Grid-Guided Sparse Laplacian Consensus for Robust Feature Matching

Yifan Xia and Jiayi Ma, *Senior Member, IEEE*

***Abstract*—Feature matching is a fundamental concern widely employed in computer vision applications. This paper introduces a novel and efficacious method named *Grid-guided Sparse Laplacian Consensus*, rooted in the concept of smooth constraints. To address challenging scenes such as severe deformation and independent motions, we devise grid-based adaptive matching guidance to construct multiple transformations based on motion coherence. Specifically, we obtain a set of precise yet sparse seed correspondences through motion statistics, facilitating the generation of an adaptive number of candidate correspondence sets. In addition, we propose an innovative formulation grounded in graph Laplacian for correspondence pruning, wherein mapping function estimation is formulated as a Bayesian model. We solve this utilizing EM algorithm with seed correspondences as initialization for optimal convergence. Sparse approximation is leveraged to reduce the time-space burden. A comprehensive set of experiments are conducted to demonstrate the superiority of our method over other state-of-the-art methods in both robustness to serious deformations and generalizability for various descriptors, as well as generalizability to multi motions. Additionally, experiments in geometric estimation, image registration, loop closure detection, and visual localization highlight the significance of our method across diverse scenes for high-level tasks.**

***Index Terms*—Feature matching, correspondence pruning, graph Laplacian, motion coherence.**

I. Introduction

CONSTRUCTING reliable correspondences across images with related contents is essential for many computer vision tasks, such as image registration and fusion, structure-from-motion, panoramic stitching, image retrieval, loop closure detection, and simultaneous localization and mapping [1], [2], [3], [4], [5]. Traditional feature matching utilizes local descriptors (*e.g.*, SIFT [6]) to establish point-to-point correspondences based on salient structures. Academic endeavors have witnessed the application of deep learning techniques in the domain of descriptor construction, including HardNet [7] and SuperPoint [8]. Despite the substantial progress made in the field of descriptor design, it is important to acknowledge that correspondences primarily reliant on local information remain susceptible to instability and are predisposed to the presence of outliers. For example, the topological geometry between feature correspondences cannot be mined using only the plain Nearest Neighbor (NN) search.

In order to mitigate this challenge, geometric constraints within the *keypoint space* are harnessed to discriminate against erroneous correspondences (outliers) while preserving correct ones (inliers). This is typically achieved through the evaluation of the consistency of correct correspondences with a global transformation model. The classical RANdom SAmple Consensus (RANSAC) method [9] identifies outliers by seeking the largest subset conforming to a task-specific model (*e.g.,* homography and epipolar geometry), including subsequent improved methods [10], [11], [12], [13], [14], [15], [16], [17]. However, it is important to note that these approaches often necessitate the a priori specification of the transformation model type, may exhibit suboptimal performance when dealing with complex nonrigid images, and can significantly compromise effectiveness in scenarios with high outlier rates.

No predefined models are required, and relaxation-based methods have been proven effective. They typically rely on neighborhood support to identify the correct correspondences. Examples of such methods include GMS [19], LPM [20], and LGSC [21]. While these methods can quickly remove mismatches, they require manual parameters setting, leading to low generalizability. Graph matching has emerged as another alternative, exemplified by LOGO [22], MCDM [23] and PFM [24]. While relaxation-based methods provide sufficient theoretical innovation, their practical performance is not clearly advantageous due to the relaxed constraints. Deep learning-based methodologies have undergone rapid and significant advancements in recent years, such as LFGC [25], OANet [26], LMCNet [27], and ConvMatch [18]. However, it is imperative to acknowledge that data-driven methodologies in this realm often exhibit a lack of interpretability and are subject to limited generalizability, contingent upon the nature and extent of the training data employed.

To address the aforementioned issues, non-parametric fitting methods have been introduced. The representative classic methods include ICF [28] and VFC [29]. They alleviate mismatches by assessing the consistency of feature with estimated mapping functions. Building upon these foundations, recent extensions of these algorithms have further improved the accuracy and efficiency of feature matching, such as BMF [30], CRC [31], and GPMatch [32].

Nonetheless, the endeavor to formulate a comprehensive global model for motion coherence based on sparse

Received 15 November 2023; revised 24 November 2024; accepted 2 February 2025. Date of publication 17 February 2025; date of current version 25 February 2025. This work was supported by the National Natural Science Foundation of China under Grant 624B2107 and Grant 62276192. The associate editor coordinating the review of this article and approving it for publication was Dr. Fabrizio Guerrini. *(Corresponding author: Jiayi Ma.)*

The authors are with the Electronic Information School, Wuhan University, Wuhan 430072, China (e-mail: xiayifan@whu.edu.cn; jyma2010@gmail.com).

Digital Object Identifier 10.1109/TIP.2025.3539469

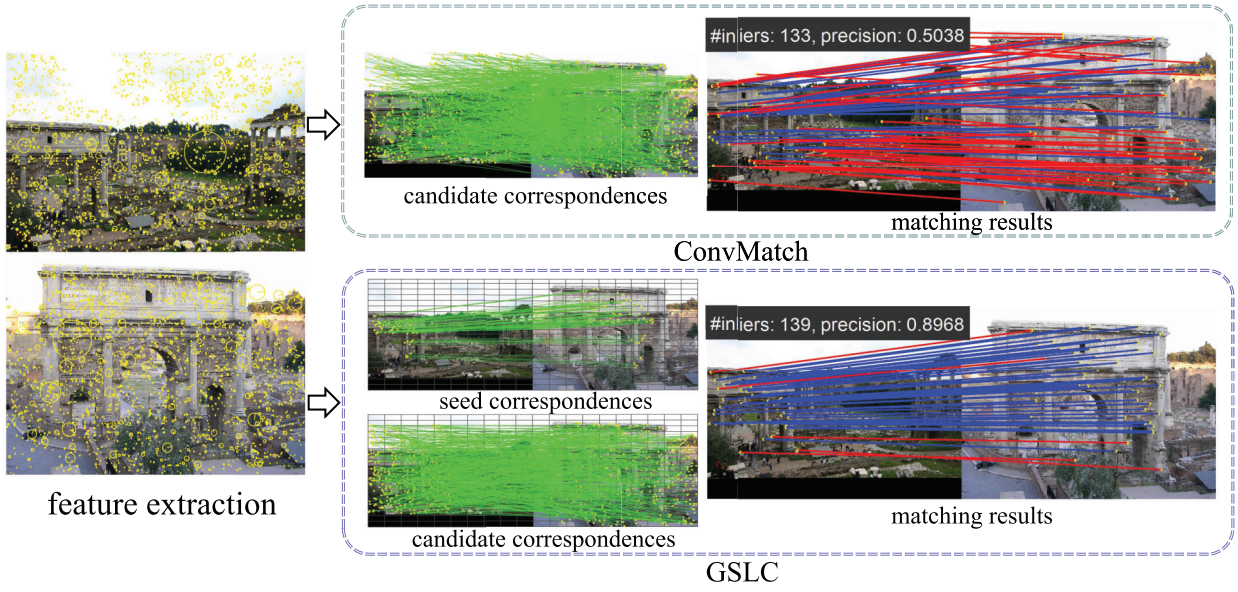

Fig. 1. A framework diagram of our GSLC compared to the state-of-the-art deep learning method ConvMatch [18]. Conventional feature matching typically involves obtaining a candidate matching set using simple tricks like NN and then employing mismatch removal methods such as ConvMatch to identify correct matches, resulting in the final output on the right side of the first row. Our GSLC, on the other hand, first processes NN results with grid statistics to obtain seed correspondences and multi-layer candidate correspondence sets. These seed correspondences are then used to guide sparse Laplacian consensus for filtering each candidate correspondence set, resulting in the final output on the right side of the second row. Best viewed in color.

correspondences between image features presents formidable challenges. Firstly, the distribution of potential correspondences derived from descriptors typically assumes a discrete, sparse, and non-uniform nature, rendering the establishment of a consistent global coherence constraint a complex undertaking. Secondly, the existence of intricate relationships between two observed scenes, encompassing elements like independent motions and wide-baseline imaging, significantly compounds the challenges confronting feature matching techniques. These hurdles pertain to the robustness required to handle complex motion patterns and the ability to generalize effectively across diverse scene contexts.

To address these problems, we propose a grid-based matching framework termed *Grid-guided Sparse Laplacian Consensus* (GSLC), as depicted in the schematic in Fig. 1. The initialization of seed correspondences serves as the foundation for subsequent sparse Laplacian consensus based on smoothing constraints, ensuring the convergence of the global optimum solution. By constructing multi-layer candidate correspondence sets, GSLC revolves around the establishment of an adaptable number of transformations grounded in smooth constraints, as opposed to relying on a single global smooth function. Empirical evidence derived from experiments attests to the robustness of our method in handling demanding scenarios, including non-rigid and wide-baseline scenes, and underscores its capacity for effective generalization across diverse image contexts, encompassing scenarios characterized by pronounced deformations and independent motions. Moreover, the utility of our approach is corroborated by its performance in tasks such as geometric matrix estimation, image registration, loop closure detection, and visual localization tasks, which are integral to high-level vision applications.

In summary, the paper makes two major contributions:

- We introduce a grid-based matching guidance framework that systematically constructs an adaptable number of transformations grounded in principles of smoothness. This approach is specifically designed to contend with motion discontinuities arising from factors such as independent motions or abrupt depth changes for robustness to various scenarios.
- We devise an innovative formulation of motion coherence that leverages the graph Laplacian and introduces a novel representation of the mapping function. This formulation enables us to frame the correspondence pruning problem as a Bayesian model. We employ the EM algorithm and a sparse approximation technique to enhance both robustness and computational efficiency in this context.

The rest of the paper is organized as follows. The researches related to our work are presented in Section II. Section III introduces the details of our method, encompassing grid-based adaptive matching guidance, sparse Laplacian consensus, and implementation details. Section IV delineates the experimental results pertaining to feature matching, geometric matrix estimation, image registration, and visual localization, aiming to substantiate the superiority of our method. The concluding remarks are summarized in Section V.

II. Related Work

Traditional pipeline of feature matching typically begins with feature detection and description, where salient feature points are extracted based on the pixel structure of images. Classical feature extraction and description methods, such as SIFT [6], SURF [33], and KAZE [34], are widely used due to their low implementation cost. However, with the advent of deep learning, the effectiveness of handcrafted

method has been surpassed by data-driven techniques such as HardNet [7], SOSNet [35], SuperPoint [8], and Aslfeat [36]. In parallel, classical matching methods, characterized by their heuristic nature within the descriptor space, have been instrumental. These include Nearest Neighbor (NN), Mutual Nearest Neighbor (MNN), Nearest Neighbor Ratio (NNR) [6], Greedy Nearest Neighbor, and First Geometrically Inconsistent Nearest Neighbor [37]. Blob matching [38] has emerged as a comprehensive framework that amalgamates these diverse strategies. However, matching methods based on descriptor similarity often yield feature correspondence sets containing incorrect matches due to the limited discriminative capability of the feature descriptors.

To exploit the geometric transformation relationships between images, the classical RANSAC [9] employs a hypothesis-and-verification approach to estimate parametric models, such as homography or the fundamental matrix. Notable advancements in the realm of RANSAC include MLESAC [10], PROSAC [12], LO-RANSAC [11], and DegenSAC [39]. Recently, USAC [40] seamlessly integrates multiple resampling enhancements within a unified framework, while GC-RANSAC [13] incorporates the graph-cut algorithm in local optimization for enhanced efficiency. Methods such as MAGSAC [14] and its successor, MAGSAC++ [15], have demonstrated superior performance by eliminating the need for user-defined inlier-outlier thresholds and employing weighted least-squares fitting to update the quality function. Additionally, VSAC [16] introduces several innovative elements to enhance the random sampling framework, and ∇-RANSAC [17] uses relaxation techniques to estimate the sampling gradients, enabling learning the randomized robust estimation pipeline.

Since real-world image transformations are often non-rigid and the type of transformation is difficult to anticipate in advance, a class of relaxed constraint-based methods reduces the complex non-rigid transformation problem to more tractable objectives. For instance, GMS [19] grids the image region to find feature correspondences with similar motion, LPM [20] enforces local topological consistency, and LGSC [21] builds upon LPM by constructing a local graph structure model. Similar approaches include DBSCAN [41] and Adalaim [42]. Moreover, PFM [24] formulates the feature matching problem as a Markov Random Field. While these methods provide sufficient theoretical innovation, their practical performance is not clearly advantageous due to the relaxed constraints. For instance, locally-based approaches can impede subsequent image pose estimation due to their disregard of global information.

Estimating global transformations between images using non-parametric models has proven to be a robust approach for addressing the feature matching problem. Classic ICF [28] mitigates mismatches by evaluating their consistency with estimated correspondence functions. VFC [29] employs Tikhonov regularizers in a reproducing kernel Hilbert space to interpolate a vector field, offering a variant with sparse approximation to enhance computational efficiency. Bilateral Motion Fields (BMF) [30] capitalizes on the bilateral domain to reformulate a piecewise smooth constraint into a global modeling framework. Compact Fourier bases are employed in the CRC [31] to create a smooth function. GPMatch [32] introduces Gaussian process regression model via variational learning for feature matching. Furthermore, progressive smoothness consensus [43], incrementally establishes correspondences between two sets of features through a Bayesian-driven matching expansion process.

Deep learning techniques are now widely integrated with feature matching. Learning to find good correspondence [25] introduces a multi-layer perception-based network for correspondence estimation. Subsequently, a series of noteworthy follow-up studies have contributed to this domain, including OANet [26] and Nm-Net [44]. LMCNet [27] leverages the concept of learning motion coherence to effectively filter out mismatches. ConvMatch [18] employs Convolutional Neural Network (CNN) as the backbone to capture contextual information, which is instrumental in the estimation of motion fields. Another approach involves direct establishment of correspondences from two sets of feature points, including SuperGlue [45], SGMNet [46], and ClusterGNN [47]. However, these data-driven techniques are challenged by high hardware costs, and their effectiveness is closely tied to the quality of training data. The performance of such models is limited by the trade-off between accuracy and generalization, which is influenced by the model's degree of fitting.

III. Methodology

Given two feature point sets $\{t_i\}_{i=1}^{T}$ and $\{r_i\}_{i=1}^{R}$ from a target image and a reference image, respectively, our goal is to establish one-to-one correspondences between two sets. This is primarily addressed through two fundamental components, namely, *Grid-based Adaptive Matching Guidance* and *Sparse Laplacian Consensus*, which are predicated upon core assumptions as outlined below:

1) *Motion coherence* signifies that valid correspondences exhibit similar motion patterns. Each accurate correspondence demonstrates motion coherence with at least one high-confidence correspondence connecting salient structural elements.
2) Undergoing any image transformation, *motion consistency* among accurate correspondences within localized regions is maintained, with the exception of areas encompassing structural boundaries.
3) Structural boundaries, which usually involve independent motions and dramatic depth changes, do not affect all surrounding correspondences. Ideally, high-confidence correspondences can guide the discovery of correspondences in the vicinity of structural boundaries.

A. Grid-Based Adaptive Matching Guidance

Seed correspondences play a pivotal role in delineating candidate correspondence regions and providing guidance for subsequent smooth estimation; hence, they are expected to remain accurate even sparse. The acquisition of these seed correspondences should be expeditious, as elaborated follows.

To commence, we initiate the computation of feature descriptor distance matrix $\mathbf{D} \in \mathbb{R}^{T\times R}$, where $\mathrm{D}_{i,j}$ represents

the Euclidean distance of descriptors associated to keypoint t_i and keypoint r_j. D_{i,j^w} indicates the w-th lowest value on i-th row of matrix $\mathbf{D}$. Similarly, $\mathrm{D}_{i^w,j}$ denotes the w-th lowest value on j-th column. Denoting the subscript $\Downarrow$ as extracting the index pair (i, j) from $D_{i,j}$, we employ the intersection of MNN and NNR to establish *initial correspondence* set $\mathcal{S}_0$:

$$\begin{aligned} \mathcal{S}_{MNN} &= \{\mathrm{D}_{i,j^1} = \mathrm{D}_{i^1,j}\}_{\Downarrow}, \\ \mathcal{S}_{NNR} &= \{\mathrm{D}'_{i,j^1} < \tau\}_{\Downarrow}, \\ \mathcal{S}_0 &= \mathcal{S}_{MNN} \cap \mathcal{S}_{NNR}, \end{aligned} \tag{1}$$

where $\mathrm{D}'_{i,j^1} = \frac{\mathrm{D}_{i,j^1}}{\mathrm{D}_{i,j^2}}$ and $\tau \leq 1$ is a preset constant.

1) Motion Statistical Constraints: Beside the descriptor space, we employ grid-based motion statistics in geometric space, which have been proven effective [19], to efficiently obtain reliable seed correspondences. Both target and reference images are divided into $n_c \times n_c$ grids, and the score s is calculated as the number of correspondences from $\mathcal{S}_0$ distributed on a grid pair and requires only a cost of $O(N)$, where N is the cardinality of *initial correspondence* set $\mathcal{S}_0$. Based on differences in motion statistics between correct and false correspondences [19], seed correspondence set $\mathcal{S}_1$ can be simply determined by a threshold η:

$$\mathcal{S}_1 = \{c_i : s_i > \eta\}, \tag{2}$$

where $\eta = \alpha \sqrt{|\mathcal{S}_0|}$ and c_i is a feature correspondence of $\mathcal{S}_0$, s_i is the score of a grid pair where c_i is located, $\overline{|\mathcal{S}_0|} = |\mathcal{S}_0|/n_c^2$, and α is a hyperparameter empirically set to 1.

2) Selection of Candidate Correspondences: The process of selecting appropriate candidate correspondences is of paramount importance following the acquisition of a sparse and reliable seed correspondence set $\mathcal{S}_1$. Conventional matching often employs simple tricks (*e.g.,* NN) to form a candidate correspondence set. However, this approach may impose limitations on the subsequent correspondence pruning. For instance, such candidate correspondence set under motion discontinuities would render smoothness-based global modeling vulnerable. Consequently, we adopt a strategy that utilizes grid pairs as a guiding framework, allowing us to construct candidate correspondences with adaptive distributions.

A grid pair can be denoted by $\left(C_{a_m,b_m}, C_{a'_m,b'_m}\right)$, where a_m and b_m are elements of the set $\{1, 2, \ldots, n_c\}$, as are a'_m and b'_m. To quantify the distance between two grid pairs, denoted as $(C_{a_i,b_i}, C_{a'_i,b'_i})$ and $(C_{a_j,b_j}, C_{a'_j,b'_j})$, we employ the Chebyshev distance between motion vectors of them, which is defined as follows:

$$\max(|a'_j - a_j - a'_i + a_i|, |b'_j - b_j - b'_i + b_i|). \tag{3}$$

In this case, we organize grid pairs that satisfy Eq. (2), guided by two fundamental principles: (i) Within each group of grid pairs, there is a requirement that for every grid pair, there must be at least one other grid pair within the same group with a distance no greater than μ (a predetermined constant integer); (ii) Conversely, the distance between grid pairs belonging to different groups must exceed μ.

As illustrated in Fig. 2, seed correspondences (lines) belonging to different groups of grid pairs are visually differentiated by distinct colors. We identify the minimal rectangle areas that encompass a group of grid pairs in two images and denote them with a pair of dashed boxes. The dimensions of a solid box is expanded by 2μ, with its center coinciding with that of the dashed box. Subsequently, a candidate correspondence set is chosen from a pair of solid boxes, denoted as a block pair. And candidate correspondences are usually chosen among the nearest neighbor descriptor matches for each feature point that align with block pairs. Afterwards, each candidate correspondence set would undergo a refinement process through an outlier pruning method.

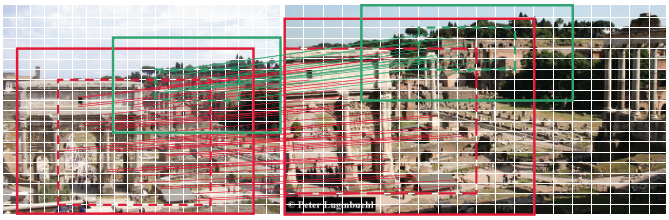

Fig. 2. Grid-based candidate correspondence selection. Lines represent seed correspondences, and the grids in which they are distributed are grouped into different blocks and indicated by dashed boxes. Candidate correspondences are selected within the solid boxes, which are expanded from the dotted boxes.

In summary, the grid-based adaptive matching guidance offers a straightforward and effective means to generate an adaptive number of candidate correspondence sets. This method can prove particularly advantageous for the estimation of smooth functions in complex scenarios, as will be substantiated in subsequent experimental evaluations.

B. Sparse Laplacian Consensus

A candidate correspondence set $\mathcal{S}_{c,j} = \{c_i = (\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^{N_j}$, where $\mathbf{x}_i \in \mathbb{R}^2$ and $\mathbf{y}_i \in \mathbb{R}^2$, denotes spatial coordinates of two corresponding keypoints. Our objective is to identify the inlier set, denoted as $\mathcal{I}_j \subseteq \mathcal{S}_{c,j}$.

Motion consistency is observed among inliers, *i.e,* closer true correspondences exhibits more similar motions. Based on this observation, we propose a novel graph-based formulation for encoding candidate correspondences to effectively filter out outliers.

1) Laplacian Motion Coherence: To describe a candidate correspondence set, we construct a fully-connected undirected graph $G = \{V, E\}$ (also called Markov network), where each node in V denotes a candidate correspondence and E includes the edges connecting any two nodes. The weights of edges are computed by $w_{i,j} = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|_2^2}{\delta^2}\right)$, where δ is a preset constant. Afterwards, the adjacency matrix of graph G is defined as $\boldsymbol{W} = [w_{i,j}]$, the degree matrix as $\boldsymbol{\Delta} = diag\left(\left[d_i = \sum_j w_{i,j}\right]\right)$, and the Laplacian matrix as $\mathbf{L} = \boldsymbol{\Delta} - \boldsymbol{W}$. Notably, $\boldsymbol{v} = [v_{i,j}]$ indicates that the components of matrix or vector $\boldsymbol{v}$ are $v_{i,j}$, and $diag(\boldsymbol{v})$ indicates that the components of $\boldsymbol{v}$ form a diagonal matrix. Our goal is to fit a mapping function $\mathbf{f}$ from the candidate correspondence set $\mathcal{S}_{c,j}$, *i.e.,* $\forall i \in \mathbb{N}_{N_i}, \mathbf{y}_i = \mathbf{f}(\mathbf{x}_i)$. For convenience, the number of candidate correspondences in what follows in this section is denoted by N rather than N_j. Hence, we formulate this problem as:

$$\min_{\mathbf{f}} \sum_{i}^{N} \|\mathbf{y}_i - \mathbf{f}(\mathbf{x}_i)\|_2^2 + \frac{\lambda}{2} \sum_{i,j} w_{i,j} \|\mathbf{f}(\mathbf{x}_i) - \mathbf{f}(\mathbf{x}_j)\|_2^2, \tag{4}$$

where $\|\mathbf{y}_i - \mathbf{f}(\mathbf{x}_i)\|_2^2$ penalizes the deviation of the mapping coordinate $\mathbf{f}(\mathbf{x}_i)$ from the output coordinate $\mathbf{y}_i$, λ is a balance factor, and $w_{i,j}\|\mathbf{f}(\mathbf{x}_i) - \mathbf{f}(\mathbf{x}_j)\|_2^2$ is a smoothness term which penalizes the mapping variation by $w_{i,j}$.

By aggregating $\mathbf{y}_i$ and $\mathbf{f}(\mathbf{x}_i)$ into matrix forms $\mathbf{Y} = [\mathbf{y}_i] \in \mathbb{R}^{N\times 2}$ and $\mathbf{F} = [\mathbf{f}(\mathbf{x}_i)] \in \mathbb{R}^{N\times 2}$, Eq. (4) becomes:

$$\min_{\mathbf{F}} \|\mathbf{Y} - \mathbf{F}\|_F^2 + \lambda \mathrm{tr}(\mathbf{F}^\top \mathbf{L}\mathbf{F}), \tag{5}$$

where $\|\cdot\|_F$ is the Frobenius norm, $\mathrm{tr}(\cdot)$ denotes the trace, and $\mathbf{F}^\top\mathbf{L}\mathbf{F}$ is used as a regularization term which imposes a smooth constraint on the mapping coordinates $\mathbf{F}$ as the graph signal [48].

Theorem 1: According to the regularization theorem [49], a smoothness-based function interpolating a sample set can be solved by minimizing a regularized risk functional, which is based on the Tikhonov regularization [50] in a Reproducing Kernel Hilbert Space [51] $\mathcal{H}$ as follows:

$$\min_{\mathbf{f}\in\mathcal{H}} \sum_{i=1}^{N} \|\mathbf{y}_i - \mathbf{f}(\mathbf{x}_i)\|_2^2 + \lambda\|\mathbf{f}\|_{\mathcal{H}}^2, \tag{6}$$

where $\|\cdot\|_{\mathcal{H}}$ represents the norm of $\mathcal{H}$.

With K as a matrix-valued kernel function, the representer theorem [52] derives a solution as:

$$\mathbf{f}(\mathbf{x}) = \sum_{n=1}^{N} \mathbf{c}_i K(\mathbf{x}, \mathbf{x}_i), \tag{7}$$

where $\mathbf{c}_i \in \mathbb{R}^2$ is a parameter vector. In particular, the common scalar kernel is radial basis function as $K(\mathbf{x}, \mathbf{x}_i) = exp(-\beta\|\mathbf{x} - \mathbf{x}_i\|_2^2)$ with β as a constant factor.

Clearly, assuming that $\beta = \frac{1}{\delta^2}$, we can represent the mapping function $\mathbf{f}$ as follows:

$$\mathbf{f}(\mathbf{x}_i) = \sum_{n=1}^{N} w_{i,n}\mathbf{c}_n. \tag{8}$$

By aggregating $\mathbf{c}_n$ into a matrix form $\mathbf{c} = [\mathbf{c}_n] \in \mathbb{R}^{N\times 2}$, the matrix $\mathbf{F}$ of mapping coordinates can be denoted as:

$$\mathbf{F} = \mathbf{W}\mathbf{c}. \tag{9}$$

In this case, problem Eq. (5) can be rewritten as follows:

$$\min_{\mathbf{c}} \|\mathbf{Y} - \mathbf{W}\mathbf{c}\|_F^2 + \lambda \mathrm{tr}(\mathbf{c}^\top\mathbf{W}^\top\mathbf{L}\mathbf{W}\mathbf{c}). \tag{10}$$

Observing that the function (10) is convex, the optimal solution of coefficient $\mathbf{c}$ can be obtained by solving the following linear system with $\mathbf{I}$ as an identity matrix:

$$(\mathbf{I} + \lambda\mathbf{L})\mathbf{W}\mathbf{c} = \mathbf{Y}. \tag{11}$$

2) Mixture Model Construction: As candidate correspondences often contain outliers in high proportions, directly applying Eq. (10) with the assumption that all are inliers is problematic. Therefore, we make the following reasonable assumptions: (i) all candidate correspondences are subject to independent identical distributions; (ii) the noise of inliers in each component is Gaussian with zero mean and uniform standard deviation σ; (iii) the distribution of outliers is uniform $\frac{1}{a}$, where a is a constant representing the volume of output space. With a latent variable $z_i \in \{0, 1\}$ indicating whether the i-th correspondence $(\mathbf{x}_i, \mathbf{y}_i)$ is an inlier ($z_i = 1$) or an outlier ($z_i = 0$), the number of inliers follows a binomial distribution $B(N, \gamma)$, *i.e.*, $p(z_i = 1) = \gamma$, and $p(z_i = 0) = 1 - \gamma$. Let $\mathbf{X} = [\mathbf{x}_i] \in \mathbb{R}^{N\times 2}$, the joint likelihood function is

$$\begin{aligned} p(\mathbf{Y}|\mathbf{X}, \boldsymbol{\theta}) &= \prod_{i=1}^{N} \sum_{z_i} p(\mathbf{y}_i, z_i|\mathbf{x}_i, \theta) \\ &= \prod_{i=1}^{N} \left(\frac{\gamma}{2\pi\sigma^2} e^{-\frac{\|\mathbf{y}_i - \mathbf{f}(\mathbf{x}_i)\|^2}{2\sigma^2}} + \frac{1-\gamma}{a} \right), \end{aligned} \tag{12}$$

where $\boldsymbol{\theta} = \{\mathbf{f}, \sigma^2, \gamma\}$ contains unknown parameters. $p(\theta)$ expresses the prior distribution of θ in a Bayesian view, and thus the Maximum A Posteriori (MAP) takes the form:

$$\boldsymbol{\theta}^* = \arg\max_{\boldsymbol{\theta}} p(\boldsymbol{\theta}|\mathbf{X}, \mathbf{Y}) = \arg\max_{\boldsymbol{\theta}} p(\mathbf{Y}|\mathbf{X}, \boldsymbol{\theta}) p(\boldsymbol{\theta}), \tag{13}$$

which has an equivalent energy function as follows:

$$E(\boldsymbol{\theta}) = -\ln p(\boldsymbol{\theta}) - \prod_{i=1}^{N} \ln \sum_{z_i} p(\mathbf{y}_i, z_i|\mathbf{x}_i, \theta). \tag{14}$$

3) Expectation-Maximization Solution: We utilize the versatile and efficient EM algorithm [53] to cope with the existence of latent variables, which iterates between two steps: an expectation step (E-step) and a maximization step (M-step).

After omitting some terms that are independent of $\boldsymbol{\theta}$ from Eq. (14), we can obtain the complete-data log likelihood as:

$$\begin{aligned} \mathcal{Q}(\boldsymbol{\theta}, \boldsymbol{\theta}^{old}) = &-\frac{1}{2\sigma^2} \sum_{i=1}^{N} p_i \|\mathbf{y}_i - \mathbf{f}(\mathbf{x}_i)\|^2 - \ln\sigma^2 \sum_{i=1}^{N} p_i \\ &+ \ln(1-\gamma) \sum_{i=1}^{N} (1 - p_i) + \ln\gamma \sum_{i=1}^{N} p_i - \ln p(\boldsymbol{\theta}), \end{aligned} \tag{15}$$

where $p_i = P(z_i = 1|\mathbf{x}_i, \mathbf{y}_i, \theta^{old})$ indicates the posterior probability of z_i. In this case, E-step and M-step are expressed as below.

E-step: Following *i.i.d.* assumption for each correspondence, the posterior probabilities can be derived by current parameters as follows:

$$p_i = \frac{\gamma e^{-\frac{\|\mathbf{y}_i - \mathbf{f}(\mathbf{x}_i)\|^2}{2\sigma^2}}}{\gamma e^{-\frac{\|\mathbf{y}_i - \mathbf{f}(\mathbf{x}_i)\|^2}{2\sigma^2}} + (1-\gamma)\frac{2\pi\sigma^2}{a}}. \tag{16}$$

M-step: Based on $\boldsymbol{\theta}^{new} = \arg\max_{\boldsymbol{\theta}} \mathcal{Q}(\boldsymbol{\theta}, \boldsymbol{\theta}^{old})$, the updating rules can be obtained by setting the derivatives of Eq. (15) about σ^2 and γ to zero. In the case where $\mathbf{P} = diag([p_i])$ and $tr(\cdot)$ denotes the trace, we can obtain:

$$\sigma^2 = \frac{tr((\mathbf{Y} - \mathbf{F})^\top \mathbf{P}(\mathbf{Y} - \mathbf{F}))}{2tr(\mathbf{P})}, \tag{17}$$

$$\gamma = \frac{\mathrm{tr}(\mathbf{P})}{N}. \tag{18}$$

Considering the terms related to $\mathbf{f}$, we obtain the following functional under flat priors for σ and γ:

$$\varepsilon(\mathbf{f}) = \frac{1}{2\sigma^2} \sum_{i=1}^{N} p_i \|\mathbf{y}_i - \mathbf{f}(\mathbf{x}_i)\|^2 - \ln p(\mathbf{f}). \tag{19}$$

Let the mapping function $\mathbf{f}$ be represented by Eq. (8) and Eq. (9), the term $-\ln p(\mathbf{f})$ translates to $\lambda \mathrm{tr}(\mathbf{c}^\top \mathbf{W}^\top \mathbf{L}\mathbf{W}\mathbf{c})$. The functional (19) can be reformulated as:

$$\min_{\mathbf{c}} \frac{1}{2\sigma^2}\|\mathbf{P}^{1/2}(\mathbf{Y}-\mathbf{W}\mathbf{c})\|_F^2 + \lambda \mathrm{tr}(\mathbf{c}^\top \mathbf{W}^\top \mathbf{L}\mathbf{W}\mathbf{c}). \tag{20}$$

It is easy to deduce that Eq. (20) is convex, and the optimal solution for the coefficients $\mathbf{c}$ can be derived as follows:

$$(\mathbf{P} + 2\lambda\sigma^2\mathbf{L})\mathbf{W}\mathbf{c} = \mathbf{P}\mathbf{Y}. \tag{21}$$

Once the EM algorithm converges, the inlier set can be obtained by a threshold on the posterior probability, *i.e.*:

$$\mathcal{I} = \{(\mathbf{x}_i, \mathbf{y}_i) : p_i > \epsilon, i \in \mathbb{N}_N\}, \tag{22}$$

where ϵ is a preset constant and empirically set to 0.85.

If there are more than one candidate correspondence sets $\{\mathcal{S}_{c,j}\}_{j=1}^{J}$, the final matching result is a concatenation of multiple inlier sets $\{\mathcal{I}_j\}_{j=1}^{J}$.

4) Seed Correspondence Initialization: Candidate correspondences typically exhibit a high outlier ratio. To mitigate this interference, seed correspondences $\mathcal{S}_1$ can be employed for initialization to avoid EM convergence to local optima. Specifically, the posterior probability p_i of correspondence $(\mathbf{x}_i, \mathbf{y}_i)$ is initialized by the following rule:

$$p_i^{init} = \begin{cases} 1, & (\mathbf{x}_i, \mathbf{y}_i) \in \mathcal{S}_1, \\ \zeta, & otherwise, \end{cases} \tag{23}$$

where ζ is a constant (*e.g.,* 10^{-4}) for computing stability.

Seed correspondence set $\mathcal{S}_1$ facilitates the EM algorithm in converging to the global optimum, thereby ensuring the efficiency and accuracy, as demonstrated in the following experiments.

5) Sparse Approximation: Linear system (21) requires high space complexity (*i.e.,* $O(N^2)$) and time complexity (*i.e,* $O(N^3)$). To mitigate the computational burden, a sparse approximation is utilized to search for a solution in a much smaller space without a significant decrease in accuracy, as demonstrated in [54] and [55]. Assuming $M \ll N$, we randomly select a point set $\{\tilde{\mathbf{x}}_m : m \in \mathbb{N}_M\}$ from the sample set $\{\mathbf{x}_i : i \in \mathbb{N}_N\}$. It has been observed that arbitrary selection of a subset does not essentially reduce the accuracy of motion coherence fitting, and this strategy would significantly enhance the efficiency of smooth function estimation rather without over-fitting. In this scenario, the mapping function $\mathbf{f}$ takes the following form:

$$\mathbf{f}(\mathbf{x}_i) = \sum_{m=1}^{M} w_{i,m}\mathbf{c}_m, \tag{24}$$

with the coefficients $\widetilde{\mathbf{c}} = [\mathbf{c}_m] \in \mathbb{R}^{M\times 2}$. In the case of $\widetilde{\mathbf{W}} = [w_{i,m}] \in \mathbb{R}^{N\times M}$, coefficients $\widetilde{\mathbf{c}}$ can be derived by:

$$(\widetilde{\mathbf{W}}^\top \mathbf{P}\widetilde{\mathbf{W}} + 2\lambda\sigma^2\mathbf{A}^\top\widetilde{\mathbf{L}}\mathbf{A})\widetilde{\mathbf{c}} = \widetilde{\mathbf{W}}^\top\mathbf{P}\mathbf{Y}, \tag{25}$$

where $\mathbf{A} \in \mathbb{R}^{M\times M}$ is the adjacency matrix of sub-graph $\widetilde{G}$ constructed by the point subset, and $\widetilde{\mathbf{L}} \in \mathbb{R}^{M\times M}$ is the associated Laplacian matrix. Eventually, the sparse approximation aids in achieving $O(N)$ complexity in both the time and space for our correspondence pruning method.

In summary, the devised grid-based adaptive matching guidance establishes multiple candidate correspondence sets, and subsequent sparse Laplacian consensus filters out outliers based on motion coherence. We call this feature matching framework as *Grid-guided Sparse Laplacian Consensus*.

C. Differences From Related Matching Methods

Our GSLC is inspired by GMS [19] and regularized methods such as VFC [29] and CRC [31], yet exhibits distinct differences from them. GMS employs grid-based statistical constraints to eliminate incorrect matches from a candidate correspondence set and utilizes a strategy involving multiple scales and rotations in the spatical domain. However, its relaxed constraints limit its effectiveness in estimating real motion fields and perform well primarily when there is an abundance of candidate matches; otherwise, the results may exhibit significant inaccuracies. In contrast, our GSLC primarily uses grid constraints as a pre-processing step to obtain seed correspondences and candidate correspondence sets, providing assistance for subsequent non-parametric error elimination through sparse Laplacian consensus.

Unlike methods such as VFC and CRC, our GSLC innovatively formulates a smooth estimation problem through graph Laplacians, and the grid-based adaptive matching guidance addresses the challenge faced by traditional regularization methods in handling multi-motion scenarios. Comparative experimental results, validating this assertion, are presented in IV-A4. In subsequent experimental comparisons, GMS and CRC, used as reference methods, consistently falls short of GSLC in terms of matching performance.

D. Implementation Details

Our method involves five main parameters: τ, δ, λ, and M. Threshold τ is to determine the initial correspondence set $\mathcal{S}_0$ and normally set to 1. Parameter δ determines the weights $w_{i,j}$ of edges and the mapping function $\mathbf{f}$. We normalize the data to $[0, 1]^2$ during correspondence pruning and empirically set δ to 1. A single parameter δ is sufficient due to the grid-based adaptive matching guidance, which has already separated feature correspondences with significantly different motions. λ controls the trade-off for smoothness term and is set to 0.01. M is the number of randomly chosen points for smooth function estimation and empirically set to 20.

IV. Experimental Results

We conduct comprehensive experiments to evaluate GSLC, including general feature matching, homography & fundamental matrix estimations, and image registration task. Descriptor distance computation is performed by open source toolbox VLFeat [56]. Experiments are run on a desktop with a 2.90 GHz Intel Core CPU, 32 GB memory, and MATLAB R2022a code.

A. General Feature Matching

General feature matching refers to constructing point-to-point correspondences between two images. Three evaluation

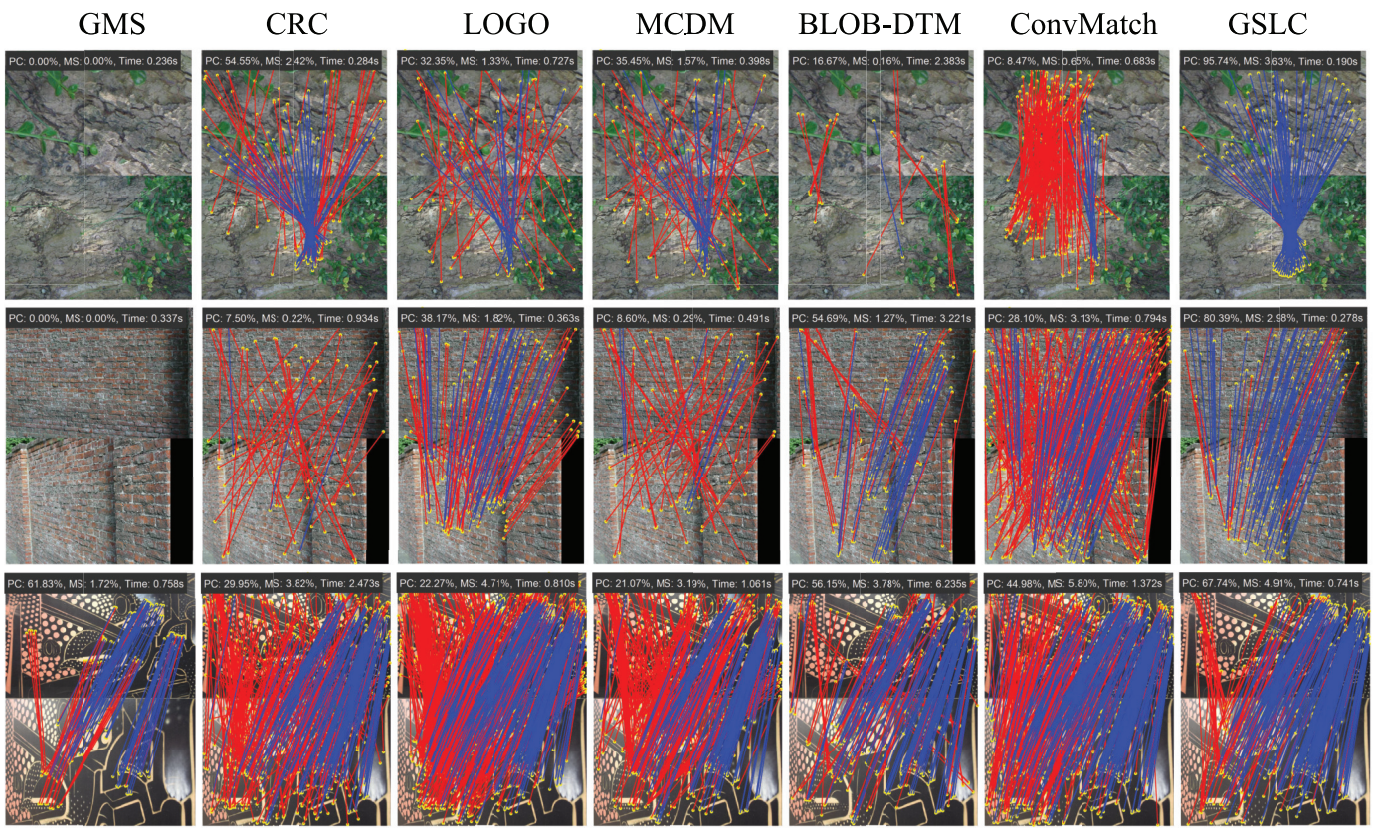

Fig. 3. Qualitative comparisons on VGG and Hannover. PC (precision), MS (matching score), and Time are used as evaluation metrics. The yellow dots represent feature points, and the blue lines denote correct correspondences with the red lines false correspondences. Best viewed in color and zoomed in.

metrics proposed by Heinly et al. [57] are adopted: *putative match ratio*, *precision*, and *matching score*. The putative match ratio, PMR = $\frac{\#correspondences}{\#features}$, which is the ratio of obtained correspondences to all feature points of target image, thus denoting the selectivity of methods. The precision, PC = $\frac{\#correct\ correspondences}{\#correspondences}$, which represents the correctness of the produced correspondences. The matching score, MS = $\frac{\#correct\ correspondences}{\#features}$, which shows the proportion of correct correspondences to features in the target image.

1) Comparisons With Existing Methods: For a comprehensive evaluation of our approach, we select twelve state-of-the-art or commonly used algorithms for comparison: NNR [6], GMS [19], MAGSAC++ [15], LPM [20], GLPM [58], LOGO [22], MCDM [23], CRC [31], GPMatch [32], OANet [26], CLNet [59], BLOB-DTM [38], and ConvMatch [18]. In particular, NNR is a classic heuristic filter with the threshold as 0.8. OANet, CLNet, and ConvMatch are deep learning based matching methods and run on the CPU for fairness of evaluation. Classic SIFT [6] is chosen for feature extraction and description. In particular, correspondence pruning methods are performed on the results of NN filter.

Datasets: The well-known VGG[1] [60] and its extended version Hannover[2] [61] contain totally 13 groups of images with the medium resolution, and each group constitutes 5 image pairs with different deformation levels. These two datasets provide homography matrices as the ground-truth, and a correspondence is considered as true if the distance to mapping coordinates by ground-truth is less than 5 pixels.

Discussions of Results: Fig. 3 shows a qualitative comparison of our method with six representative methods. Our method is demonstrated to be robust to various deformations, accurate for images with large view changes, and capable of constructing substantial matches. The results are quantified in Table I, where our method achieves the highest matching score with the greatest precision compared to other state-of-the-art methods. Simultaneously, our method exhibits only a slightly slower processing time compared to NNR and GMS, thus ensuring real-time performance.

2) Ablation Study: To specifically analyze the effect of technical components in our method, we conduct an ablation study, the results of which are presented in Table II. NNR [6], GMS [19], CRC [31], and ConvMatch [18] are chosen as comparison methods, where NNR is the commonly employed heuristic approach, GMS is a voting-based filtering method via grid constraints, CRC is a mapping function estimation method based on smoothness consistency, and ConvMatch is an advanced deep learning algorithm. The quantitative results of ablation study are presented in Table II.

Grid-based candidate correspondence construction (G.C3) is tailored to generate adaptive number of candidate correspondence sets, effectively addressing the challenge of significant motion discrepancies within scenes with severe deformations. From Table II, Grid-based candidate

[1]https://www.robots.ox.ac.uk/~vgg/research/affine/
[2]http://www.tnt.uni-hannover.de/project/feature_evaluation/

TABLE I

THE QUANTITATIVE RESULTS OF FEATURE MATCHING ON VGG AND HANNOVER DATASETS, WHERE OUR GSLC IS COMPARED WITH OTHER TWELVE METHODS INCLUDING NNR [6], GMS [19], MAGSAC++ [15], LPM [20], GLPM [58], LOGO [22], MCDM [23], CRC [31], OANET [26], CLNET [59], BLOB-DTM [38], GPMATCH [32] AND CONVMATCH [18]. PUTATIVE MATCH RATIO (PMR, %), PRECISION (PC, %), MATCHING SCORE (MS, %), AND TIME (MS) ARE USED AS EVALUATION METRICS, WHERE THE **bold** INDICATES THE BEST IN ADDITION TO PMR DENOTING THE SELECTIVITY OF MATCHERS

Methods	NNR	GMS	MAGSAC++	LPM	GLPM	LOGO	MCDM	CRC	OANet	CLNet	BLOB-DTM	GPMatch	ConvMatch	**GSLC**
PMR	37.43	24.43	26.57	26.46	28.19	24.43	33.43	29.51	34.93	22.47	32.84	25.98	29.94	28.75
PC	52.74	63.89	73.17	64.16	71.30	57.83	57.88	70.28	68.63	69.66	60.21	69.33	72.04	**73.70**
MS	24.17	21.90	25.39	21.55	25.27	21.93	25.06	25.34	25.53	16.00	24.79	22.70	25.62	**25.79**
Time	**452.4**	456.7	629.1	681.8	1364	1342	617.4	1358	629.5	1213	4790	560.4	889.6	556.1

TABLE II

RESULTS OF ABLATION STUDY. VGG AND HANNOVER ARE TESTING DATASETS, AND PC (%), MS (%), AND TIME (MS) ARE REGARDED AS THE METRICS TO EVALUATE THE MATCHING RESULTS. W/O G.C3 STANDS FOR NO GRID-BASED CANDIDATE CORRESPONDENCE CONSTRUCTION. W/O S.C.I STANDS FOR NO SEED CORRESPONDENCE INITIALIZATION. W/O S.A STANDS FOR NO SPARSE APPROXIMATION. **Bold** INDICATES THE BEST

Methods		PC	MS	Time
NNR [6]		52.74	24.17	452.4
GMS [19]		63.89	21.90	456.7
CRC [31]		70.28	25.34	1358.2
ConvMatch [18]		72.04	25.62	889.6
GSLC	w/o G.C3	72.74	25.43	**364.2**
	w/o S.C.I	61.53	7.44	619.8
	w/o S.A.	**74.09**	**26.03**	7888.1
	full	73.70	25.79	556.1

TABLE III

RESULTS OF DESCRIPTOR GENERALITY TESTING. VGG AND HANNOVER ARE THE DATASETS, WHILE MS (%) AND PC (%) ARE REGARDED AS THE METRICS TO EVALUATE THE MATCHING RESULTS UNDER DIFFERENT DESCRIPTORS INCLUDING SIFT [6], KAZE [34], HARDNET [7], SUPOINT [8], AND ASLFEAT [36]. **Bold** INDICATES THE BEST

Method		NNR	GMS	CRC	ConvMatch	**GSLC**
SIFT [6]	PC	52.74	63.89	70.28	72.04	**73.70**
	MS	24.17	21.90	25.34	25.62	**25.79**
KAZE [34]	PC	52.45	58.57	61.57	64.73	**65.08**
	MS	25.61	27.88	29.59	26.15	**29.86**
HardNet [7]	PC	38.02	42.02	39.70	37.41	**42.28**
	MS	15.03	17.56	18.05	18.19	**18.22**
SuPoint [8]	PC	55.72	58.33	57.78	56.71	**58.67**
	MS	30.65	32.71	34.89	35.06	**35.10**
ASLFeat [36]	PC	63.05	64.38	67.42	57.92	**68.33**
	MS	36.18	37.22	46.94	20.23	**47.43**
Average	PC	52.40	57.44	59.35	57.76	**61.61**
	MS	26.33	27.45	30.96	25.05	**31.28**

correspondence construction can improve the matching accuracy. **Seed correspondence initialization** (S.C.I) leverages reliable correspondences as a guide to prevent the smoothness estimation from converging to local optima. Table II demonstrates that seed correspondence initialization significantly enhances matching performance in terms of both PC and MS. **Sparse approximation** (S.A.) accomplishes motion coherence fitting within a significantly reduced computational space, which can substantially reduce GSLC's computational time without fundamentally sacrificing accuracy.

3) Descriptor Generality Testing: To assess the matching performance of our GSLC under different descriptors, we conduct generality testing experiments. In addition to SIFT, we consider traditional descriptor KAZE [34], as well as deep learning-based descriptors HardNet [7], SuperPoint [8], and ASLFeat [36]. For brevity, NNR [6], GMS [19], CRC [31], and ConvMatch [18] are used as comparative methods to quantitatively illustrate the effectiveness of our GSLC. The NNR threshold is set to 1.25, and initial matches are constructed using the nearest neighbor approach. The matching results are presented in Table III, where it can be observed that GSLC consistently achieves the best matching performance when dealing with different descriptors compared to other methods, demonstrating its robustness and versatility.

4) Matching on Independent Motions: Independent motions are often encountered in practical scenarios, presenting a challenge for matching methods in terms of geometric constraints. Differences in motions of different objects make it difficult to maintain geometric consistency. The public dataset Adelaide [62] contains 38 image pairs, and each image pair exhibits multiple motions or homographies. This dataset provides the initial correspondences with coordinates and scores as well as correct correspondences as the ground truth. As it provides no descriptor information, the matching can only be handled by indirect matching methods, while our method can use the scores as a prior to construct seed correspondences. In particular, Adelaide dataset provides sparse matches to be removed for each image pair, leading to generally quick processing times for matching methods. Qualitative comparisons are shown in Fig. 4, where our method can find inliers accurately under the grid-based adaptive matching guidance. When faced with multiple independently moving objects, our GSLC significantly outperforms other smoothness-based methods VFC [29] and CRC [31] and deep learning-based OANet [26]. With *precision* (Pre.), *recall* (Rec.), *F1-score* (F1.), and Time as metrics, the quantitative results are present in Table IV, where our method exhibits the highest recall

Fig. 4. Qualitative comparisons on representative image pairs from Adelaide [62]. The blue lines denote correct correspondences and the red lines represent false correspondences.

TABLE IV

FEATURE MATCHING ON ADELAIDE WITH MULTIPLE MOTIONS OR HOMOGRAPHIES. PRE. (%), REC. (%), F1. (%), AND TIME (MS) ARE REGARDED AS THE METRICS TO EVALUATE THE MISMATCH REMOVAL RESULTS. **Bold** INDICATES THE BEST

Methods	Pre.	Rec.	F1.	Time
GMS [19]	91.86	86.27	88.81	**0.82**
LPM [20]	**98.86**	89.47	93.76	8.19
MAGSAC++ [15]	90.74	78.37	81.07	46.5
CRC [31]	92.68	86.23	88.73	2.02
LOGO [22]	98.23	84.59	90.38	21.8
MCDM [31]	97.74	89.86	93.38	13.4
GPMatch [32]	87.97	87.13	87.57	20.5
OANet [26]	88.16	88.15	86.52	98.1
CLNet [59]	74.65	33.55	45.21	87.0
ConvMatch [18]	51.62	32.99	39.32	25.3
GSLC	94.87	**93.71**	**94.04**	19.2

and F1-score thus superior performance compared to other methods.

5) Analysis of Sparse Approximation: To comprehensively analyze the impact of sparse approximation on feature matching, we applied Sparse Laplacian Consensus (III-B) to the outlier removal tasks. The datasets used include VGG+Hannover [60], [61] and Adelaide [62]. In the VGG+Hannover dataset, initial correspondences were obtained using the NNR filter with a threshold of 1.25 applied to SIFT descriptors, and the ground truth (GT) was determined based on the homography matrix provided by the dataset. For the Adelaide dataset, both the initial correspondences and GT were directly provided. Using F1-score and Time as evaluation metrics, we assessed the performance of outlier removal under different values of M. Quantitative experimental results are presented in Fig. 5, which demonstrate that sparse approximation significantly reduces runtime without compromising matching accuracy, with outstanding performance observed when $M = 20$.

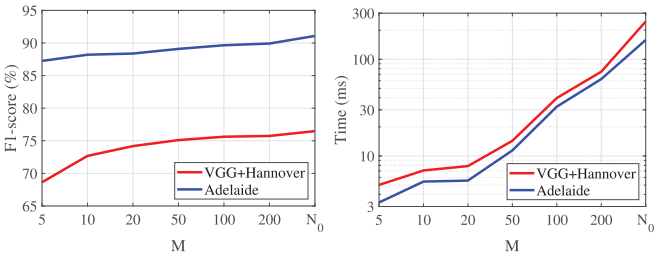

Fig. 5. Analysis of sparse approximation with respect to the parameter M focusing on the runtime associated with the outlier removal described in III-B, where N_0 represents the number of candidate correspondences.

B. Geometric Matrix Estimation

Subsequently, our method is evaluated in homography & fundamental matrix estimation and compared with other state-of-the-art methods. Based on the point-to-point correspondences between two images after feature matching, a model estimator (*e.g.,* RANSAC [9]) can derive a geometric model (homography or fundamental matrix) describing image transformations. During the geometric matrix estimation experiments, the candidate correspondences required by certain methods are uniformly the results from Nearest Neighbor filter.

Datasets: Hpatches[3] benchmark [63] comprises 580 image pairs, each accompanied by accurate ground-truth homography matrices. To provide a comprehensive assessment, the dataset is structured into a total of 116 distinct scenes, with each scene encompassing one source image and five corresponding target images, yielding five image pairs per scene. Additionally, it is noteworthy that a maximum limit of 4000 keypoints has been enforced for each individual image, ensuring consistency in the evaluation process.

For fundamental matrix estimation, FM-Bench[4] offers a comprehensive collection of four substantial datasets: *TUM* [64], which is extensively utilized in Simultaneous Location and Mapping applications, featuring indoor scenes with short-baseline images at a resolution of 480 × 640; *KITTI* [65], obtained through a camera affixed to a moving vehicle, comprises outdoor scenes with short-baseline images at a resolution of 370×1226; *T&T* [66] encompasses diverse scenes and objects, and includes image pairs with wide baselines at resolution of 1080 × 1920 or 1080 × 2048; *CPC* [67] gathers unstructured landmark images from Flicker, featuring

[3]https://github.com/hpatches/hpatches-dataset
[4]https://github.com/JiawangBian/FM-Bench

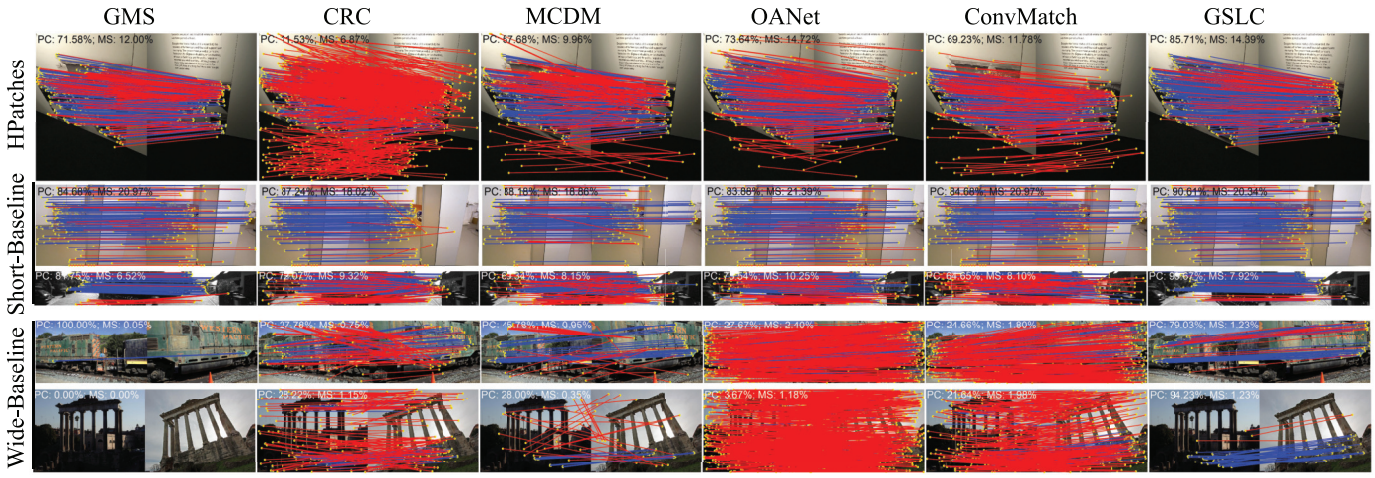

Fig. 6. Qualitative comparisons on representative image pairs from Hpatches, Short-Baseline (TUM and KITTI), and Wide-Baseline (T&T and CPC) datasets. The yellow dots represent feature points, and the blue lines denote correct correspondences with the red lines false correspondences. Best viewed in color.

wide-baseline image pairs at varying resolutions. Each of these four datasets has 1000 pairs of images and provides the fundamental matrices as the ground truth. The maximum number of extracted features in each image is set to 4000.

Metrics: Beside PC and MS, the *accuracy* (Acc.) of geometric matrix estimation, measured by the ratio of accurate estimates to all estimates, is an important metric. Homography error is computed as in [8] with an error of less than 4 pixels considered accurate. For fundamental matrix estimation, an estimated fundamental matrix with Normalized Symmetric Geometry Distance (NSGD) less than 0.05, as defined in FM-Bench [68], is considered accurate. The accuracy (Acc.) is defined as the ratio of accurate estimates to all estimates.

Discussion of Results: Qualitative matching results are shown in Fig. 6. With GMS [19], CRC [31], MCDM [23], OANet [26], CLNet [59], and ConvMatch [18] as competitors, our GSLC demonstrates superior matching performance in variety of scenarios (including indoor and outdoor, wide baseline and short baseline), especially in terms of precision. The quantitative comparison results of homography estimation are presented in Table V, where it can be seen that our GSLC obtains the highest matching scores and estimation accuracies with minimal time consumption, which is a clear advantage over other state-of-the-art methods.

Table VI and Table VII present the results of the fundamental matrix estimations. The former pertains to short baseline image pairs (*TUM* and *KITTI*), while the latter is for wide baseline image pairs (*T&T* and *CPC*). In the case of short baseline images, our GSLC exhibits highest matching score and accuracy, highlighting its advantages in geometric estimation. Furthermore, for wide baseline images, our GSLC demonstrates competitiveness with deep learning methods trained on similar images, underscoring the robustness of our method across different image types.

C. Compared to Learning-Based Direct Matching

Learning-based direct matching methods, such as SuperGlue [45] and SGMNet [46], have gained significant attention in recent years. These approaches utilize feature descriptor information to train matching models for accurate results. However, they exhibit two major drawbacks: high operational costs requiring substantial computational power, and reliance on pre-training with specific descriptors, limiting their generalizability across diverse descriptor types. In contrast, our GSLC does not necessitate training on specific descriptors and can seamlessly integrate with various descriptors while maintaining constant hyperparameters.

TABLE V

THE QUANTITATIVE COMPARISON OF HOMOGRAPHY ESTIMATION ON HPATCHES DATASET. PC (%), MS (%), ACC. (%), AND TIME (MS) ARE REGARDED AS THE METRICS TO EVALUATE THE EXPERIMENTAL RESULTS. **Bold** INDICATES THE BEST RESULTS

Methods	PC.	MS.	Acc.	Time
GMS [19]	80.28	22.36	72.59	518
LPM [20]	72.15	16.17	66.03	537
MAGSAC++ [15]	**86.65**	26.02	79.76	717
CRC [31]	77.85	25.99	73.45	530
LOGO [22]	65.70	21.32	73.79	2161
MCDM [31]	64.64	25.87	76.38	819
GPMatch [32]	82.55	24.80	75.69	573
OANet [26]	82.78	26.23	80.17	928
CLNet [59]	79.86	19.20	74.31	1107
ConvMatch [18]	85.06	25.91	78.28	924
GSLC	84.02	**26.50**	**85.34**	**273**

To quantitatively assess our GSLC method against these state-of-the-art techniques, we focus on pose estimation experiments, including homography and fundamental matrix estimation, presenting results that incorporate learning-based descriptors such as XFeat [69] and Aslfeat [36]. The experiments are run on CPUs to ensure fairness of comparisons.

Table VIII presents the homography estimation results of GSLC compared to state-of-the-art learning-based methods. Our approach demonstrates competitive matching accuracy

TABLE VI

THE QUANTITATIVE COMPARISON OF FUNDAMENTAL MATRIX ESTIMATION ON TWO SHORT-BASELINE IMAGE DATASETS (I.E., TUM AND KITTI). RED INDICATES THE BEST, BLUE RANKS THE SECOND, AND GREEN REPRESENTS THE THIRD

Methods	*TUM*			*KITTI*		
	Acc.	PC	MS	Acc.	PC	MS
NNR [6]	49.1	67.64	27.50	86.6	77.12	30.43
GMS [19]	62.0	91.69	24.08	90.3	95.67	28.01
LPM [20]	64.2	94.57	24.84	89.9	91.68	30.42
MSC+ [15]	65.0	89.67	30.04	88.4	93.65	31.06
MCDM [23]	62.7	83.66	30.17	90.9	94.52	30.75
CRC [31]	65.6	94.49	26.12	89.3	92.39	31.54
LOGO [22]	67.8	81.43	23.75	90.3	83.13	29.28
GPMatch [32]	61.0	95.13	25.50	85.4	93.55	27.34
OANet [26]	68.7	93.36	28.78	90.7	97.44	31.76
CLNet [59]	66.9	94.97	23.34	89.3	98.69	20.96
ConvMatch [18]	68.6	95.52	30.07	90.6	95.48	31.92
GSLC	70.7	88.11	30.49	91.0	94.68	32.00

TABLE VII

THE QUANTITATIVE COMPARISON OF FUNDAMENTAL MATRIX ESTIMATION ON TWO WIDE-BASELINE IMAGE DATASETS (I.E., T&T AND CPC). RED INDICATES THE BEST, BLUE RANKS THE SECOND, AND GREEN REPRESENTS THE THIRD

Methods	*T&T*			*CPC*		
	Acc.	PC	MS	Acc.	PC	MS
NNR [6]	27.8	24.83	4.42	6.00	20.62	2.83
GMS [19]	49.8	53.77	3.89	19.0	53.28	2.63
LPM [20]	50.2	42.95	3.00	38.8	52.98	2.76
MSC+ [15]	57.2	35.29	2.54	26.0	26.36	1.22
MCDM [23]	69.0	47.01	4.58	30.4	36.40	2.41
CRC [31]	59.4	39.94	4.59	23.4	33.49	2.63
LOGO [22]	50.6	44.80	2.25	20.1	37.77	1.33
GPMatch [32]	36.4	29.93	3.18	10.5	19.45	1.37
OANet [26]	67.6	44.26	4.81	37.1	39.07	4.00
CLNet [59]	71.8	51.72	5.71	45.3	55.67	3.69
ConvMatch [18]	72.3	76.08	4.55	44.6	79.41	2.84
GSLC	70.6	54.38	4.61	42.6	60.00	2.91

on image pairs conforming to homography, yielding matches more favorable for homography estimation (higher Acc.) compared to SuperGlue and SGMNet. Notably, GSLC accomplishes this in approximately 10% of the computation time, significantly enhancing its practical utility.

Table IX presents the fundamental matrix estimation results, comparing GSLC with SuperGlue and SGMNet. Using only SIFT as the feature descriptor, our GSLC demonstrates competitive matching precision and pose estimation accuracy on short-baseline dataset (TUM and KITTI). Due to the versatility of GSLC, *i.e.,* applicable to other descriptors without training, its combination with learning-based descriptors (Aslfeat and XFeat) achieves superior matching performance and pose estimation accuracy. Most importantly, GSLC incurs an order of magnitude lower computational cost, enhancing its practicality and usability.

TABLE VIII

HOMOGRAPHY ESTIMATION RESULTS. COMPARE OUR GSLC WITH LEARNING-BASED DIRECT MATCHING METHODS, SUPERGLUE [45] AND SGMNET [46]. THE **bold** INDICATES THE BEST

Feature	Matcher	PC	MS	Acc.	Time
SIFT	SuperGlue	47.71	23.97	57.59	6.7492
	SGMNet	84.63	32.11	81.21	3.3040
	GSLC	84.02	26.50	85.34	**0.2731**
XFeat	GSLC	74.60	57.89	84.31	0.4903
Aslfeat	GSLC	**86.91**	**61.62**	**88.62**	0.5295

D. Image Registration Task

To further exploit the application value of our method, we carry out the image registration task. After obtaining feature point matches, TPS [70] estimates the transformation to align the common regions of two images. In this experiment, the candidate correspondences for indirect matching methods are uniformly established by NNR with a threshold as 0.8.

Datasets and Metrics: RS [20] and *720Yun* [71] contain 92 image pairs and are chosen as experimental datasets including nonrigid transformation. Based on 20 pairs of landmark pixels randomly selected, *Root Mean Square Error* (RMSE), *MAximum Error* (MAE), and *MEdian Error* (MEE) are employed as metrics. The definitions of these metrics are as follows:

$$\text{RMSE} = \sqrt{1/M\sum_{i=1}^{M}(r_i - \mathcal{F}(s_i))^2}, \tag{26}$$

$$\text{MAE} = \max\left\{\sqrt{(r_i - \mathcal{F}(s_i))^2}\right\}_{i=1}^{M}, \tag{27}$$

$$\text{MEE} = \text{median}\left\{\sqrt{(r_i - \mathcal{F}(s_i))^2}\right\}_{i=1}^{M}. \tag{28}$$

Discussions of Results: Our method accurately aligns the images pixel by pixel, as demonstrated in Fig. 7. Compared to CRC [31] and CLNet [59], the registration presentation of our GSLC shows a distinct lack of splicing traces, and this difference is evident at the boundaries of the grids as indicated by the yellow dashed boxes in Fig. 7. Quantitative comparisons are shown in Table X, where our method has the lowest errors compared to other methods and is therefore advantageous in registration error reduction.

E. Loop Closure Detection

To further substantiate the practical utility of GSLC, we conduct the loop closure detection task based on feature matching [4], specifically targeting the recognition and validation of loop closures during robot navigation. The test datasets include CC (one of Oxford [72], 1237 images) and K00 (from KITTI Vision Suite [65], 4541 images).

DELG [73] is employed to extract global image features, given its superiority in LCD-related tasks, such as image

TABLE IX

RESULTS OF FUNDAMENTAL MATRIX ESTIMATION ON FMBENCH DATASETS. COMPARE OUR GSLC WITH LEARNING-BASED DIRECT MATCHING METHODS, SUPERGLUE [45] AND SGMNET [46]. OUR GSLC IS TRAINING-FREE AND INDEPENDENT OF SPECIFIC DESCRIPTORS, ALLOWING EASY ADAPTATION TO ANY FEATURE DESCRIPTORS. THE **bold** INDICATES THE BEST

Feature	Matcher	TUM				KITTI				T&T				CPC			
		PC	MS	Acc.	Time	PC	MS	Acc.	Time	PC	MS	Acc.	Time	PC	MS	Acc.	Time
SIFT	SuperGlue	92.1	41.0	67.0	1.43	94.2	44.2	90.7	3.57	62.4	11.8	**83.0**	18.4	66.8	10.3	46.7	17.3
	SGMNet	**96.3**	40.5	68.8	0.96	**97.1**	40.5	90.7	1.81	**76.8**	15.6	81.0	9.32	**82.9**	14.5	54.7	6.82
	GSLC	88.1	30.5	70.7	**0.06**	94.7	32.0	**91.0**	**0.13**	54.4	4.61	70.6	0.56	60.0	2.91	42.6	0.52
Aslfeat	GSLC	94.3	**87.2**	**72.5**	0.22	94.6	66.6	90.7	0.59	59.0	**26.4**	79.4	0.49	70.3	19.2	**66.1**	0.48
XFeat	GSLC	89.4	83.4	70.5	0.18	88.8	**70.4**	88.9	0.42	55.5	26.1	**83.0**	**0.40**	61.6	**23.1**	56.0	**0.42**

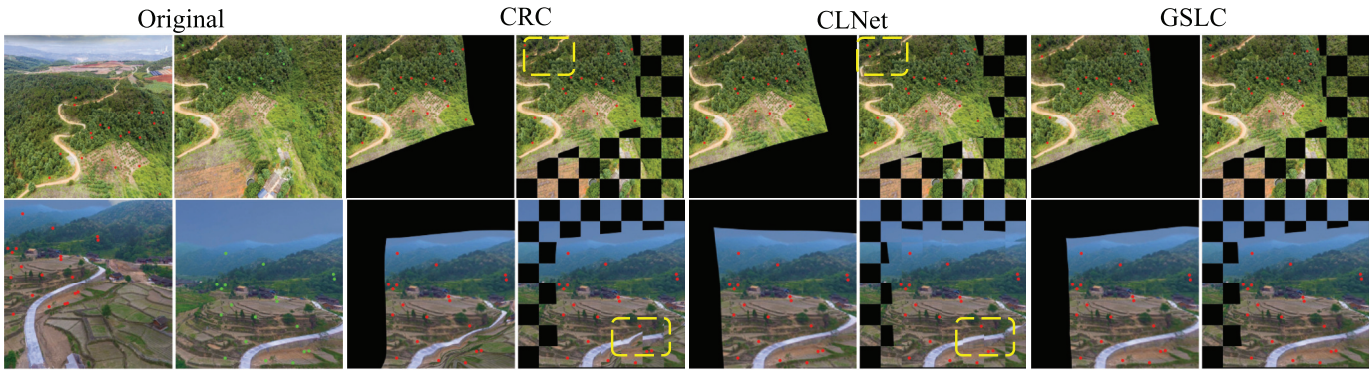

Fig. 7. Image registration examples of our GSLC compared to CRC [31] and CLNet [59], where the common regions of original images are stitched together checkerboard style. In each row, original image pairs contain the target image (left) and reference image (right), and two images corresponding each method are warped target image and stitched registration result, respectively. Green and red dots represent the landmarks. Best viewed zoomed in.

TABLE X

THE QUANTITATIVE RESULTS OF IMAGE REGISTRATION. THE AVERAGE VALUES AND STANDARD DEVIATIONS OF RMSE, MAE, AND MEE ARE USED FOR EVALUATION. **Bold** INDICATES THE BEST

Method	RMSE	MAE	MEE
NNR	266.4(±146.9)	533.5(±247.6)	355.7(±211.0)
GMS	223.3(±224.8)	450.1(±536.0)	296.7(±279.3)
LPM	95.96(±110.0)	189.6(±213.9)	128.0(±151.2)
MSC++	171.2(±135.6)	353.8(±273.4)	227.0(±185.8)
CRC	139.5(±141.3)	264.7(±252.2)	188.6(±195.8)
LOGO	199.4(±199.4)	405.9(±383.5)	264.4(±273.2)
MCDM	179.0(±173.9)	348.6(±315.5)	237.6(±239.7)
GPMatch	64.35(±173.9)	127.9(±189.7)	85.84(±135.6)
OANet	94.22(±131.8)	186.7(±256.9)	126.5(±179.4)
CLNet	65.54(±136.5)	125.5(±238.1)	88.48(±190.9)
ConvMatch	69.31(±132.2)	136.9(±236.5)	91.97(±183.9)
GSLC	**61.08(±103.5)**	**124.6(±204.7)**	**78.70(±134.5)**

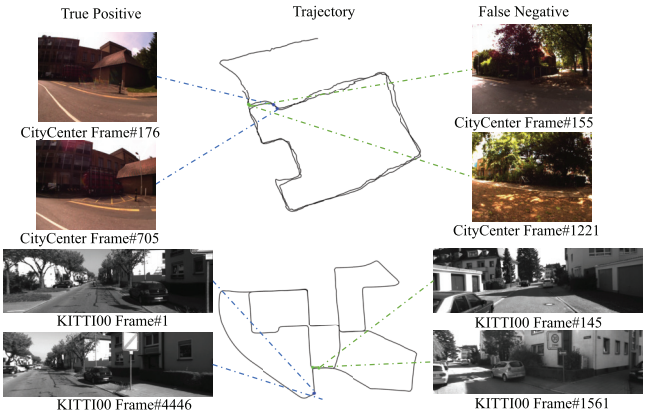

Fig. 8. Visual examples of LCD tasks using our GSLC, with the left images depicting correctly recognized loop closures and the right images showcasing instances of unrecognized loop closures.

retrieval. Specifically, the L2-norm distance of global features is used to select candidate frames for each query image. Subsequently, SIFT [6] is uniformly employed as the local feature to establish point-to-point matches between the query image and its candidate frames. Only when a sufficient number of matches (greater than a certain threshold) are retained, a loop closure event is identified.

For all loop closures ultimately identified, their correctness is determined based on Ground Truth (GT). In the context of precision and recall, an identification is considered a true inlier when it is situated within a proximity of 10 adjacent frames relative to the loop-closing samples from GT. The qualitative examples are exhibited in Fig. 8, demonstrating that our method exhibits limited matching performance for image pairs characterized by drastic variations in illumination and pronounced changes in viewing angles. The maximum recall rate as 100% precision is maintained as the quantitative

TABLE XI

THE QUANTITATIVE RESULTS OF LOOP CLOSURE DETECTION. THE MAXIMUM RECALL (%) AT 100% PRECISION OF TEN FEATURE MATCHING METHODS ON CC AND K00 SEQUENCES. **Bold** INDICATES THE BEST

	GMS [19]	LPM [20]	GLPM [58]	LOGO [22]	MCDM [23]	CRC [31]	OANet [26]	BLOB-DTM [38]	ConvMatch [18]	GSLC
CC	83.30	80.67	93.10	89.29	78.77	92.68	91.09	91.29	86.28	**93.60**
K00	94.29	90.71	89.85	94.42	93.91	27.79	87.65	93.65	89.85	**94.91**

TABLE XII

THE QUANTITATIVE RESULTS OF VISUAL LOCALIZATION, WHERE THE ACCURACY IS OBTAINED AT DIFFERENT THRESHOLDS, *i.e.*, (0.25M,2°), (0.5M,5°), AND (1.0M,10°). THE **bold** DENOTES THE BEST

Method	Day	Night
	(0.25m,2°)/(0.5m,5°)/(1.0m,10°)	
MNN	82.3/88.2/91.5	43.9/49.0/60.2
LFGC [25]	83.1/92.2/96.2	69.4/79.6/89.8
OANet [26]	83.3/92.5/96.6	71.4/80.6/89.8
CLNet [59]	83.3/92.4/**97.0**	71.4/80.6/**90.9**
ConvMatch [18]	83.5/92.7/96.8	**72.0**/80.9/90.6
GSLC	**83.7**/**92.8**/96.3	71.5/**81.0**/90.2

evaluation metric under various threshold values. The experimental results are presented in Table XI. Our GSLC exhibits the best performance due to its superior matching accuracy compared to other nine comparators. Therefore, GSLC significantly enhances the efficiency of loop-closure detection tasks, demonstrating its high utility in advanced visual tasks.

F. Visual Localization

To demonstrate the practicality of our GSLC, we evaluate its performance on visual localization tasks using the HLoc pipeline [74], which estimates the 6-DOF pose of a query image relative to a 3D model. The assessment focuses on challenging conditions such as large viewpoint changes and day-night illumination variations. Experiments were conducted on the Aachen Day-Night dataset [75], consisting of 4,328 reference images and 922 query images, including 824 captured during the day and 98 at night using mobile cameras.

We report the proportion of successfully localized queries based on standard distance and orientation thresholds [74], as shown in Table XII. SIFT [6] was used to extract up to 4,096 keypoints per image, which were matched using various feature matching methods. In addition to our proposed GSLC, competitive methods include classic MNN and representative learning-based methods. The SfM model, built from daytime images with known poses, was employed to register nighttime queries via 2D-2D matching within the COLMAP framework [76]. As a non-learning-based approach, GSLC exhibited competitive performance in visual localization tasks, showcasing stable results in daytime scenarios.

V. CONCLUSION

This paper proposes a novel feature matching method called *Grid-guided Sparse Laplacian Consensus*. Unlike previous global modeling based smooth constraints, a grid-based matching guidance strategy is designed to overcome the difficulties of independent motions. We also propose a novel motion coherence formulation based on graph Laplacian, where a new representation for the mapping function is derived. In the context of solving the constructed mixture model using the EM algorithm, seed correspondence initialization and sparse approximation ensure rapid and accurate convergence to the global optimum. The experiments demonstrate the superiority of our method over state-of-the-art methods, including deep-learning based methods. Our GSLC excels in terms of robustness to severe deformations, generalizability to various descriptors, suitability for handing independent motions, and practical applicability in high-level vision tasks.

Although our method effectively establishes a significant number of accurate correspondences, there remains room for improvement in terms of coverage areas, which may impact performance in geometric estimation tasks, particularly in challenging wide-baseline scenarios.

REFERENCES

[1] J. Ma, X. Jiang, A. Fan, J. Jiang, and J. Yan, "Image matching from handcrafted to deep features: A survey," *Int. J. Comput. Vis.*, vol. 129, no. 1, pp. 23–79, Jan. 2021.
[2] S. Jiang, W. Jiang, and B. Guo, "Leveraging vocabulary tree for simultaneous match pair selection and guided feature matching of UAV images," *ISPRS J. Photogramm. Remote Sens.*, vol. 187, pp. 273–293, May 2022.
[3] M. Brown and D. G. Lowe, "Automatic panoramic image stitching using invariant features," *Int. J. Comput. Vis.*, vol. 74, no. 1, pp. 59–73, Aug. 2007.
[4] K. Zhang, X. Jiang, and J. Ma, "Appearance-based loop closure detection via locality-driven accurate motion field learning," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 3, pp. 2350–2365, Mar. 2022.
[5] R. Mur-Artal, J. M. M. Montiel, and J. D. Tardos, "ORB-SLAM: A versatile and accurate monocular SLAM system," *IEEE Trans. Robot.*, vol. 31, no. 5, pp. 1147–1163, Oct. 2015.
[6] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, Nov. 2004.
[7] A. Mishchuk, D. Mishkin, F. Radenovic, and J. Matas, "Working hard to know your neighbor's margins: Local descriptor learning loss," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 4829–4840.
[8] D. DeTone, T. Malisiewicz, and A. Rabinovich, "SuperPoint: Self-supervised interest point detection and description," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2018, pp. 224–236.
[9] M. A. Fischler and R. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, no. 6, pp. 381–395, 1981.
[10] P. H. S. Torr and A. Zisserman, "MLESAC: A new robust estimator with application to estimating image geometry," *Comput. Vis. Image Understand.*, vol. 78, no. 1, pp. 138–156, Apr. 2000.
[11] O. Chum, J. Matas, and J. Kittler, "Locally optimized RANSAC," in *Pattern Recognition*. Berlin, Germany: Springer, Jan. 2003, pp. 236–243.
[12] O. Chum and J. Matas, "Matching with PROSAC-progressive sample consensus," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 1, Jun. 2005, pp. 220–226.

[13] D. Barath and J. Matas, "Graph-cut RANSAC," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 6733–6741.

[14] D. Barath, J. Matas, and J. Noskova, "MAGSAC: Marginalizing sample consensus," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 10197–10205.

[15] D. Baráth, J. Noskova, M. Ivashechkin, and J. Matas, "MAGSAC++, a fast, reliable and accurate robust estimator," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 1301–1309.

[16] M. Ivashechkin, D. Barath, and J. Matas, "VSAC: Efficient and accurate estimator for H and F," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 15223–15232.

[17] T. Wei, Y. Patel, A. Shekhovtsov, J. Matas, and D. Barath, "Generalized differentiable RANSAC," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2023, pp. 17649–17660.

[18] S. Zhang and J. Ma, "ConvMatch: Rethinking network design for two-view correspondence learning," in *Proc. AAAI Conf. Artif. Intell.*, 2023, pp. 1–12.

[19] J. Bian, W.-Y. Lin, Y. Matsushita, S.-K. Yeung, T.-D. Nguyen, and M.-M. Cheng, "GMS: Grid-based motion statistics for fast, ultra-robust feature correspondence," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2828–2837.

[20] J. Ma, J. Zhao, J. Jiang, H. Zhou, and X. Guo, "Locality preserving matching," *Int. J. Comput. Vis.*, vol. 127, no. 5, pp. 512–531, May 2019.

[21] X. Jiang, Y. Xia, X.-P. Zhang, and J. Ma, "Robust image matching via local graph structure consensus," *Pattern Recognit.*, vol. 126, Jun. 2022, Art. no. 108588.

[22] Y. Xia and J. Ma, "Locality-guided global-preserving optimization for robust feature matching," *IEEE Trans. Image Process.*, vol. 31, pp. 5093–5108, 2022.

[23] J. Ma, A. Fan, X. Jiang, and G. Xiao, "Feature matching via motion-consistency driven probabilistic graphical model," *Int. J. Comput. Vis.*, vol. 130, no. 9, pp. 2249–2264, Sep. 2022.

[24] S. Lee, J. Lim, and I. H. Suh, "Progressive feature matching: Incremental graph construction and optimization," *IEEE Trans. Image Process.*, vol. 29, pp. 6992–7005, 2020.

[25] K. M. Yi, E. Trulls, Y. Ono, V. Lepetit, M. Salzmann, and P. Fua, "Learning to find good correspondences," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 2666–2674.

[26] J. Zhang et al., "Learning two-view correspondences and geometry using order-aware network," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 5845–5854.

[27] Y. Liu, L. Liu, C. Lin, Z. Dong, and W. Wang, "Learnable motion coherence for correspondence pruning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2021, pp. 3237–3246.

[28] X. Li and Z. Hu, "Rejecting mismatches by correspondence function," *Int. J. Comput. Vis.*, vol. 89, no. 1, pp. 1–17, Aug. 2010.

[29] J. Ma, J. Zhao, J. Tian, A. L. Yuille, and Z. Tu, "Robust point matching via vector field consensus," *IEEE Trans. Image Process.*, vol. 23, pp. 1706–1721, 2014.

[30] W.-Y. Lin, M. Cheng, J. Lu, H. Yang, N. Minh, and P. H. S. Torr, "Bilateral functions for global motion modeling," in *Proc. Eur. Conf. Comput. Vis.*, Jan. 2014, pp. 341–356.

[31] A. Fan, X. Jiang, Y. Ma, X. Mei, and J. Ma, "Smoothness-driven consensus based on compact representation for robust feature matching," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 8, pp. 4460–4472, Aug. 2023.

[32] Y. Lu, J. Ma, L. Fang, X. Tian, and J. Jiang, "Robust and scalable Gaussian process regression and its applications," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 21950–21959.

[33] H. Bay, T. Tuytelaars, and L. Van Gool, "SURF: Speeded up robust features," in *Proc. Eur. Conf. Comput. Vis.*, 2006, pp. 404–417.

[34] P. F. Alcantarilla, A. Bartoli, and A. J. Davison, "KAZE features," in *Proc. 12th Eur. Conf. Comput. Vis. (ECCV)*, Oct. 2012, pp. 214–227.

[35] Y. Tian, X. Yu, B. Fan, F. Wu, H. Heijnen, and V. Balntas, "SOSNet: Second order similarity regularization for local descriptor learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 11016–11025.

[36] Z. Luo et al., "ASLFeat: Learning local features of accurate shape and localization," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 6589–6598.

[37] K. Lenc, J. Matas, and D. Mishkin, "A few things one should know about feature extraction, description and matching," in *Proc. Comput. Vis. Winter Workshop (CVWW)*, 2014, pp. 67–74.

[38] F. Bellavia, "SIFT matching by context exposed," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 2, pp. 2445–2457, Feb. 2023.

[39] O. Chum, T. Werner, and J. Matas, "Two-view geometry estimation unaffected by a dominant plane," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2005, pp. 772–779.

[40] R. Raguram, O. Chum, M. Pollefeys, J. Matas, and J.-M. Frahm, "USAC: A universal framework for random sample consensus," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 2022–2038, Aug. 2013.

[41] X. Jiang, J. Ma, J. Jiang, and X. Guo, "Robust feature matching using spatial clustering with heavy outliers," *IEEE Trans. Image Process.*, vol. 29, pp. 736–746, 2020.

[42] L. Cavalli, V. Larsson, M. R. Oswald, T. Sattler, and M. Pollefeys, "Handcrafted outlier detection revisited," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2020, pp. 770–787.

[43] Y. Xia, J. Jiang, Y. Lu, W. Liu, and J. Ma, "Robust feature matching via progressive smoothness consensus," *ISPRS J. Photogramm. Remote Sens.*, vol. 196, pp. 502–513, Feb. 2023.

[44] C. Zhao, Z. Cao, C. Li, X. Li, and J. Yang, "NM-Net: Mining reliable neighbors for robust feature correspondences," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 215–224.

[45] P.-E. Sarlin, D. DeTone, T. Malisiewicz, and A. Rabinovich, "SuperGlue: Learning feature matching with graph neural networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 4938–4947.

[46] H. Chen et al., "Learning to match features with seeded graph matching network," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 6281–6290.

[47] Y. Shi, J.-X. Cai, Y. Shavit, T.-J. Mu, W. Feng, and K. Zhang, "ClusterGNN: Cluster-based coarse-to-fine graph neural network for efficient feature matching," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 12517–12526.

[48] A. Ortega, P. Frossard, J. Kovačević, J. M. F. Moura, and P. Vandergheynst, "Graph signal processing: Overview, challenges, and applications," *Proc. IEEE*, vol. 106, no. 5, pp. 808–828, May 2018.

[49] F. Girosi, M. Jones, and T. Poggio, "Regularization theory and neural networks architectures," *Neural Comput.*, vol. 7, no. 2, pp. 219–269, Mar. 1995.

[50] A. N. Tikhonov, "On the solution of ill-posed problems and the method of regularization," *Dokl. Akad. Nauk*, vol. 151, no. 3, pp. 501–504, 1963.

[51] N. Aronszajn, "Theory of reproducing kernels," *Trans. Amer. Math. Soc.*, vol. 68, no. 3, pp. 337–404, 1950.

[52] C. A. Micchelli and M. Pontil, "On learning vector-valued functions," *Neural Comput.*, vol. 17, no. 1, pp. 177–204, Jan. 2005.

[53] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Royal Stat. Soc. B, Methodol.*, vol. 39, no. 1, pp. 1–22, 1977.

[54] C. Williams and M. Seeger, "Using the Nyström method to speed up kernel machines," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 13, 2000, pp. 661–667.

[55] J. Ma, J. Zhao, J. Tian, X. Bai, and Z. Tu, "Regularized vector field learning with sparse approximation for mismatch removal," *Pattern Recognit.*, vol. 46, no. 12, pp. 3519–3532, Dec. 2013.

[56] A. Vedaldi and B. Fulkerson, "VLFeat: An open and portable library of computer vision algorithms," in *Proc. 18th ACM Int. Conf. Multimedia*, 2010, pp. 1469–1472.

[57] J. Heinly, E. Dunn, and J.-M. Frahm, "Comparative evaluation of binary features," in *Proc. Eur. Conf. Comput. Vis.*, 2012, pp. 759–773.

[58] J. Ma, J. Jiang, H. Zhou, J. Zhao, and X. Guo, "Guided locality preserving feature matching for remote sensing image registration," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 8, pp. 4435–4447, Aug. 2018.

[59] C. Zhao, Y. Ge, F. Zhu, R. Zhao, H. Li, and M. Salzmann, "Progressive correspondence pruning by consensus learning," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, Oct. 2021, pp. 6464–6473.

[60] K. Mikolajczyk et al., "A comparison of affine region detectors," *Int. J. Comput. Vis.*, vol. 65, nos. 1–2, pp. 43–72, Oct. 2005.

[61] K. Cordes, B. Rosenhahn, and J. Ostermann, "High-resolution feature evaluation benchmark," in *Proc. Int. Conf. Comput. Anal. Images Patterns*, Jan. 2013, pp. 327–334.

[62] H. S. Wong, T.-J. Chin, J. Yu, and D. Suter, "Dynamic and hierarchical multi-structure geometric model fitting," in *Proc. IEEE Int. Conf. Comput. Vis.*, Nov. 2011, pp. 1044–1051.

[63] V. Balntas, K. Lenc, A. Vedaldi, and K. Mikolajczyk, "HPatches: A benchmark and evaluation of handcrafted and learned local descriptors," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 5173–5182.

[64] J. Sturm, N. Engelhard, F. Endres, W. Burgard, and D. Cremers, "A benchmark for the evaluation of RGB-D SLAM systems," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Oct. 2012, pp. 573–580.
[65] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "Vision meets robotics: The KITTI dataset," *Int. J. Robot. Res.*, vol. 32, no. 11, pp. 1231–1237, 2013.
[66] A. Knapitsch, J. Park, Q.-Y. Zhou, and V. Koltun, "Tanks and temples: Benchmarking large-scale scene reconstruction," *ACM Trans. Graph.*, vol. 36, no. 4, pp. 1–13, 2017.
[67] K. Wilson and N. Snavely, "Robust global translations with 1DSfM," in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 61–75.
[68] J.-W. Bian et al., "An evaluation of feature matchers for fundamental matrix estimation," in *Proc. Brit. Mach. Vis. Conf.*, Jan. 2019, pp. 1–14.
[69] G. Potje, F. Cadar, A. Araujo, R. Martins, and E. R. Nascimento, "XFeat: Accelerated features for lightweight image matching," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2024, pp. 2682–2691.
[70] G. Donato and S. Belongie, "Approximate thin plate spline mappings," in *Proc. Eur. Conf. Comput. Vis.*, 2002, pp. 21–31.
[71] L. Liang et al., "Image registration using two-layer cascade reciprocal pipeline and context-aware dissimilarity measure," *Neurocomputing*, vol. 371, pp. 1–14, Jan. 2020.
[72] M. Cummins and P. Newman, "FAB-MAP: Probabilistic localization and mapping in the space of appearance," *Int. J. Robot. Res.*, vol. 27, no. 6, pp. 647–665, 2008.
[73] B. Cao, A. Araujo, and J. Sim, "Unifying deep local and global features for image search," in *Proc. Eur. Conf. Comput. Vis.*, Aug. 2020, pp. 726–743.
[74] P.-E. Sarlin, C. Cadena, R. Siegwart, and M. Dymczyk, "From coarse to fine: Robust hierarchical localization at large scale," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 12716–12725.
[75] T. Sattler et al., "Benchmarking 6DOF outdoor visual localization in changing conditions," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 8601–8610.
[76] J. L. Schonberger and J.-M. Frahm, "Structure-from-motion revisited," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2016, pp. 4104–4113.

Yifan Xia received the B.E. degree in information and communication engineering from Wuhan University, Wuhan, China, in 2021, where he is currently pursuing the Ph.D. degree with the Electronic Information School. His current research interests include computer vision and image processing.

Jiayi Ma (Senior Member, IEEE) received the B.S. degree in information and computing science and the Ph.D. degree in control science and engineering from the Huazhong University of Science and Technology, Wuhan, China, in 2008 and 2014, respectively. He is currently a Professor with the Electronic Information School, Wuhan University, Wuhan. He has co-authored more than 400 refereed journal and conference papers, including *Cell*, IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, and IJCV. He was a recipient of the Information Fusion Best Paper Award in 2024 and the Hsue-Shen Tsien Paper Award in 2023. He is an Area Editor of *Information Fusion* and an Associate Editor of IEEE/CAA JOURNAL OF AUTOMATICA SINICA, *Neurocomputing*, *Geo-spatial Information Science*, and *Image and Vision Computing*.