

Received February 1, 2020, accepted February 15, 2020, date of publication February 21, 2020, date of current version March 2, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2975670

# A GMS-Guided Approach for 2D Feature Correspondence Selection

QUNFANG TANG<sup>1,2</sup>, JIE YANG<sup>1</sup>, WENJING JIA<sup>3</sup>, XIANGJIAN HE<sup>3</sup>, (Senior Member, IEEE),  
QINGNIAN ZHANG<sup>4</sup>, AND HAIBO LIU<sup>2</sup>

<sup>1</sup>Hubei Key Laboratory of Broadband Wireless Communication and Sensor Networks, Wuhan University of Technology, Wuhan 430070, China

<sup>2</sup>School of Electrical Information Engineering, Hunan Institute of Technology, Hengyang 421002, China

<sup>3</sup>Global Big Data Technologies Centre (GBDTC), University of Technology Sydney, Sydney, NSW 2007, Australia

<sup>4</sup>School of Transportation, Wuhan University of Technology, Wuhan 430070, China

Corresponding author: Jie Yang (jiejyang509@163.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 51879211, in part by the Hunan Provincial Natural Science Foundation of China under Grant 2017JJ3053, and in part by the Hunan Provincial Science Research Project of China under Grant 17A051.

**ABSTRACT** Feature correspondence selection, which aims to seek as many true matches (*i.e.*, inliers) as possible from a given putative set while minimizing false matches (*i.e.*, outliers), is crucial to many feature-matching based tasks in computer vision. It remains a challenging problem how to deal with putative sets with low inlier ratios. To address this problem, in this paper, we propose a novel correspondence selection strategy, which is guided by Grid-based Motion Statistics (GMS). We first adopt the GMS to generate a small correspondence set with a high inlier ratio. Then, an accurate geometric model is built using the above correspondence set. Finally, the built geometric model is used to filter the given putative correspondence set to obtain true correspondences. The experimental results on benchmark datasets demonstrate that our proposed approach outperforms the state-of-the-art approaches for putative sets with various inlier ratios, especially for cases with low inlier ratios.

**INDEX TERMS** Correspondence selection, grid-based motion statistics (GMS), geometric model, outlier.

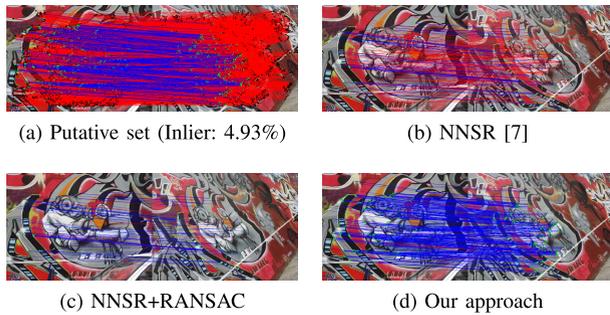
## I. INTRODUCTION

Establishing reliable feature correspondences (*i.e.*, matches) between two image feature sets is not only a prerequisite but also a critical step in many computer vision tasks, such as Simultaneous Localization and Mapping (SLAM) [1], object tracking [2], stereo matching [3], image stitching [4], and Structure-From-Motion (SFM) [5]. Generally, finding good feature correspondences is a two-stage strategy, typically carried out in the context of the feature matching. In the first stage, a set of putative correspondence (*i.e.*, putative set) is constructed by using similarity constraints of local image descriptors [6] (*e.g.*, SIFT [7], ORB [8]), where ORB is widely used in real-time applications for its fast speed. However, the putative set inevitably contains a large number of false matches (*i.e.*, outliers) due to ambiguities of the local descriptors, and the large variances in illumination, perspective and clarity. This inconsistency makes massive high-level vision tasks impossible. Therefore, correspondence selection

is widely adopted during the second stage, *e.g.*, [9]–[12], to eliminate the outliers from the constructed putative set. The main goal of the correspondence selection is to seek as many true matches (*i.e.*, inliers) as possible from a given putative set while minimizing false matches.

During the past few decades, although a variety of methods have been proposed to deal with the mismatch removal problem, it remains a challenging problem when dealing with putative sets with low inlier ratios. Almost all of the existing methods perform well in putative sets with higher inlier ratios but poorly in putative sets with lower inlier ratios. To address this problem, the classic pipelines aim to boost the inlier ratio of the putative set using the Nearest Neighbor Similarity Ratio (NNSR) [7] to prune poor-quality correspondences before implementing the correspondence selection approach. However, NNSR is not generally applicable to other descriptors than SIFT [13]. As shown in Fig. 1(b), when the putative set is generated from ORB features, only a small portion of true matches as well as a large portion of false matches are preserved after using NNSR. Therefore, as shown in Fig. 1(c), we can get only a small amount of

The associate editor coordinating the review of this manuscript and approving it for publication was Huazhu Fu.



**FIGURE 1.** The correspondence selection results of the classic approaches and our proposed approach. (a) The putative set generated from the ORB features; (b) The result of NNSR [7]; (c) The result of NNSR+RANSAC; (d) The result of our approach. We draw them in blue if they conform to the ground-truth, and in red otherwise.

true matches when we further filter putative set using the correspondence selection method. This is problematic when the image pairs themselves contain very limited true matches, such as matching low-overlapping images in image stitching. Therefore, it is necessary to develop an outlier removal method that can handle putative sets with low inlier ratios and also be applicable to descriptors other than SIFT.

To solve the above-mentioned issue, in this paper, we develop a new strategy that can preserve most of the existing true matches when the putative set contains a large number of outliers or few inliers. Inspired by Grid-based Motion Statistics (GMS) [14], which is a simple, efficient (as efficient as NNSR) and ultra-robust correspondence selection method, our strategy is referred to as a GMS-guided (GMS-G) approach. As we all know, for an outlier removal method based on geometrical constraint, the more accurate the geometric model is, the better the correspondence selection performance has. However, an accurate geometric model cannot be constructed from a putative set with a high outlier ratio. Furthermore, if we want to retain the maximum number of inliers, the putative set selected for the correspondence selection method should include as many true matches as possible. Thus, in our approach, instead of building the geometric model directly from the putative set, we propose to construct the geometric model from a smaller but more reliable correspondence set. In order to do so, we propose to filter the initial putative set using the idea of GMS while retaining all possible true matches to yield the largest number of inliers. As demonstrated in comparative experiments, our strategy can significantly boost true matches and remove false matches, and effectively handle putative sets with extremely low inlier ratios.

The contributions made in this paper are in the following three aspects. (i) We propose a GMS-guided correspondence selection framework that can greatly boost the true matches while rejecting false matches, improving both precision and recall; (ii) Our method is especially advantageous when dealing with putative correspondence sets involving a large number of outliers or a small number of inliers; (iii) Our approach is highly efficient and can be applied to real-time applications.

## II. RELATED WORK

Existing 2D feature correspondence selection works can be generally classified into three major categories, *i.e.*, parametric approaches, non-parametric approaches and learning based approaches.

### A. PARAMETRIC APPROACHES

The parametric model based methods try to seek consistent correspondences grounded on parametric geometric models. The classic Random Sample Consensus algorithm (RANSAC) [15] and its variants (*e.g.*, PROSAC [16], USAC [17], GC-RANSAC [18], MAGSAC [19]) estimate a geometric model (*e.g.*, a homographic matrix or fundamental matrix) to distinguish between inliers and outliers. They abide by a hypothesize-and-verify strategy, aiming to find the smallest consistent inlier set to fit a given geometric model and estimate a pre-defined transformation by resampling randomly [20]. Correspondences consistent with the geometric model are considered as inliers; otherwise outliers.

### B. NON-PARAMETRIC APPROACHES

The non-parametric methods are independent of parametric model assumptions. Some non-parametric model-based methods use the feature similarity constraint or geometric constraint to search the corresponding inliers, such as the Nearest Neighbor Similarity Ratio (NNSR) [7] and Locality Preserving Matching (LPM) [21], [22]. The NNSR [7], using the feature similarity constraint, assigned a penalty equal to the ratio of the closest to the second-closest feature distance to each correspondence, where correspondences with low ratios were treated as inliers. The LPM [21], [22] introduced a mathematical model to select inlier correspondences having similar local neighborhood structures. There are also some constraint-independent non-parametric methods, such as Vector Field Consensus (VFC) [23], [24], Manifold Regularization-based Robust Point Matching (MR-RPM) [25], Robust Feature Matching based on Spatial Clustering Algorithm with Noisy samples (RFM-SCAN) [26] and Grid-based Motion Statistics (GMS) [14]. The VFC assumed that the noise around inliers and outliers fell in different distributions and estimated the probability of inliers by maximum likelihood estimation for parameters in the mixture probabilistic model. The MR-RPM approach [25] enforced the motion field to be smooth under manifold regularization and conquered the matching problem from a robust motion field interpolation perspective. It has shown a promising performance on addressing the deformable matching problem. The RFM-SCAN approach [26] can address image pairs undergoing any transformation models. This method cast the matching problem into spatial clustering with outliers, and customized the classic density-based spatial clustering of applications with noise (DBSCAN) [27] to automatically determine the number of clusters and eliminate the outliers simultaneously. The GMS approach [14] rejected false matches by counting the

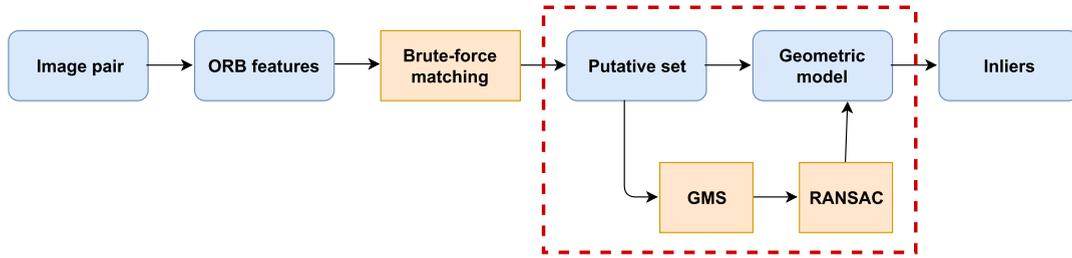


FIGURE 2. GMS-guided correspondence selection strategy.

number of matches in small neighborhoods and has achieved real-time performance with an efficient grid-based score estimator. Our work in this paper is inspired by this approach.

### C. LEARNING BASED APPROACHES

Some new approaches based on deep learning technique have also been well studied and achieved great success in recent years, such as the LFGC [28], LMR [29], and NM-Net [30]. The LFGC [28] firstly proposed a learning approach to find good correspondences by finding geometrically consistent correspondences with a deep network. The LMR [29] casted the mismatch removal into a two-class classification problem, and learned a general classifier to determine the correctness of an arbitrary putative match. The NM-Net [30] proposed a deep classification network that fully mined compatibility-specific locality for correspondence selection.

The approach proposed in this paper belongs to the parametric approaches. Aiming to address the challenges of putative sets with low inlier ratios, inspired by the GMS approach [14] we first generate a rather accurate geometric model using those reliable correspondence set of high inlier ratios. Our approach can greatly boost the true matches while avoiding false matches, improving both precision and recall.

## III. METHOD

### A. THE FRAMEWORK OF GMS-GUIDED STRATEGY

The framework of our GMS-guided strategy for correspondence selection is shown in Fig. 2, where the modules enclosed in the red dash box indicate our contribution.

In this strategy, using the GMS algorithm, an accurate and small correspondence set can first be obtained by filtering the putative set. Then, with the RANSAC estimator, these accurate correspondences are used to construct an accurate geometric model. Finally, this accurate geometric model is adopted to filter the initial putative set to obtain inliers, where almost all of the true matches are retained with very few mismatches.

### B. THE GMS ALGORITHM

Now, we briefly review the popular feature matching method GMS [14] used in our GMS-guided strategy.

The GMS algorithm proves that the number of features also contributes to the quality of correspondences besides feature

descriptiveness. It shows that the quantity of correspondences in a small neighborhood around a true match is larger than that around a false match under the smooth motion.

Suppose that there is an image pair  $\{I_a, I_b\}$  having  $\{N, M\}$  features respectively.  $\mathcal{C} = \{\mathbf{c}_i\}_{i=1}^N$  is the set of all nearest neighbor feature matches from  $I_a$  to  $I_b$ , and  $\mathcal{C}$  has cardinality  $|\mathcal{C}| = N$ .

In the GMS algorithm, first  $I_a$  and  $I_b$  are divided into  $20 \times 20$  non-overlapping cells (grids). For each cell in  $I_a$ , the cell containing the maximum amount of correspondences is grouped in  $I_b$ .  $S_{ij}$  is a measure of neighborhood support in cell-pair  $(i, j)$  and their small neighborhoods (eight cell-pairs), which can be estimated as:

$$S_{ij} = \sum_{k=1}^{k=9} |C_{ik^k j^k}|, \quad (1)$$

where  $|C_{ik^k j^k}|$  is the amount of correspondences in the cell-pair  $(i^k, j^k)$ .

All correspondences in  $(i, j)$  are considered as inliers if

$$cell - pair \{i, j\} \in \begin{cases} \mathcal{T}, & \text{if } S_{ij} > \tau_i = \alpha \sqrt{n_i} \\ \mathcal{F}, & \text{otherwise} \end{cases} \quad (2)$$

where  $\mathcal{T}$  indicates true correspondences (inliers) and  $\mathcal{F}$  indicates false correspondences (outliers),  $\tau_i$  is a threshold approximated by  $\alpha \sqrt{n_i}$ , and  $\alpha$  is a given parameter and  $n_i$  is the average (of the nine cell-pairs) amount of correspondences.

### C. THE GEOMETRIC MODEL ESTIMATION

In this section, we provide the details of the geometric model estimation.

There are several geometric models in two-view image geometry theory [31], such as affine, homography and epipolar geometry. For 2D correspondence selection, we are interested in estimating the homography, which suits more general scenes than other geometric models.

Assume that images are obtained by a perspective pinhole camera, and points are presented by Cartesian coordinates. Suppose that  $\mathbf{x} = (u; v)^T$  and  $\mathbf{x}' = (u'; v')^T$  are two points of a correspondence. A planar projective transformation or

homography that maps  $\mathbf{x}$  to  $\mathbf{x}'$  can be expressed by:

$$\begin{bmatrix} u' \\ v' \\ 1 \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \text{ or } \begin{bmatrix} \mathbf{x}' \\ 1 \end{bmatrix} = H [\mathbf{x}1], \quad (3)$$

where  $H$  is a non-singular  $3 \times 3$  matrix.

Since the transformation is defined up to a scaling factor, it can be normalized by scaling  $h_{33} = 1$ , and as such  $H$  can be parameterized by eight parameters. Typically, homography is estimated between images by finding feature correspondences in those images, and can be computed from four correspondences.

Our goal of building an accurate geometric model to distinguish between outliers and inliers is now converted to estimating the homography matrix  $H$  from a putative set. Obviously, if we can successfully estimate the homography matrix  $H$ , the outliers can be easily identified. Furthermore, if we can calculate  $H$  from a putative set with a high inlier ratio, we can get a more accurate geometric model, which can help us to better remove outliers.

#### D. THE GMS-GUIDED CORRESPONDENCE SELECTION

In this section, we explain the process of our GMS-guided correspondence selection strategy in steps.

##### 1) STEP 1: GENERATING A PUTATIVE SET

As is well-known, the ORB feature outperforms others in real-time applications due to its excellent trade-off between robustness and efficiency. In our GMS-guided correspondence selection strategy, ORB features are extracted from a given image pair  $\{I_a, I_b\}$ . Moreover, the brute-force matching tries to match all descriptors, it can not only always find the best correspondence, but also be accelerated significantly with modern GPU hardware. Therefore, the putative set  $\mathcal{C}$  introduced in Part B is built with brute-force Hamming distance comparisons (Nearest-neighbor matching), which can be described as:

$$\mathcal{C} = \{\mathbf{c}_i\}_{i=1}^N = \{(\mathbf{x}_i, \mathbf{x}'_i)\}_{i=1}^N, \quad (4)$$

where  $\mathbf{x}_i$  and  $\mathbf{x}'_i$  are two feature points of correspondence of  $\mathbf{c}_i$ .

Note that, this putative set usually contains a large number of matches, covering possibly all true matches. Meanwhile, with a low inlier ratio, it also has many false matches due to the limited discrimination ability of the binary ORB features and external obstructions such as repetitive patterns and noise.

##### 2) STEP 2: IMPROVING THE ACCURACY OF THE PUTATIVE SET BY GMS ALGORITHM

As aforementioned, our goal is to estimate an accurate geometric model to distinguish between outliers and inliers. Nevertheless, the putative set  $\mathcal{C}$  contains a large number of outliers.

As the most commonly used robust estimation method for homographies, the RANSAC algorithm [15] has been widely adopted to build geometric models. However, building

a geometric model directly from a putative set with high outlier ratios can result in inaccurate geometric models even failure using RANSAC. Therefore, in our GMS-guided strategy, we propose to use the process of the GMS (detailed in Part B) to boost the inlier ratio of the putative set  $\mathcal{C}$  and obtain a reliable correspondence set  $\mathcal{C}_m$  as:

$$\mathcal{C}_m = \{(\mathbf{x}_i, \mathbf{x}'_i)\}_{i=1}^M, \quad (5)$$

where  $M$  is the maximum number of the putative set  $\mathcal{C}_m$ , and indicates relatively more inliers.

In order to accelerate the convergence of RANSAC algorithm and build a more accurate geometric model in the next step, we rank  $\mathcal{C}_m$  by similarity and select  $L$  top-ranking correspondences from  $M$  (supposing that  $L$  is not larger than  $M$ ) to obtain a new set as:

$$\mathcal{C}_s = \{(\mathbf{x}_i, \mathbf{x}'_i)\}_{i=1}^L. \quad (6)$$

Here,  $L$  is empirically set to 500 in this paper and  $\mathcal{C}_s$  is a smaller and more accurate correspondence set.

##### 3) STEP 3: ESTIMATING THE GEOMETRIC MODEL

In this step, we use the RANSAC algorithm [15] to estimate an accurate geometric model (*i.e.*, the homography matrix  $H$ ).

The RANSAC algorithm firstly randomly samples several correspondences (at least 4) from the putative set  $\mathcal{C}'$  (if the number of  $\mathcal{C}_m > 500$ ,  $\mathcal{C}' = \mathcal{C}_s$ ; otherwise,  $\mathcal{C}' = \mathcal{C}_m$ ) and generates the model hypothesis  $H_i$  for those samples at the  $i$ -th iteration. Then, the hypothesis  $H_i$  is verified via the following object function:

$$O_i = \sum_{\mathbf{c} \in \mathcal{C}'} h_i(\mathbf{c}), \quad (7)$$

where  $h(\cdot)$  is a binary function defined by:

$$h_i(\mathbf{c}) = \begin{cases} 1, & \text{if } \|\mathbf{x}'_i - \mathbf{y}_i\|_2 \leq t_{\text{ransac}} \\ 0, & \text{otherwise,} \end{cases} \quad (8)$$

where  $\mathbf{y}_i = \rho \left( H_i \begin{bmatrix} \mathbf{x}_i \\ 1 \end{bmatrix} \right)$  is a reprojected feature point of  $\mathbf{x}_i$ , and  $t_{\text{ransac}}$  is the reprojection error with a default value of 3 pixels. After  $n_{\text{ransac}}$  iterations, the model with the maximum object function is selected as the final model  $H$ .

##### 4) STEP 4: REMOVING OUTLIERS USING THE GEOMETRIC MODEL

Finally, the established geometric model  $H$  is used to filter the putative set  $\mathcal{C}$  to remove the outliers.

For each correspondence  $(\mathbf{x}_i, \mathbf{x}'_i)$  in the putative set  $\mathcal{C}$ , the Euclidean distance, denoted as  $\mathbf{d}(\mathbf{x}'_i, \mathbf{y}'_i)$ , between points  $\mathbf{y}'_i$  and  $\mathbf{x}'_i$  is calculated. Specifically, if the distance is less than a pre-defined inlier threshold  $t$ , *i.e.*,

$$\mathbf{d}(\mathbf{x}'_i, \mathbf{y}'_i) < t, \quad (9)$$

the correspondence is identified as an inlier, where  $\mathbf{y}'_i = \rho \left( H \begin{bmatrix} \mathbf{x}_i \\ 1 \end{bmatrix} \right)$  is a reprojected feature point of  $\mathbf{x}_i$  under the

geometric model  $H$ , and  $\mathbf{d}(\cdot, \cdot)$  is the Euclidian distance between the two points, and the inlier threshold  $t$  is set to 2.5 following the practice of [32].

It should be noted that, in this step, we filter the putative set  $\mathcal{C}$  instead of  $\mathcal{C}_m$  to yield inliers, which is the key difference between our approach and the GMS+RANSAC approach. Because the reliable correspondence set  $\mathcal{C}_m$  is generated by the GMS algorithm, which selects true matches by simply computing the number of neighborhood correspondences, it is inevitable to remove true isolate matches with no neighborhood correspondences supporting or true matches with rare neighborhood correspondences supporting. Hence, a portion, and even a large portion, of the true matches in  $\mathcal{C}_m$  will be removed if the putative set distributes sparsely. So, if we filter  $\mathcal{C}_m$  to produce inliers, we will not obtain an inlier set including the maximum number of true matches. Instead,  $\mathcal{C}$  is generated by brute-force matching, containing almost all of the true matches. Therefore, filtering  $\mathcal{C}$  can significantly boost the true matches by enlarging the putative set and benefit those matching problems where the image pairs involve very few true matches.

---

#### Algorithm 1 GMS-Guided Algorithm

---

**Input:** An image pair  $\{I_a, I_b\}$

- 1: Detect ORB feature points;
- 2: Establish a putative set  $\mathcal{C} = \{(\mathbf{x}_i, \mathbf{x}'_i)\}_{i=1}^N$  using brute-force matching;
- 3: Obtain a reliable correspondence set  $\mathcal{C}_m = \{(\mathbf{x}_i, \mathbf{x}'_i)\}_{i=1}^M$  using GMS to filter putative set  $\mathcal{C}$ ;
- 4: **if** The numbers of the reliable correspondence set  $\mathcal{C}_m > L$  **then**
- 5:     Select the first  $M$  correspondences with the highest similarity combining  $\mathcal{C}_s = \{(\mathbf{x}_i, \mathbf{x}'_i)\}_{i=1}^L$  from  $\mathcal{C}_m$ ;
- 6:     Calculate a Homography matrix  $H$  using  $\mathcal{C}_s$ ;
- 7: **else**
- 8:     Calculate the Homography matrix  $H$  using  $\mathcal{C}_m$ ;
- 9: **end if**
- 10: Obtain the geometric model matrix  $H$ ;
- 11: Filter  $\mathcal{C}$  using  $H$ ;
- 12: **if** The Euclidian distance  $\mathbf{d}(\mathbf{x}'_i, \mathbf{y}_i) < 2.5$  **then**
- 13:     Correspondence  $(\mathbf{x}_i, \mathbf{x}'_i)$  is identified as an inlier.
- 14: **end if**

**Output:** Inlier set

---

We summarize our GMS-guided strategy in Algorithm 1. Because the geometric model constructed by this procedure is very accurate, our GMS-guided method can retain the most correct matches and contains few mismatches.

## IV. EXPERIMENTS

In this section, we evaluate the performance of our proposed GMS-guided approach and compare it with the state-of-art approaches on benchmark datasets.

## A. EXPERIMENT SETUP

### 1) IMPLEMENTATION

In our experiments, we use the open-source toolbox OpenCV 3.4 to extract ORB features, and brute-force matching is adopted to generate the putative set uniformly. All experiments are conducted on a 2.2GHZ Intel Core i7 CPU with a 16GB memory laptop.

We compare our algorithm with the state-of-the-art approaches including RANSAC [15], GMS [14], GMS+RANSAC (referred to as ‘GMS-R’), VFC [23], LFGC [28], LMR [29], and LPM [21]. The GMS+RANSAC method adopts GMS to obtain a putative set, and then uses RANSAC algorithm to remove outliers from the obtained putative set, which is listed for comparison to demonstrate the superiority of our strategy.

In the RANSAC algorithm, the number of  $n_{\text{ransac}}$  is set to 10,000 to get a good balance between performance and speed. In order to optimize experimental results, the scale and rotation variables in the GMS algorithm are both set to True. All other parameters not mentioned here are either default values in OpenCV functions or as described in the original publications, and are consistent throughout all experiments.

### 2) DATASETS

In this paper, three benchmark datasets, *i.e.*, the VGG dataset [33], the Heinly dataset [32] and the Symbench dataset [34], are adopted for conducting comparative experiments.

The VGG dataset [33] is a hybrid dataset and contains 48 images of eight scenes. These eight scenes cover a range of special interferences such as blur (‘bike’, ‘tree’), viewpoint change (‘graffiti’, ‘wall’), zoom and rotation (‘bark’, ‘boat’), lighting change (‘leuven’), and JPEG compression (‘ubc’). Each scene consists of six images, and the first image in each scene is a reference image to other images. Therefore, the generality to different conditions and the robustness of specific nuisances can be reflected from experimental results conducted on this dataset.

The Heinly dataset [32] comprises of 40 images with dense or sparse viewpoint change, illumination, pure large-scale zoom or rotation. So, it can be used to evaluate the performance under the condition of a geometrical structure deformation (pure zoom or rotation).

The Symbench dataset [34] is composed of 46 image pairs, and each pair includes the same object with lighting changes or different rendering styles. It is meant that the dataset causes image quality variations and gives rise to potential errors in the putative set. The performance in the context of image quality variation can be specifically evaluated. Moreover, all three datasets provide ground-truth.

### 3) EVALUATION CRITERIA

Same as in [35], Precision, Recall and F-measure are used to measure the performance of the evaluated algorithms. In the following formulas, the putative set, the ground-truth of the

**TABLE 1.** The Inlier Ratio (IR) of the putative set and the comparison of performance, *i.e.*, Precision (P), Recall (R) and F-measure (F), obtained with six state-of-the-art correspondence selection methods and our GMS-guided approach on the seven typical image pairs selected from the benchmark datasets shown in Fig. 3. Results highlighted in red and boldface are the best.

Image	IR (%)	Evaluation	VFC [23]	RANSAC [15]	LPM [21]	LMR [29]	GMS [14]	GMS-R	Our GMS-G
Archeasy	1.38	P (%)	0	0	49.44	37.10	73.33	<b>100</b>	<b>100</b>
		R (%)	0	0	31.88	59.42	7.97	7.97	<b>100</b>
		F (%)	-	-	38.77	45.68	14.38	14.77	<b>100</b>
Daynight	6.15	P (%)	0	88.66	75.82	71.90	55	92.08	<b>99.19</b>
		R (%)	0	35.61	56.59	63.25	90.24	75.61	<b>99.02</b>
		F (%)	-	50.81	64.80	67.30	68.35	83.04	<b>99.10</b>
Graffiti	4.93	P (%)	40.36	69.25	37.21	47.40	38.12	87.14	<b>93.13</b>
		R (%)	<b>98.65</b>	43.75	38.34	72.30	60.14	51.52	93.92
		F (%)	57.28	53.62	37.77	57.26	46.66	64.76	<b>93.52</b>
Venice	6.65	P (%)	0	73.77	36.34	45.74	50.66	84.89	<b>92.03</b>
		R (%)	0	33.83	43.98	76.69	86.47	71.80	<b>95.49</b>
		F (%)	-	46.39	39.80	57.30	63.89	77.80	<b>93.73</b>
Boat	16.27	P (%)	81.35	92.46	82.73	86.93	66.59	92.89	<b>95.12</b>
		R (%)	<b>99.45</b>	92.75	87.46	95.27	72.16	59.43	95.94
		F (%)	89.49	92.61	85.03	90.91	69.26	72.49	<b>95.53</b>
Townsquare	1.12	P (%)	0	0	5.92	23.21	73.33	84.85	<b>89.25</b>
		R (%)	0	0	9.09	47.27	60.00	50.91	<b>75.45</b>
		F (%)	-	-	7.17	31.14	66.00	63.64	<b>81.77</b>
Metz	1.20	P (%)	6.14	0	10.22	12.84	13.01	66.09	<b>82.03</b>
		R (%)	28.33	0	15.83	31.67	74.17	63.33	<b>87.50</b>
		F (%)	10.04	-	12.42	18.27	22.14	64.68	<b>84.68</b>
Average	5.39	P (%)	18.26	46.31	42.53	46.45	52.86	86.85	<b>92.96</b>
		R (%)	32.34	29.42	40.45	63.7	64.45	54.37	<b>92.47</b>
		F (%)	22.4	34.78	40.82	52.55	50.10	63.03	<b>92.62</b>

putative set and the selected correct correspondence set are represented as  $C_{inlier}$ ,  $C_{inlier}^{GT}$ , and  $C_{inlier}^{correct}$  respectively. Then, the evaluation metrics, *i.e.*, Precision, Recall and F-measure, are defined as:

$$\text{Precision} = \frac{|C_{inlier}^{correct}|}{|C_{inlier}|}, \tag{10}$$

$$\text{Recall} = \frac{|C_{inlier}^{correct}|}{|C_{inlier}^{GT}|}, \tag{11}$$

and

$$\text{F-measure} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}, \tag{12}$$

where  $|\cdot|$  denotes the cardinality of a set.

A correspondence  $(\mathbf{x}_i, \mathbf{x}'_i)$  in putative set belongs to  $C_{inlier}^{GT}$  if

$$\mathbf{d}(\mathbf{x}'_i, \mathbf{x}_i^{gt}) < t, \tag{13}$$

where  $\mathbf{x}_i^{gt} = \rho \left( H_{gt} \begin{bmatrix} \mathbf{x}_i \\ 1 \end{bmatrix} \right)$  is a reprojected feature point of  $\mathbf{x}_i$ , and  $H_{gt}$  is the ground-truth homography matrix.

## B. EXPERIMENTAL RESULTS

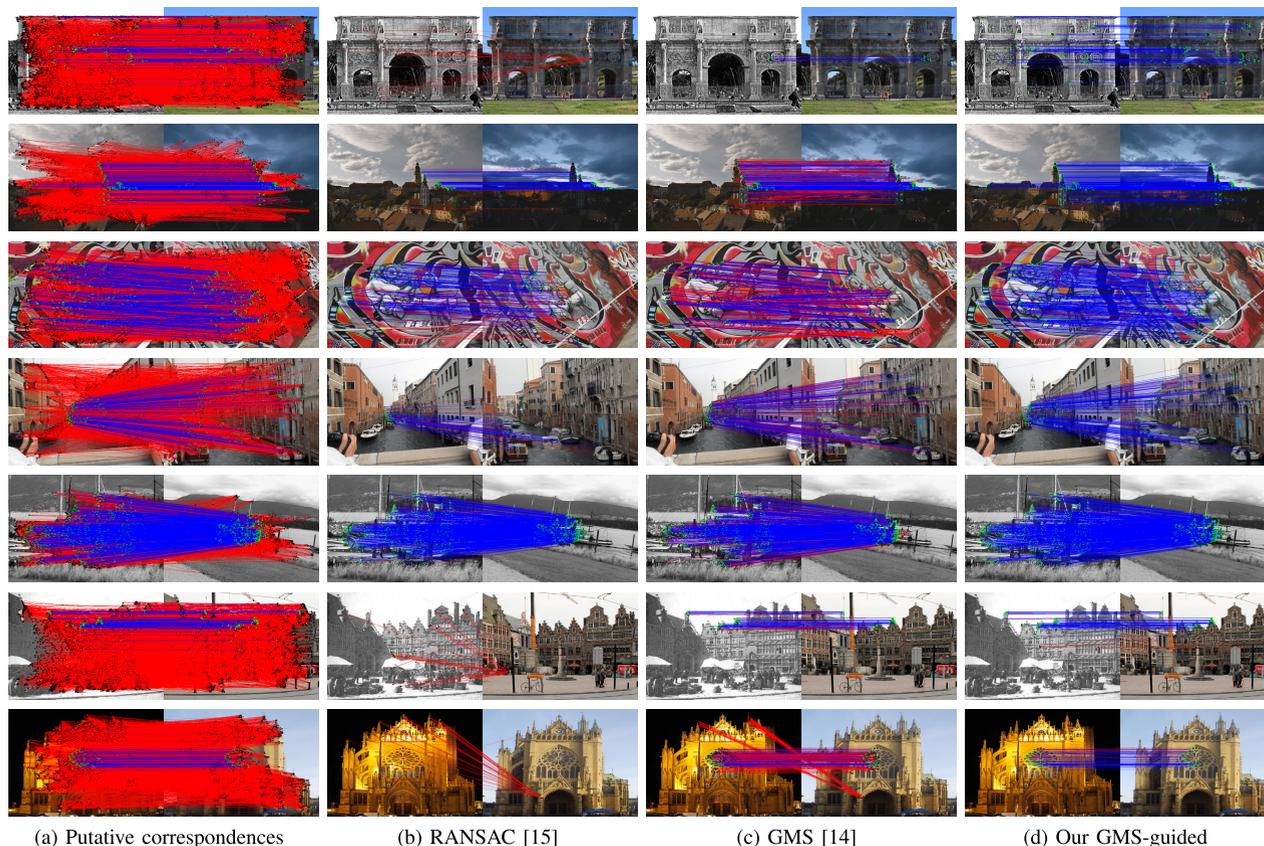
### 1) PERFORMANCE ON PUTATIVE SETS WITH LOW INLIER RATIOS

In this part, to demonstrate the performance of the correspondence selection from putative sets with low inlier ratios, experiments comparing six classic and state-of-the-art correspondence selection methods including RANSAC [15], GMS [14], GMS+RANSAC, LPM [21], VFC [23], LMR [29] and our GMS-guided algorithm are conducted.

Note that the selected typical image pairs contain different types of large variations or multiple types of changes, such as rendering and noise ('Archeasy' and 'Townsquare'), significant lighting change ('Daynight' and 'Metz'), large viewpoint change ('Graffiti'), large scale change ('Venice'), zoom and rotation ('Boat'). Moreover, the ORB feature adopted in our approach has relatively lower discrimination ability than the floating-points features such as SIFT. This is true especially for image pairs 'Archeasy', 'Townsquare' and 'Metz', where the variation of the rendering style makes it more challenging to maintain the features' descriptiveness. Furthermore, we set a relatively strict inlier threshold (2.5 pixels) to identify the inliers. Therefore, the obtained putative sets in our experiments have extremely low inlier ratios.

Table 1 presents the Precision, Recall, and F-measure statistics of the seven comparison approaches on the seven typical image pairs, along with the initial inlier ratios of the putative sets. As it shows, the inlier ratios of the seven typical image pairs ('Archeasy', 'Daynight', 'Graffiti', 'Venice', 'Boat', 'Townsquare' and 'Metz' from top to bottom respectively) are only 1.38%, 6.15%, 4.93%, 6.65%, 16.27%, 1.12% and 1.2%, respectively. These extremely low inlier ratios of the putative sets make the correspondence selection task very challenging.

Fig. 3 shows some examples of the correspondence selection results obtained by our GMS-guided approach and two approaches on the seven image pairs. In the figure, the blue lines indicate the true matches and the red lines indicate the false matches. Note that, due to space limitations, only the results of the two methods that are most relevant to ours (*i.e.*, GMS [14] and RANSAC [15]), and our GMS-guided algorithm are shown in this figure.



**FIGURE 3.** Samples of the correspondence selection results obtained with RANSAC [15], GMS [14] and our GMS-guided approach on seven typical image pairs ('Archcasy', 'Daynight', 'Graffiti', 'Venice', 'Boat', 'Townsquare' and 'Metz' from top to bottom respectively) containing different types of variations. The inlier ratios in putative correspondence sets of the seven image pairs (from top to bottom) are 1.38%, 6.15%, 4.93%, 6.65%, 16.27%, 1.12%, and 1.20% respectively. The blue lines indicate true matches and the red lines indicate the false matches.

Several observations can be made from the results shown in Table 1 and Fig. 3:

First, except for our GMS-guided approach, the performance of all correspondence selection algorithms deteriorate dramatically because of the low inlier ratios. For instance, the VFC algorithm fails on four image pairs (*i.e.*, 'Archeasy', 'Townsquare', 'Metz', and 'Venice'), and the RANSAC algorithm becomes ineffective on three image pairs (*i.e.*, 'Archeasy', 'Townsquare', and 'Metz'). The main reason is that both of the inlier number and the inlier ratio in the putative sets of these images are very small. In other words, even many high-performing methods are unable to select correct correspondences effectively in putative sets with low inlier ratios.

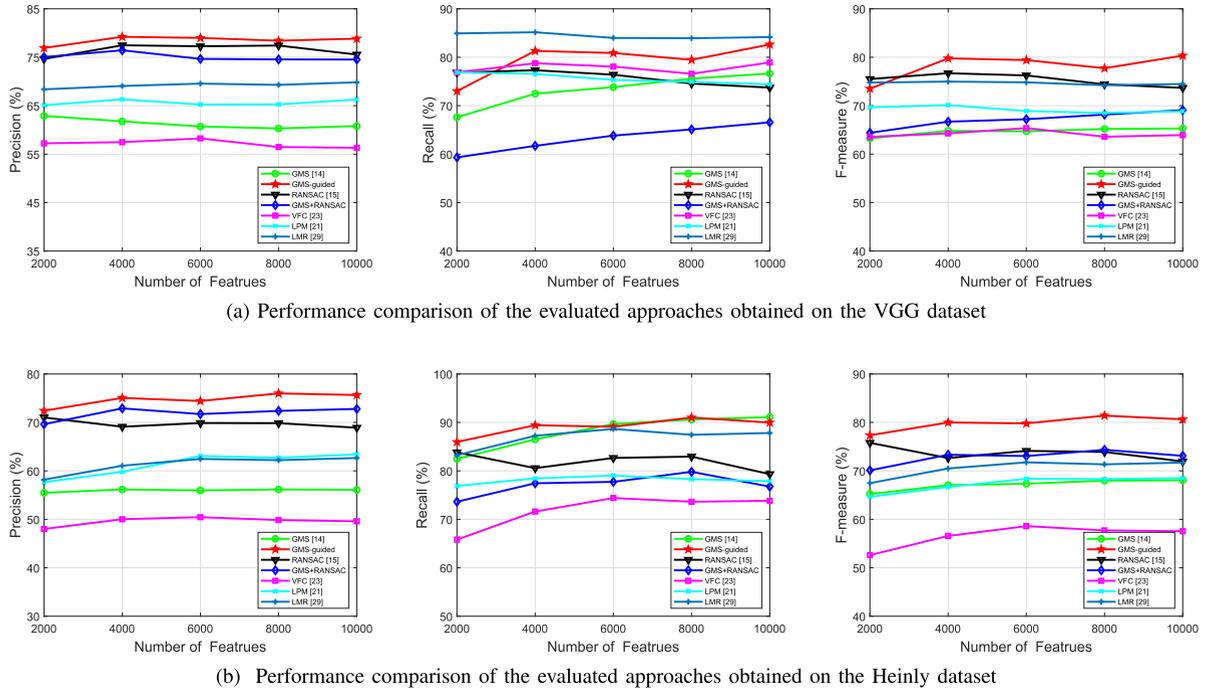
On the contrary, our GMS-guided algorithm can process putative sets with extremely low inlier ratios without any problem. For example, our GMS-guided algorithm has selected all correct matches without any false matches when the inlier ratio in the putative set is only just 1.38% in the image pair 'Archeasy' (see the first row of Fig. 3), and also achieved almost perfect performance on the image pair 'Daynight' (with an inlier ratio of 6.15%). In addition, in terms of the average Precision, Recall and F-measure, in all the seven image pairs of an average inlier ratio of 5.39%,

the performance obtained with our GMS-guided approach, has achieved 92.96%, 92.47%, and 92.62% respectively, which have significantly exceeded those of the comparative approaches.

Last, among the methods compared, GMS and GMS+RANSAC have both achieved relatively good performance. But GMS has yielded relatively low Precision because it remains a fraction of false matches when removing the mismatches using the motion smoothness constraint (as shown in the third column of Fig. 3). The Precision of GMS + RANSAC is higher, but its Recall is lower, because this method limits false matches at the cost of eliminating most of the true matches at the GMS stage. In contrast, the excellent results of our GMS-guided approach indicate that it can better solve the problem of selecting correct correspondences effectively in putative sets with low inlier ratios.

## 2) ROBUSTNESS

To assess the robustness of the evaluated algorithms, Fig. 4 reveals the trends of Precision, Recall and F-measure curves among the RANSAC [15], GMS [14], GMS+RANSAC, LPM [21], VFC [23], LMR [29] and our GMS-guided algorithm on the VGG [33] and Heiny [32] datasets.



**FIGURE 4. Performance comparison of the evaluated algorithms on VGG (a) and Heiny (b) datasets, in terms of Precision (left), Recall (middle) and F-measure (right) under different correspondence numbers in putative sets. The average inlier ratios in the putative sets are 34.28% on VGG dataset and 33.62% on Heiny dataset respectively.**

In the experiments, different numbers of ORB feature points (such as 2,000, 4,000, 6,000, 8,000 and 10,000) are selected to get the corresponding numbers of putative correspondences. The average inlier ratio in putative sets is 34.28% on VGG dataset and 33.62% on Heiny dataset. From Fig. 4, we can see that except for the Recall on the VGG dataset, our GMS-guided approach has achieved the highest performance in terms of Precision, Recall, and F-measure on both datasets. The RANSANC algorithm produced satisfying results because we set a big enough number of iterations and the inlier ratios of the putative sets are relatively high in our testing. LPM and LMR’s performance is not outstanding because they are better and more suitable for putative sets with higher inlier ratios. Therefore, in our experiments, they have not shown any competitive advantages. VFC and GMS both have high Recall but low Precision. In contrast, our method has the best Precision-Recall trade-off.

Another thing worth of noting is that, the curves associated with Precision, Recall, and F-measure fluctuate slightly with the increase of the number of ORB features in Fig. 4. On the VGG dataset, the inlier ratio in the putative set reaches to the maximum (35.37%) when the number of ORB features is 4000, then drops to the minimum (33.2%) when the number of ORB features is 10,000. On the Heiny dataset, the inlier ratios in putative set have only minor changes and the trend decreases as the number of ORB features increases.

It can be seen clearly that on both datasets, the Recall curve of GMS rises with the increasing number of ORB features,

indicating that GMS needs more matches for statistics to select the true matches. On the Heiny dataset (in Fig. 4(b)), we can see that with the increase of the ORB feature number, all three curves of VFC and LPM rise at first, reach the highest values when the ORB feature number is 6000, and then drop. This shows that, for LPM and VFC, a larger number of matches or true matches will lead to good performance to some extent, and 6000 matches is an optimal value in our experiments.

On the VGG dataset (in Fig. 4(a)), the trend of the three curves of RANSANC is consistent with the trend of initial ratio when the number of ORB features increases, showing that its performance is affected more by the initial inlier ratio. If we only consider the influence of the number of ORB features, LMR is most robust, and our method has the best performance among all methods.

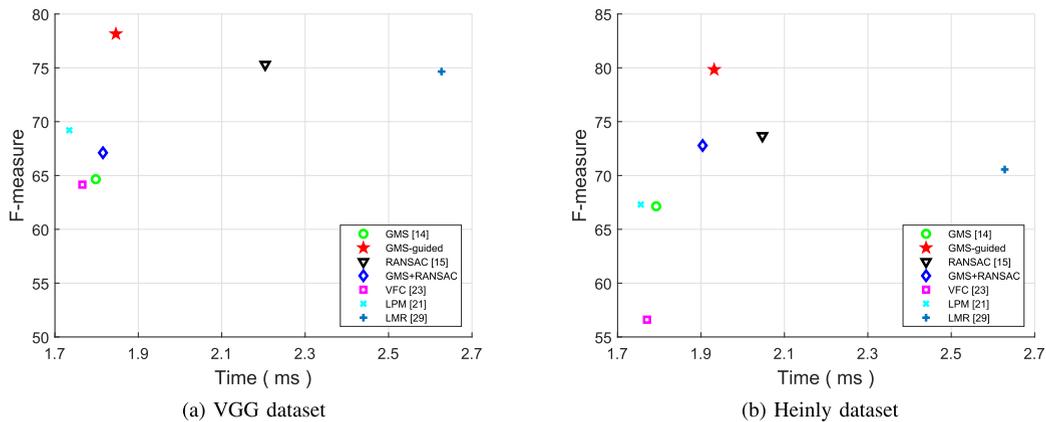
To further test the robustness of our GMS-guided approach, some specific nuisances (*e.g.*, zoom, rotation, blur, viewpoint change, light change, and JPEG compression) are adopted. Table 2 compares the average results obtained with different algorithms in terms of Precision, Recall and F-measure against various nuisances. As can be seen from Table 2, our GMS-guided approach has obtained the highest accuracy on Precision and F-measure besides viewpoint change. Also, only our algorithm has achieved both high Precision and high Recall, among all the tested algorithms. LMR shows excellent performance in three nuisances, *i.e.*, zoom, rotation, and viewpoint change. VFC always has a high Recall under various nuisances.

**TABLE 2.** Performance comparison of evaluated algorithms for images of various inlier ratio of the putative set (IR) against different nuisances. Precision (P), Recall (R) and F-measure (F). First, second, third best results are highlighted in color.

Case	IR (%)	Evaluation	VFC [23]	GMS [14]	LPM [21]	LMR [29]	RANSAC [15]	GMS-R	Our GMS-G
Zoom and rotation (boat)	28.03	P (%)	63.67	63.36	70.21	<b>77.03</b>	76.45	<b>82.60</b>	<b>85.79</b>
		R (%)	<b>95.90</b>	79.63	74.44	<b>94.27</b>	74.74	70.08	<b>96.08</b>
		F (%)	72.45	70.26	72.24	<b>83.66</b>	<b>75.46</b>	75.38	<b>89.23</b>
Blur (bike)	44.41	P (%)	65.02	63.21	70.91	73.65	<b>83.41</b>	<b>85.76</b>	<b>92.51</b>
		R (%)	<b>99.85</b>	<b>96.79</b>	85.67	93.14	80.37	83.10	<b>94.48</b>
		F (%)	77.71	75.45	76.72	81.52	<b>81.74</b>	<b>84.40</b>	<b>93.42</b>
Viewpoint change (wall)	24.09	P (%)	58.50	59.86	61.34	63.92	<b>67.79</b>	<b>67.88</b>	<b>70.50</b>
		R (%)	<b>75.96</b>	69.31	<b>76.24</b>	<b>89.12</b>	72.24	62.62	75.03
		F (%)	65.81	64.00	67.44	<b>72.73</b>	<b>69.84</b>	64.85	<b>72.57</b>
Light change (lenven)	59.87	P (%)	78.90	76.04	86.81	90.72	<b>93.29</b>	<b>92.91</b>	<b>98.36</b>
		R (%)	<b>100</b>	<b>96.89</b>	85.09	90.49	86.87	84.76	<b>97.75</b>
		F (%)	88.18	85.18	85.91	<b>88.66</b>	<b>91.85</b>	88.47	<b>98.05</b>
Rotation (ceil)	44.02	P (%)	62.18	62.85	68.80	68.00	<b>72.92</b>	<b>73.42</b>	<b>75.34</b>
		R (%)	94.66	<b>97.48</b>	91.54	<b>98.26</b>	88.68	83.65	<b>97.45</b>
		F (%)	74.35	75.74	77.94	<b>79.63</b>	<b>79.54</b>	77.71	<b>84.36</b>
JPEG compression (ubc)	69.41	P (%)	88.44	84.89	<b>95.75</b>	<b>96.01</b>	94.68	94.53	<b>99.92</b>
		R (%)	<b>99.99</b>	<b>99.04</b>	97.88	97.19	84.75	75.88	<b>99.32</b>
		F (%)	<b>93.54</b>	91.14	<b>96.72</b>	96.58	89.34	84.13	<b>99.62</b>
Zoom (venice)	23.70	P (%)	41.34	58.41	<b>70.78</b>	<b>85.02</b>	60.99	64.60	<b>91.15</b>
		R (%)	48.03	<b>86.53</b>	67.78	65.51	67.40	<b>90.79</b>	<b>88.49</b>
		F (%)	42.22	69.39	68.97	<b>73.41</b>	63.70	<b>74.77</b>	<b>89.49</b>

**TABLE 3.** Average F-measure of our GMS-guided and seven competitors in putative sets with different inlier ratios (IRs) on the VGG dataset. The result highlighted in red are the best.

Descriptor	IR (%)	F-measure (%)							
		LFGC [28]	GMS-R	GMS [14]	VFC [23]	LPM [21]	RANSAC [15]	LMR [29]	Our GMS-G
ORB	52.24	81.95	89.06	92.24	86.84	94.77	94.89	96.44	<b>97.06</b>
SIFT	88.10	87.07	87.57	85.58	97.48	95.98	95.30	97.10	<b>98.49</b>
SURF	57.61	76.85	84.91	81.69	91.00	91.54	94.84	<b>95.72</b>	94.04



**FIGURE 5.** Efficiency vs. F-measure plots on VGG (a) and Heiny (b) datasets. The efficiency-axis is shown logarithmically for clarity.

All in all, the results in Table 2 and Fig. 4 show that our GMS-guided algorithm has the best robustness compared with the-state-of-the-art methods.

### 3) GENERALITY OF DIFFERENT FEATURE DESCRIPTORS

To this end, we evaluate the putative sets constructed from different feature descriptors such as SIFT [7], ORB [8] and SURF [36] on the VGG dataset. The average values of F-measure of the eight competitors are summarized in Table 3.

It should be noted that we add the LFGC for comparison in this table, some of the data are cited from literature [29], and we select the best result (LMR-RF-10) from the LMR paper for comparison.

As shown in Table 3, the inlier ratios of the correspondence sets are 52.24%, 88.10%, and 57.61% respectively. Our GMS-guided algorithm can achieve the best performance with both SIFT and ORB feature descriptors and the second best performance with SURF feature descriptors. As a learning based approach, the LFGC method performs poorly

because it is designed for large baseline image matching and aims to accurately recover the transformation matrix.

Thus, we can draw a conclusion that our GMS-guided approach does not rely on any specific feature descriptors, and can work well on putative sets with both high and low inlier ratios.

#### 4) TIME EFFICIENCY

For the convenience of the reader, both of the selection performance and efficiency are taken into consideration. Fig. 5 (a) and (b) presents the efficiency v.s. F-measure plots on VGG and Heinly datasets respectively. Note that the run time does not include the time of initial matching.

As shown in Fig. 5, our GMS-guided algorithm has achieved the best F-measure performance and good computing speed. In fact, for the GMS series (such as the GMS, GMS-guided, and GMS+RANSAC algorithm), there is little difference in their speeds. So, our GMS-guided algorithm can achieve a good balance between the selection performance and efficiency. Besides, our GMS-guided algorithm only requires dozens of milliseconds for mismatch removal from thousands of putative matches to achieve comparable results. Therefore, our proposed GMS-guided algorithm outperforms the state of the arts on both effectiveness and efficiency in terms of dealing with the correspondence sets of both high and low inlier ratios.

## V. CONCLUSION

In this paper, we have proposed a GMS-guided correspondence selection strategy to handle the extreme outlier problem, which has significantly boosted true matches without sacrificing accuracy. The comparative results on testing benchmark datasets for correspondence selection have demonstrated that our GMS-guided algorithm can effectively improve the correspondence selection performance and obtain good computing speed. Therefore, it can be applied in Real-time Visual Simultaneous Localization and Mapping (VSLAM) systems. The results in putative sets with low inlier ratios also indicate that our method is effective for addressing the problem of the putative sets with low inlier ratios. This means that our method can also be used in matching low-overlapping images in image stitching.

## REFERENCES

- [1] S. Benhimane and E. Malis, "Real-time image-based tracking of planes using efficient second-order minimization," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, vol. 1, Oct. 2004, pp. 943–948.
- [2] S. Hare, A. Saffari, and P. H. S. Torr, "Efficient online structured output learning for keypoint-based object tracking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 1894–1901.
- [3] H. Hirschmuller, "Stereo processing by semiglobal matching and mutual information," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 2, pp. 328–341, Feb. 2008.
- [4] J. Jia and C.-K. Tang, "Image stitching using structure deformation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 4, pp. 617–631, Apr. 2008.
- [5] N. Snavely, S. M. Seitz, and R. Szeliski, "Modeling the world from Internet photo collections," *Int. J. Comput. Vis.*, vol. 80, no. 2, pp. 189–210, Dec. 2007.
- [6] C. Leng, H. Zhang, B. Li, G. Cai, Z. Pei, and L. He, "Local feature descriptor for image matching: A survey," *IEEE Access*, vol. 7, pp. 6424–6434, 2019.
- [7] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, Nov. 2004.
- [8] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "ORB: An efficient alternative to SIFT or SURF," in *Proc. Int. Conf. Comput. Vis.*, Nov. 2011, vol. 11, no. 1, p. 2.
- [9] J. Ma, J. Zhao, J. Tian, Z. Tu, and A. L. Yuille, "Robust estimation of nonrigid transformation for point set registration," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 2147–2154.
- [10] Y. Lin, Z. Lin, and H. Zha, "The shape interaction matrix-based affine invariant mismatch removal for partial-duplicate image search," *IEEE Trans. Image Process.*, vol. 26, no. 2, pp. 561–573, Feb. 2017.
- [11] G. Wang, Z. Wang, Y. Chen, Q. Zhou, and W. Zhao, "Context-aware Gaussian fields for non-rigid point set registration," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 5811–5819.
- [12] G. Lv, "A novel correspondence selection technique for affine rigid image registration," *IEEE Access*, vol. 6, pp. 32023–32034, 2018.
- [13] J. L. Schonberger, H. Hardmeier, T. Sattler, and M. Pollefeys, "Comparative evaluation of hand-crafted and learned local features," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1482–1491.
- [14] J. Bian, W.-Y. Lin, Y. Matsushita, S.-K. Yeung, T.-D. Nguyen, and M.-M. Cheng, "GMS: Grid-based motion stitching for fast, ultra-robust feature correspondence," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4181–4190.
- [15] M. A. Fischler and R. C. 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.
- [16] 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.
- [17] 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.
- [18] D. Barath and J. Matas, "Graph-cut RANSAC," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 6733–6741.
- [19] 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.
- [20] X. Jiang, J. Ma, and J. Chen, "Progressive filtering for feature matching," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2019, pp. 2217–2221.
- [21] 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, Sep. 2018.
- [22] 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.
- [23] J. Zhao, J. Ma, J. Tian, J. Ma, and D. Zhang, "A robust method for vector field learning with application to mismatch removing," in *Proc. CVPR*, Jun. 2011, pp. 2977–2984.
- [24] 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, no. 4, pp. 1706–1721, Apr. 2014.
- [25] J. Ma, J. Wu, J. Zhao, J. Jiang, H. Zhou, and Q. Z. Sheng, "Nonrigid point set registration with robust transformation learning under manifold regularization," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 12, pp. 3584–3597, Dec. 2019.
- [26] X. Jiang, J. Ma, J. Jiang, and X. Guo, "Robust feature matching using spatial clustering with heavy outliers," *IEEE Trans. Image Process.*, vol. 29, no. 2, pp. 736–746, Aug. 2019.
- [27] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," in *Proc. KDD*, vol. 96, 1996, no. 34, pp. 226–231.
- [28] 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.
- [29] J. Ma, X. Jiang, J. Jiang, J. Zhao, and X. Guo, "LMR: Learning a two-class classifier for mismatch removal," *IEEE Trans. Image Process.*, vol. 28, no. 8, pp. 4045–4059, Aug. 2019.
- [30] 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.

[31] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. Cambridge, U.K.: Cambridge Univ. Press, 2003.

[32] J. Heinly, E. Dunn, and J.-M. Frahm, "Comparative evaluation of binary features," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*. Berlin, Germany: Springer, 2012, pp. 759–773.

[33] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. V. Gool, "A comparison of affine region detectors," *Int. J. Comput. Vis.*, vol. 65, nos. 1–2, pp. 43–72, Oct. 2005.

[34] D. C. Hauage and N. Snavely, "Image matching using local symmetry features," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 206–213.

[35] C. Zhao, J. Yang, Y. Xiao, and Z. Cao, "Comparative evaluation of 2D feature correspondence selection algorithms," 2019, *arXiv:1904.13383*. [Online]. Available: <http://arxiv.org/abs/1904.13383>

[36] H. Bay, T. Tuytelaars, and L. Van Gool, "Surf: Speeded up robust features," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*. Berlin, Germany: Springer, 2006, pp. 404–417.



**WENJING JIA** received the Ph.D. degree in computing sciences from the University of Technology Sydney, Australia, in 2007. Her research interests are mainly in the areas of image processing/analysis, computer vision, and pattern recognition, particularly car license plate detection and text detection.



**XIANGJIAN HE** (Senior Member, IEEE) received the Ph.D. degree in computing sciences from the University of Technology Sydney, Australia, in 1999. Since 1999, he has been with the University of Technology Sydney, where he is currently a Full Professor and the Director of the Computer Vision and Pattern Recognition Laboratory.



**QUNFANG TANG** received the master's degree from Hunan University, China, in 2010. She is currently pursuing the Ph.D. degree in communication and information engineering with the Wuhan University of Technology, Wuhan, China. Her research interests are mainly in the areas of image processing, computer vision, and pattern recognition.



**QINGNIAN ZHANG** received the Ph.D. degree in mechanical design and theory from the Wuhan University of Technology, China, in 2002. He is currently a Professor with the School of Transportation, Wuhan University of Technology. His research interests are mainly in the areas of smart transportation, data statistics, and model analysis.



**JIE YANG** received the Ph.D. degree in precision instrument and mechanics from Shanghai Jiao Tong University, China, in 1999. She is currently a Professor with the Hubei Key Laboratory of Broadband Wireless Communication and Sensor Networks, Wuhan University of Technology, China. Her research interests are mainly in the areas of image processing, information hiding, cryptography, and multimedia communication.



**HAIBO LIU** received the Ph.D. degree in communication and information engineering from the Wuhan University of Technology, China, in 2015. His research interests are mainly in the areas of image processing, pattern recognition, and blind signal processing.

...