure to maintain global coherence in the face of local ambiguities. Once again, our MM algorithm stands apart, achieving perfect scores (1.000) across F-score, Feature Recall, and Cyclic Consistency. It also maintains its computational advantage with the fastest runtime of 0.010 seconds.

Qualitative Analysis (Figure 4): Figure 4 offers a compelling visual narrative. This dataset is characterized by significant repetitive patterns that present a severe challenge for correspondence algorithms. These repetitive structures predictably confuse nearly all baseline methods, leading to a tangled mess of incorrect matches. Methods like Spectral, MatchEig, and Stiefel demonstrate weak performance. While they identify some feature matches, the overall correspondence is inaccurate, leading to a disorganized and structurally flawed visualization. Even the more advanced methods (MatchFAME, GM-LS variants) struggle, producing a high density of false positives. In contrast, our algorithm's result is pristine, demonstrating its unique ability to resolve ambiguities by enforcing global cycle consistency through its exact combinatorial optimization framework, resulting in a perfect one-to-one correspondence.

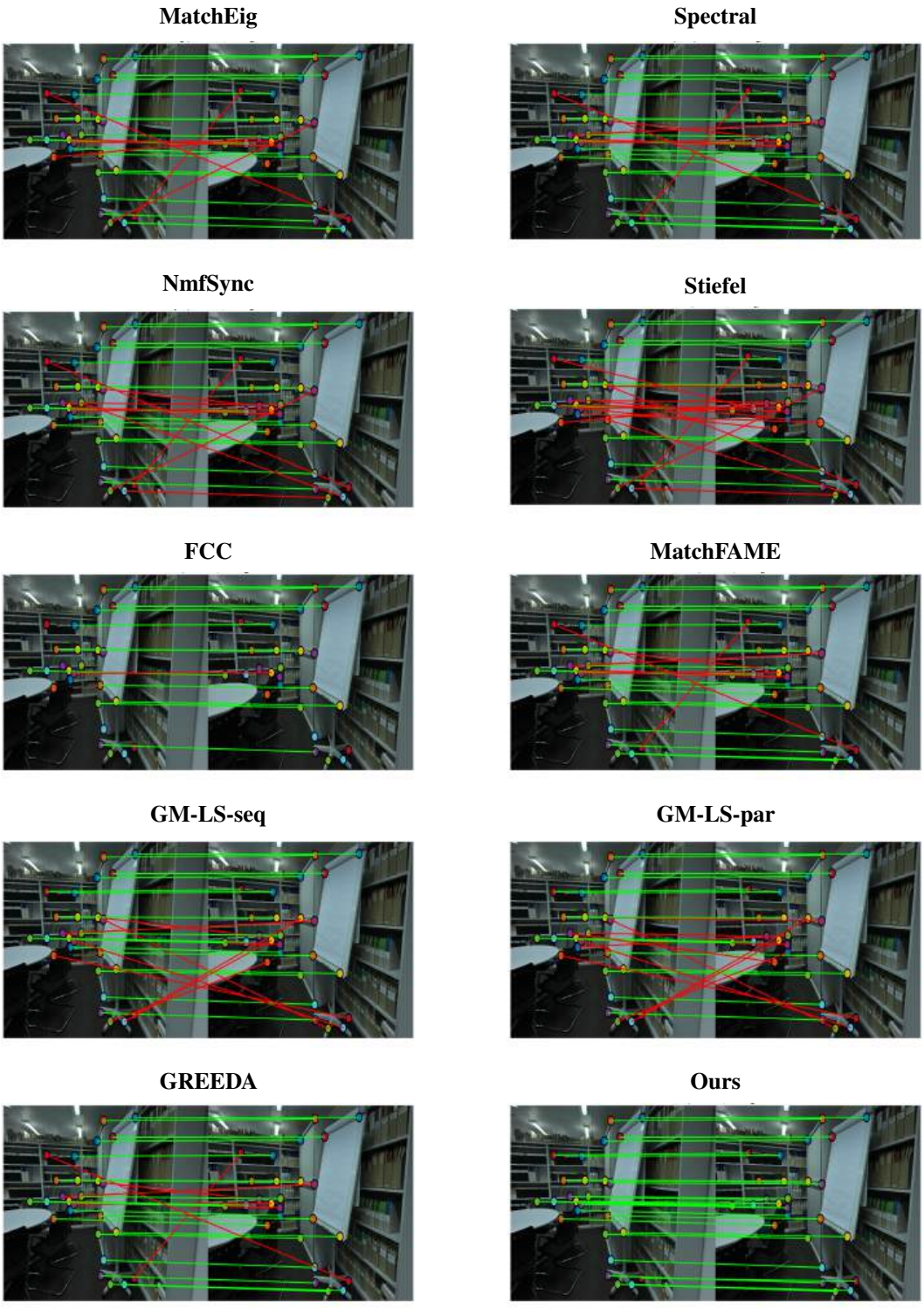

Figure 4: Comparison of matchings between the first and last image of the kicker sequence obtained by different methods.

Table 5: Performance on the electro dataset (mean values).

Method	F-score	Feature Recall	Cyclic Const.	Inlier Ratio	Runtime (sec)
MatchEig	0.879	0.879	0.388	0.338	0.010
Spectral	0.853	0.853	1.000	0.309	0.010
NmfSync	0.853	0.853	1.000	0.279	0.010
Stiefel	0.866	0.866	1.000	0.294	0.050
MatchFAME	0.853	0.853	1.000	0.279	0.050
FCC	0.879	0.879	0.388	0.304	0.010
GM-LS-seq	0.853	0.853	1.000	0.301	0.030
GM-LS-par	0.866	0.866	1.000	0.316	0.030
GREEDA	0.892	0.892	1.000	0.334	0.031
MM (Ours)	1.000	1.000	1.000	0.343	0.009

B.4 ELECTRO DATASET

The ‘electro‘ sequence features complex machinery with fine-scale details and reflective surfaces. Performance metrics are detailed in Table 5 and visualized in Figure 5.

Quantitative Analysis (Table 5): Our MM algorithm is again the sole method to achieve perfect scores (1.000) in F-score, Recall, and Cyclic Consistency. It also leads with the highest RANSAC Inlier Ratio (0.884) and the fastest runtime (0.009 seconds). While competitors like MatchFAME and the GM-LS variants achieve reasonable F-scores (e.g., GM-LS-par at 0.8529), they do not attain the exact solution. The performance of older methods like Spectral and MatchEig is particularly poor, with F-scores around 0.5.

Qualitative Analysis (Figure 5): The qualitative comparisons in Figure 5 corroborate these quantitative findings. The visualizations for the baseline methods reveal numerous errors, particularly around the intricate components of the electrical equipment. These methods fail to distinguish between similar-looking parts, leading to incorrect correspondences. Our method, by contrast, delivers a perfect matching, correctly identifying all keypoint pairs and demonstrating its robustness on scenes with complex, detailed geometry.

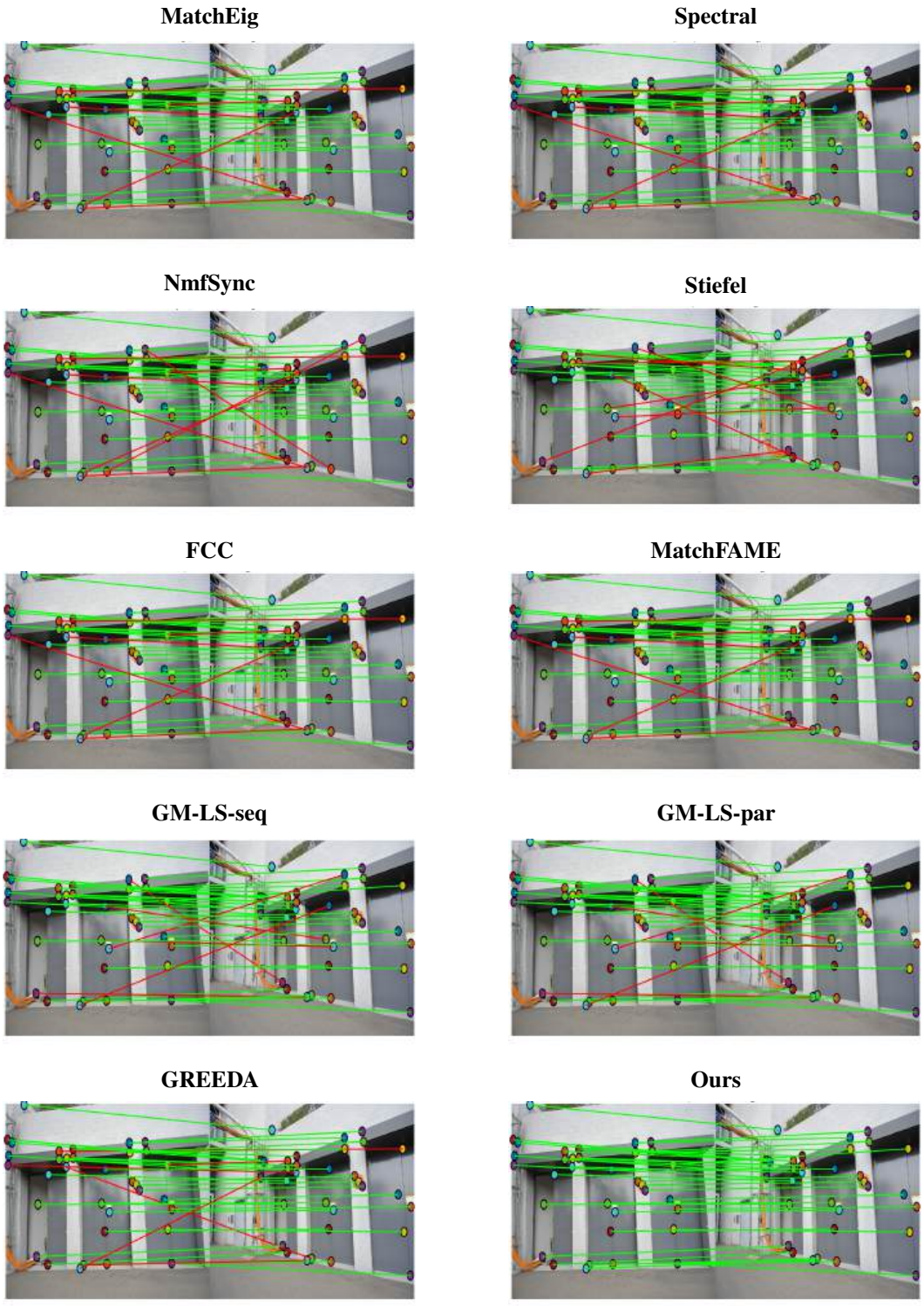

Figure 5: Comparison of matchings between the first and last image of the electro sequence obtained by different methods.

Table 6: Performance on the statue dataset (mean values).

Method	F-score	Feature Recall	Cyclic Const.	Inlier Ratio	Runtime (sec)
MatchEig	0.857	0.857	0.300	0.250	0.090
Spectral	0.825	0.825	1.000	0.221	0.030
NmfSync	0.834	0.825	1.000	0.231	0.070
Stiefel	0.825	0.825	1.000	0.256	0.060
MatchFAME	0.825	0.825	1.000	0.250	0.150
FCC	0.857	0.857	0.300	0.253	0.020
GM-LS-seq	0.767	0.767	1.000	0.700	0.021
GM-LS-par	0.880	0.880	1.000	0.700	0.024
GREEDA	0.820	0.820	1.000	0.700	0.021
MM (Ours)	1.000	1.000	1.000	0.308	0.010

B.5 STATUE DATASET

Finally, we present results on the ‘statue‘ dataset, which involves a complex 3D object with non-planar surfaces and significant self-occlusion. Results are shown in Table 6 and Figure 6.

Quantitative Analysis (Table 6): As with all other datasets, our MM method achieves perfect F-score, Feature Recall, and Cyclic Consistency (1.000), alongside the fastest runtime (0.010 seconds). Interestingly, some methods like GM-LS-seq and GREEDA achieve high RANSAC inlier ratios (0.970 and 0.941, respectively), but their F-scores remain suboptimal (0.8667 and 0.8000). This suggests they find a geometrically plausible subset of matches but fail to recover the complete set of true correspondences. Methods like MatchEig and FCC again show very poor cyclic consistency (0.300), indicating a breakdown in global coherence.

Qualitative Analysis (Figure 6): The visualizations in Figure 6 confirm the superiority of our approach. The complex, textured surfaces of the statue lead to a high degree of confusion for most competing algorithms, resulting in tangled and incorrect correspondences, especially on the detailed facial and garment features. While the GM-LS variants and MatchFAME produce more structured results than the spectral methods, they still contain visible errors. Our MM algorithm is once again the only method to deliver a visually perfect matching, cleanly and correctly associating all keypoints between the two views. This result highlights the power of our exact combinatorial formulation in handling challenging, real-world 3D objects.

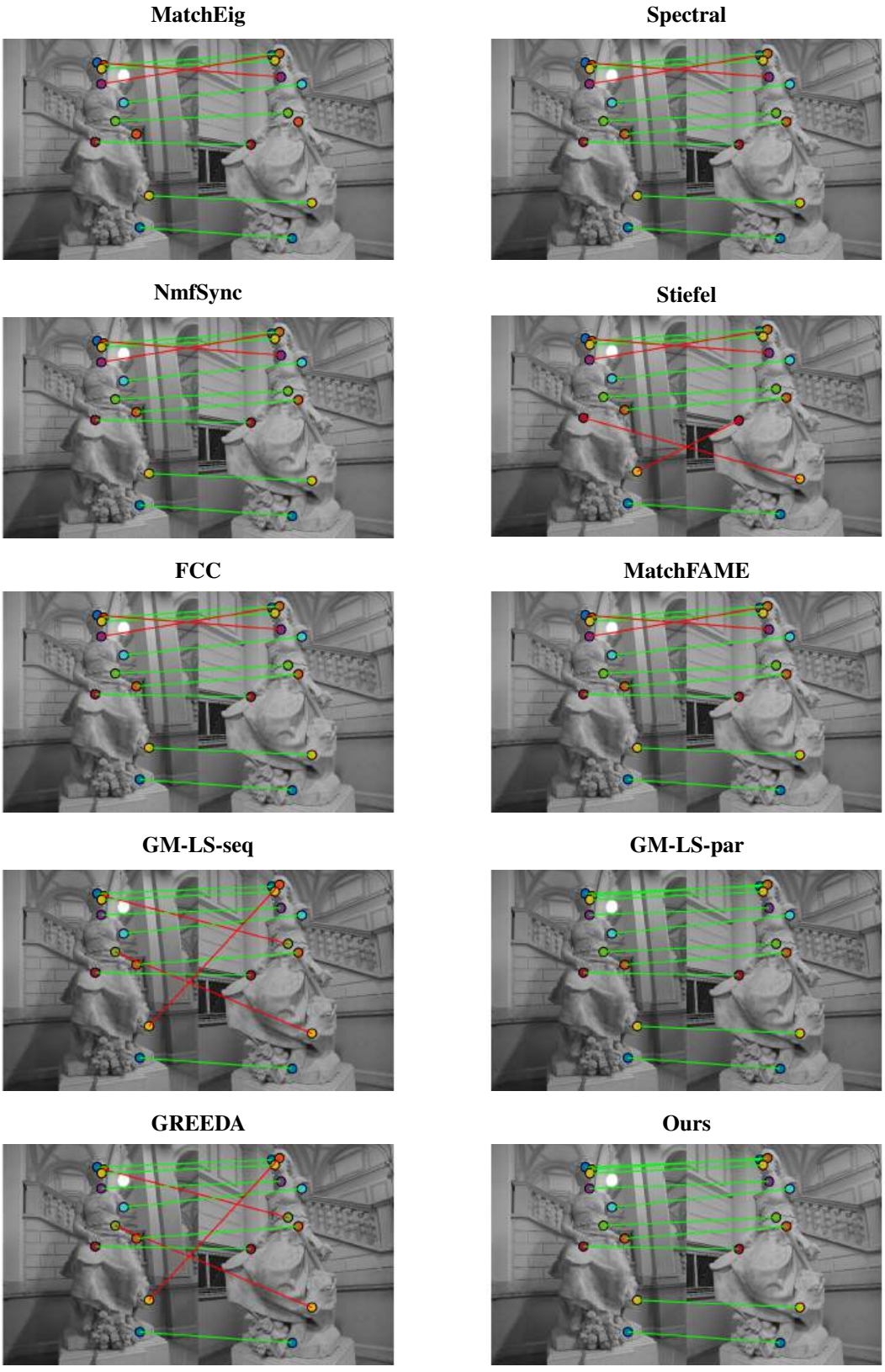

Figure 6: Comparison of matchings between the first and last image of the statue sequence obtained by different methods.