Image Feature Correspondence Selection: A Comparative Study and a New Contribution

Chen Zhao, Zhiguo Cao, Jiaqi Yang, Ke Xian, and Xin Li, *Fellow, IEEE*

***Abstract*— Image feature correspondence selection is pivotal to many computer vision tasks from object recognition to 3D reconstruction. Although many correspondence selection algorithms have been developed in the past decade, there still lacks an in-depth evaluation and comparison in the open literature, which makes it difficult to choose the appropriate algorithm for a specific application. This paper attempts to fill this gap by evaluating eight competing correspondence selection algorithms including both classical methods and current state-of-the-art ones. In addition to preselected correspondences, we have compared different combinations of detector and descriptor on four standard datasets. The diversity of those datasets cover a wide range of uncertainty factors including zoom, rotation, blur, viewpoint change, JPEG compression, light change, different rendering styles and multiple structures. We have measured the quality of competing correspondence selection algorithms in terms of four performance metrics - i.e., precision, recall, F-measure and efficiency. Moreover, we propose to combine the strengths of eight competing methods by combining their correspondence selection results. Extensive experimental results are reported to demonstrate the superiority of several fusion strategies to individual methods, which suggests the possibility of adaptively combining those methods for even better performance.**

***Index Terms*— Image feature correspondence, feature matching, correspondence selection, inliers.**

I. Introduction

FEATURE correspondence selection is a fundamental task in computer vision, pattern recognition and robotics. It is the building block to a wide range of applications such as structure-from-motion [1], simultaneous localization and mapping [2], tracking [3], image stitching [4], face verification [5], image retrieval [6], and object recognition [7]. The main purpose of correspondence selection is to retrieve as many as correct correspondences (also known as *inliers*) from the given image pair. The general process of feature matching often starts from detecting representative points (namely keypoints) - e.g., local detectors [8]–[10] can be used to extract keypoints from a given image. To establish the correspondence between two images, keypoints with similar feature descriptors have to be matched, generating a set of initial feature matches.

(a) Initial feature matches (b) Selected matching

Fig. 1. An exemplar illustration of image feature correspondence selection, where colorized lines represent correspondences between two images. (a) Initial feature matches generated by brute-force matching. (b) Feature matches after correspondence selection.

Initial feature matches often suffer from undesirable incorrect matches (as shown in Fig. 1 (a)) due to limited distinctiveness of feature descriptors or/and external interference such as noise and occlusion. This problem makes *correspondence selection* a necessity for accurate feature matching. As shown in Fig. 1 (b), matches after correspondence selection are far more consistent, which greatly facilitate high-level vision tasks such as the estimation of homography, affine transformation and essential matrix [4]. Other applications such as camera parameter estimation [1] and object tracking [3] also require correspondence selection as a preprocessing step. Nonetheless, the correspondence selection problem is difficult in real applications due to various uncertainty factors including zoom, rotation, blur, viewpoint change, JPEG compression, light change, different rendering styles, multiple structures etc.

To address these challenges, many approaches have been developed in the past two decades and can be classified into two categories [11]: *parametric* and *non-parametric*. Parametric methods seek consistent correspondences defined by parametric geometric models. Typical methods include the random sample consensus (RANSAC) [12], the progressive sample consensus (PROSAC) [13], the universal framework for random sample consensus (USAC) [14], and etc. Non-parametric methods often search correspondence inliers via either feature similarity constraint or geometric constraint, such as the nearest neighbor similarity ratio (NNSR) [8], spectral technique (ST) [15], game-theoretic matching (GTM) [16], graph-based matching [17], [18] and locality preserving

Manuscript received August 2, 2019; revised November 13, 2019; accepted December 4, 2019. Date of publication January 3, 2020; date of current version January 30, 2020. This work was supported in part by the National Natural Science Foundation of China under Grant 61876211 and in part by the 111 Project on Computational Intelligence and Intelligent Control under Grant B18024. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Jiwen Lu. *(Corresponding author: Ke Xian.)*

Chen Zhao, Zhiguo Cao, and Ke Xian are with the School of Artificial Intelligence and Automation, Huazhong University of Science and Technology, Wuhan 430074, China (e-mail: hust_zhao@hust.edu.cn; zgcao@hust.edu.cn; kexian@hust.edu.cn).

Xin Li is with Lane Department of Computer Science and Electrical Engineering, West Virginia University, Morgantown, WV 26506 USA (e-mail: xin.li@mail.wvu.edu).

Jiaqi Yang was with the School of Artificial Intelligence and Automation, Huazhong University of Science and Technology, Wuhan 430074, China. He is now with the School of Computer Science, Northwestern Polytechnical University, Xi'an, China (e-mail: jqyang@hust.edu.cn).

This article has supplementary downloadable material available at http://ieeexplore.ieee.org, provided by the authors.

Digital Object Identifier 10.1109/TIP.2019.2962678

matching (LPM) [19]. There are also constraint-independent non-parametric methods such as identifying point correspondences by correspondence function (ICF) [20], vector field consensus (VFC) [11], grid-based motion statistics (GMS) [21] and coherence based decision boundaries (CODE) [22].

Performance evaluation of competing image feature matching techniques also exists in the literature. For instance, In [23] and [24], the performance of several 2D feature descriptors is compared; in [25] a similar study is conducted for 2D feature detectors. An aggregated evaluation of both 2D detectors and descriptors can be found in [26]. Additionally, performance evaluation for a set of random sample consensus methods including the popular RANSAC and its variants are conducted in [27]. Unfortunately, existing studies are not comprehensive enough for an in-depth comparison and suffer from the following limitations. First, the critical step in correspondence selection is to reach correspondence consensus, while feature detection and description only aim at building high-quality initial feature correspondences (note that consensus is difficult to be guaranteed without correspondence selection [8]). Second, only parametric methods are evaluated in [27] (non-parametric approaches and more recent algorithms are left out).

In this paper, we present the first comprehensive evaluation for image feature correspondence selection, to the best of our knowledge. The considered methods in our evaluation range from classical algorithms to the most recent ones, covering both parametric and non-parametric approaches. More specifically, RANSAC [12] and USAC [14] are selected from the parametric family. As for non-parametric methods, we choose NNSR [8] as the representative based on descriptor similarity constraints. Additionally, ST [15], GTM [16] and LPM [19] are selected as they all rely on geometric consensus. VFC [11] and the recent GMS [21] are also taken into consideration since they eliminate outliers from the perspective of statistical measures.

In order to conduct a comprehensive comparison among those competing methods, we choose four standard datasets - i.e., VGG [28], Heinly [24], Symbench [29], AdelaideRMF [30] - because together they cover a variety of nuisances arising from real world scenarios. Among them, VGG is a hybrid dataset containing challenges including zoom, rotation, blur, JPEG compression, light and viewpoint change. Heinly contains pure zoom and rotation. Symbench involves scenes with light changes and varying rendering styles. AdelaideRMF possesses viewpoint change, multiple structures, and dynamic scenes, resulting in multiple separate local transformations. Although the size of these evaluated datasets is smaller than some public datasets -e.g., Imagenet [31], in the field of image classification, segmentation, and etc., datasets employed in our evaluation benchmark cover *most challenges* in correspondence selection and have been *widely used* in previous correspondence selection approaches [19], [32]–[34].

The performance of each competing method is quantitatively measured using precision, recall and F-measure [19], [21], [32]. The robustness against the specific nuisance and efficiency with respect to different scales of initial feature matches are also examined. In addition, the performance under preselected correspondences (with higher inlier ratios) and different detector-descriptor combinations are accessed to test their flexibility with respect to the inlier ratio and correspondence distribution changes. Based on experimental findings, we make an aggregated summary about the advantages and limitations of our evaluated methods as well as their suitable applications. Furthermore, we suggest that combining correspondence selection results of different methods is a convenient yet powerful solution to achieve even better performance. We have compared several popular strategies of combining strategies under the framework of correspondence selection.

In a nutshell, the contributions of this paper are threefold:

- We conduct a comprehensive review of the core computation steps in eight state-of-the-art image feature correspondence selection algorithms along with their connections and differences.
- We quantitatively evaluate and compare the performance (including the robustness and the efficiency) of each algorithm on four standard datasets covering a wide range of uncertainty factors such as zoom, rotation, blur, viewpoint change, JPEG compression, light change, different rendering styles and multiple structures.
- We propose a fusion-based strategy that combines the strengths of different correspondence selection methods and achieves even more robust correspondence selection results.

The remainder of this paper is organized as follows. Sect. II presents the background information related to image feature correspondence selection and performance evaluation. Sect. III presents the core computation steps of eight state-of-the-art approaches and discusses their connections. Sect. IV proposes the details of fusion-based strategies towards correspondence selection. Sect. V describes the experimental setup including test datasets, evaluation criteria and implementation details of the evaluated methods. Qualitative and quantitative experimental results are shown in Sect. VI. We include summary and discussion in Sect. VII. and draw some final conclusions in Sect. VIII.

II. Related Work

A. Correspondence Selection Methods

For parametric methods, the most well-known algorithm is arguably RANSAC presented in [12]. RANSAC iteratively explores the space of model parameters by randomly sampling and estimates the most reliable model based on the maximum number of inliers. Then, outliers can be removed using the generated model. Several variants of RANSAC such as MLESAC [35], LO-RANSAC [36], PROSAC [13] and USAC [14] were proposed in the following decades. MLESAC employs the maximum likelihood estimation rather than the inlier count to check the solutions. LO-RANSAC inserts an optimization process where the generated model is refined by the subset of inliers. A weighted sampling step is adopted instead of random sampling in PROSAC. This method sorts the raw correspondences by matching quality and generates hypotheses from the most promising correspondences. USAC extends the standard hypothesize-and-verify structure in RANSAC and presents a universal framework

that integrates advantages of previous parametric methods. In addition, some other approaches relying on local parametric structures have also been developed, such as agglomerative correspondence clustering (ACC) [37], multi-structures robust fitting (Multi-GS) [38], Hough voting and inverted Hough voting (HVIV) [39]. ACC uses Hessian-affine detector [40], which is invariant to affine transformations, to estimate the local homography matrix as constraints. The initial correspondences are then clustered based on the constraints, and the clusters with inliers are supposed to be larger than the ones constituted by outliers. Multi-GS generates a series of tentative hypotheses by random sampling and considers that two correspondences from the same local structure are inliers if they share a common list of hypotheses. HVIV employs the BPLR detector [41] to cluster correspondences and estimates the homographic transformation for each correspondence as well. The most plausible correspondence in each cluster is then selected using normalized kernel density estimation.

For non-parametric methods, their theoretical foundations are not always the same. A widely-used strategy is exploiting the consistency information of local geometric structures or appearance (feature similarity). Specifically, in [8], a nearest neighbor similarity ratio (NNSR) method was proposed to assign a ratio-based penalty to each correspondence and treats those correspondences with low ratios as inliers. In spectral technique (ST) [15], an affinity matrix is built using pairwise geometric constraints to remove mismatches in conflict with the most credible correspondences. In [16] the selection of correspondences was cast into a game theoretic framework, known as game-theoretic matching (GTM), where a natural selection process allows corresponding points that satisfy a mutual distance constraint to thrive. In [42], reweighted random walk algorithm (RRWM) was presented for graph matching. An associated graph between two sets of candidate correspondences is drawn at first, and reliable nodes indicating the consistent correspondences in this graph are then selected by the reweighted random walk algorithm. In [19], locality preserving matching (LPM) was proposed to improve inlier selection by maintaining the local neighborhood structures of those potential true matches. Some non-parametric approaches that formulate the correspondence selection problem as a statistics problem have also been used, e.g., vector field consensus (VFC) [11] and grid-based motion statistics (GMS) [21]. VFC supposes that the noise around inliers and outliers falls in different distributions. This approach estimates the probability of inliers by the maximum likelihood estimation for parameters in the mixture probabilistic model. Additionally, GMS rejects false matches by counting the quantity of matches in small neighborhoods and achieves real-time performance with an efficient grid-based score estimator.

B. Performance Evaluation

In image feature matching, some evaluations of 2D/3D local descriptors and detectors have been performed. For instance, in [23] the performance of 2D feature descriptors was evaluated under transformations of rotation, zoom, viewpoint change, blur, JPEG compression, light change and keypoint localization errors. In [26], an evaluation of several groups of 2D feature detectors and descriptors was conducted on images captured from the same 3D object with different viewpoints and lighting conditions. In [24], an evaluation of several 2D binary descriptors was performed, aiming at testing their descriptiveness under different feature detectors on several scenes with illumination change, viewpoint change, pure camera rotation and pure scale change. In [25] the performance of several 2D feature detectors was investigated on a particular dataset wherein each scene was depicted from 119 camera positions with a range of light directions. In 3D domain, [43] compared two categories (i.e., fixed-scale and adaptive-scale) of 3D feature detectors in terms of distinctiveness, repeatability and efficiency under the nuisances of viewpoint changes, clutter, occlusions and noise. In [44], the descriptiveness, robustness, compactness and efficiency of ten local geometric descriptors were tested on eight datasets with radius variations, varying mesh resolution, Gaussian noise and etc. More relevant to our work is the evaluation performed in [27], where RANSAC and a set of its variants were examined under different ratios of inliers. This paper, compared with [27], considers both parametric and non-parametric methods as well as a variety of nuisances for more comprehensive evaluation.

III. Benchmark Methods

Eight image feature correspondence selection algorithms including two parametric ones (i.e., RANSAC [12] and USAC [14]) and six non-parametric ones (i.e., NNSR [8], ST [15], GTM [16], VFC [11], GMS [21], LPM [19]) are considered in our evaluation. To facilitate our review, we start with some basic notations.

Given two images (I, I'), keypoint locations and local feature descriptors are computed as $(\mathcal{K}, \mathcal{K}')$ and $(\mathcal{F}, \mathcal{F}')$, respectively. This procedure can be done using off-the-shelf detectors and descriptors (e.g., SIFT [8]). To generate initial feature matches $\mathcal{C}$, keypoints are matched with each other based on the similarity of feature descriptors - i.e., a correspondence (match) in $\mathcal{C}$ is defined as $c = \{\mathbf{x}, \mathbf{x}', \arg\max_{\mathbf{f}'} s_{\mathcal{F}}(\mathbf{f}, \mathbf{f}')\}$ with $\mathbf{x} \in \mathcal{K}$, $\mathbf{x}' \in \mathcal{K}'$, $\mathbf{f} \in \mathcal{F}$, $\mathbf{f}' \in \mathcal{F}'$ and $s_{\mathcal{F}}$ being the feature similarity score. The objective of correspondence selection is to seek the maximum consensus (inlier) subset $\mathcal{C}_{inlier} \subseteq \mathcal{C}$. Key principles and computation steps of eight competing algorithms are briefly reviewed as follows. Values of thresholds required by those algorithms are summarized in Table I.

A. Nearest Neighbor Similarity Ratio (NNSR) [8]

NNSR directly utilizes descriptor similarities to remove less distinctive matches. Specifically, the ratio of the closest and the second-closest feature distance to each correspondence is used as a penalty. That is, a correspondence is judged as an inlier if

$$\frac{\| \mathbf{f} - \mathbf{f}'_1 \|_2}{\| \mathbf{f} - \mathbf{f}'_2 \|_2} \leq t_{nnsr}, \tag{1}$$

where $t_{nnsr} \in [0, 1]$, $\|\cdot\|_2$ denotes the L_2 norm as suggested in [8], $\mathbf{f}'_1$ and $\mathbf{f}'_2$ represent the most and the second most similar feature descriptors of $\mathbf{f}$, respectively.

TABLE I

PARAMETER SETTINGS AND IMPLEMENTATIONS OF EVALUATED ALGORITHMS, WHERE *pix* REPRESENTS THE PIXEL UNIT

No.	Algorithm	Implementation	Parameters	Setting
1	NNSR [8]	OpenCV	t_{nnsr}	Adaptive [45]
2	RANSAC [12]	OpenCV	t_{ransac}	10*pix*
			n_{ransac}	2000
3	ST [15]	MATLAB	t_{st}	0.3
4	GTM [16]	OpenCV	t_{gtm}	Adaptive [45]
			n_{gtm}	100
			λgtm	0.0001
5	USAC [14]	OpenCV	n_{usac}	850000
			$t_{\mathbf{H}}$	10*pix*
			$t_{\mathbf{F}}$	1.5*pix*
6	VFC [11]	OpenCV	β	0.1
			λ_{vfc}	3
			t_{vfc}	0.75
			γ	0.9
7	GMS [21]	OpenCV	α	4
8	LPM [19]	MATLAB	λ_{lpm}	6
			k	4

B. Random Sample Consensus (RANSAC) [12]

RANSAC follows a trial-and-error framework by repeating procedures of random sampling and checking to maximize the objective function. For image feature correspondence selection, the desired parametric model is usually a plane homography matrix or a fundamental matrix. Taking the homography matrix as an example, it first randomly samples several correspondences (at least 4) from $\mathcal{C}$ and generates the model hypothesis $\mathbf{H}_i$ for those samples at the ith iteration. Then, the quality of hypothesis $\mathbf{H}_i$ is checked via the following objective function

$$O_i = \sum_{c \in \mathcal{C}} h_i(c), \tag{2}$$

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

$$h_i(c) = \begin{cases} 1, & \text{if } \|\mathbf{x}' - \rho\left(\mathbf{H}_i \begin{bmatrix} \mathbf{x} \\ 1 \end{bmatrix}\right)\|_2 \leq t_{ransac} \\ 0, & \text{otherwise,} \end{cases} \tag{3}$$

with $\rho([a_1 \ a_2 \ a_3]^T) = [a_1/a_3 \ a_2/a_3]^T$ and t_{ransac} being a threshold that determines the confidence of an inlier. Those steps are repeated n_{ransac} times and the model with the maximum objective function is selected as the final model $\mathbf{H}^\star$. Correspondences consistent with $\mathbf{H}^\star$ (producing output value of ones using Eq. 3) are identified as inliers.

C. Spectral (ST) [15]

This method locates the most reliable element by matrix decomposition. It assumes that the connection among correct matches is much tighter than the one among mismatches. Based on this assumption, ST first builds an adjacency matrix $\mathbf{A}$ as

$$a_{ij} = \min\left(\frac{\|\mathbf{x}_i - \mathbf{x}_j\|_2}{\|\mathbf{x}'_i - \mathbf{x}'_j\|_2}, \frac{\|\mathbf{x}'_i - \mathbf{x}'_j\|_2}{\|\mathbf{x}_i - \mathbf{x}_j\|_2}\right), \tag{4}$$

where $a_{ij} \in \mathbf{A}$ is the affinity between c_i and c_j. Second, the principal eigenvector $\mathbf{v}_{st}$ of $\mathbf{A}$ is computed using singular value decomposition (SVD). Third, the maximum element in $\mathbf{v}_{st}$ is selected as v_i indicating c_i being the most reliable correspondence. Fourth, set v_i to zero and remove other components of $\mathcal{C}$ that are in conflict with c_i, i.e.,

$$a_{ij} \leq t_{st}, \tag{5}$$

where t_{st} is a predefined threshold. By repeating the third and fourth steps until $\mathcal{C}$ is empty or $v_i = 0$, the correspondences related to all elements selected from $\mathbf{v}_{st}$ are determined as inliers.

D. Game Theory Matching (GTM) [16]

GTM concentrates on extracting correspondences being consistent to the majority of $\mathcal{C}$. Specifically, this strategy interprets the filtering process as a game-theoretic framework where players attempt to obtain highest payoffs. At the beginning of this game, every two players extracted from a large population choose a pair of correspondences (served as strategies of game playing in this context) from $\mathcal{C}$. Then they will receive a payoff linearly correlated to the coherence between these correspondences. The player who gets high payoffs will receive higher supports. In general, as the game going on, players will prefer to select more reliable correspondences to pursue higher pay-offs.

Given a pair of correspondences (c_i, c_j), the payoff function is defined as

$$\Pi_{ij} = e^{-\lambda_{GTM} \max(|T_i(\mathbf{x}_i) - T_j(\mathbf{x}_i)|, |T_i(\mathbf{x}_j) - T_j(\mathbf{x}_j)|)}, \tag{6}$$

where λ_{GTM} is a selectivity parameter, $|\cdot|$ represents the L_1 norm and $T_i(\mathbf{x})$ is the similarity transformation estimated by (similarly for $T_j(\mathbf{x})$)

$$T_i(\mathbf{x}) = \rho\left(\mathbf{H}_{c_i} \begin{bmatrix} \mathbf{x} \\ 1 \end{bmatrix}\right), \tag{7}$$

where $\mathbf{H}_{c_i}$ is the homographic transformation of c_i. Note that this algorithm particularly requires the local affine transformation cue to compute the pay-off function. Next, the payoff matrix $\mathbf{P}_{GTM}$ with the element in the ith row and jth column that is defined as

$$p_{ij} = \begin{cases} \Pi_{ij}, & \text{if } i \neq j \\ 0, & \text{otherwise,} \end{cases} \tag{8}$$

can be generated. The population vector $\mathbf{q}$ is updated by the evolutionary stable states algorithm (ESS's) [46] as

$$q_i(k+1) = q_i(k) \frac{(\mathbf{P}_{GTM}\mathbf{q}(k))_i}{\mathbf{q}(k)^T \mathbf{P}_{GTM} \mathbf{q}(k)}, \tag{9}$$

where q_i represents the element in the ith row of $\mathbf{q}$ and k is the iteration number. After n_{GTM} iterations, a correspondence c_i is identified as inlier if its corresponding q_i is higher than a threshold t_{GTM}.

E. Universal RANSAC (USAC) [14]

USAC integrates a universal framework for RANSAC, where each original step is optimized by referring to the advantages of previous parametric approaches such as PROSAC [13], SPRT test [47] and LO-RANSAC [36]. Further, this algorithm inserts degeneracy and local optimization processes after generating the minimal-sample model.

During the sampling step, USAC uses a weighted sampling algorithm named PROSAC [13], where the initial correspondences are reordered at first based on the descending sort order of brute-force matching scores and correspondences with higher scores are preserved. At the checking stage of the model (homography matrix or fundamental matrix), a correspondence is judged as an inlier by Eq. (3) with the threshold $t_{\mathbf{H}}$ or the epipolar geometry constraint with the threshold $t_{\mathbf{F}}$. After generating the minimal-sample model, USAC verifies whether the model is interesting by the SPRT test [47]. The likelihood ratio can be computed after evaluating n correspondences as

$$\xi_n = \prod_{i=1}^{n} \frac{p(r_i|\mathbf{H}_b)}{p(r_i|\mathbf{H}_g)}, \tag{10}$$

where $\mathbf{H}_g$ and $\mathbf{H}_b$ respectively represent a "good" model and a "bad" model, r_i is equal to 1 if c_i is consistent with the generated model and 0 otherwise, $p(1|\mathbf{H}_g)$ is approximated by the inlier ratio and $p(r_i|\mathbf{H}_b)$ follows a Bernoulli distribution. If the ξ_n is higher than an adaptive threshold, the model will be discarded. When fitting the fundamental matrix by epipolar geometry constraint, USAC utilizes DEGENSAC [48] for degeneracy. Eventually, USAC adds a local optimization (LO-RANSAC [36]) to refine the minimal-sample model.

F. Vector Field Consensus (VFC) [11]

VFC interpolates a vector field where the posteriori probability of a correct correspondence is estimated by the Bayes rule. For a correspondence c_i, the transformation to a motion field is expressed as $(\mathbf{x}_i,\ \mathbf{x}_i') \rightarrow (\mathbf{u}_i,\ \mathbf{v}_i)$, where $\mathbf{u}_i = \mathbf{x}_i$ and $\mathbf{v}_i = \mathbf{x}_i' - \mathbf{x}_i$. In this motion field, VFC holds the assumption that the noise around inliers indicated by $z_i = 1$ follows the Gaussian distribution and the noise around outliers indicated by $z_i = 0$ follows the uniform distribution. Thus, the probability is a mixture model given by

$$p(\mathcal{U}|\mathcal{V},\theta) = \prod_{i=1}^{N} \left(\frac{\gamma}{(2\pi\sigma^2)^{D/2}} e^{-\frac{\|\mathbf{v}_i - \mathbf{f}_{vfc}(\mathbf{u}_i)\|_2}{2\sigma^2}} + \frac{1-\gamma}{a}\right), \tag{11}$$

where $\theta = \left\{\mathbf{f}_{vfc}, \sigma^2, \gamma\right\}$ is a set of unknown parameters, $\mathbf{f}_{vfc}$ is the vector field expected to be recovered, γ is the mixing coefficient of the mixture probability model, i.e, $p(z_i = 1) = \gamma$, $\mathcal{U}$ and $\mathcal{V}$ respectively are sets of $\mathbf{u}$ and $\mathbf{v}$, σ is the uniform standard deviation of Gaussian distribution, $\frac{1}{a}$ is the probability density of the uniform distribution and D is the dimension of the output space. VFC employs the EM [49] algorithm to deal with the maximum likelihood estimation with latent variables. The E-step and M-step are repeated until parameters are converged. Finally, the inlier set is generated as

$$\mathcal{C}_{inlier} = \left\{c_i : p_i > t_{vfc}\right\}, \tag{12}$$

where t_{vfc} is a predefined threshold.

G. Grid-Based Motion Statistics (GMS) [21]

GMS proves that besides feature descriptiveness, feature number also contributes to the quality of correspondences. It supposes 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. In over-large neighborhoods, regions are divided into multiple small region pairs where distributions of correspondence number are approximated by Binomial distributions. Given a correspondence c_i, the joint statistical distribution is modeled as

$$S_i \sim \begin{cases} B(Kn, p_t), & \text{if } c_i \text{ is inlier} \\ B(Kn, p_f), & \text{otherwise,} \end{cases} \tag{13}$$

where S_i is the total number of correspondences in a region pair (a, b) around c_i, K is the quantity of small region pairs, p_t is the probability that the nearest neighbor of each keypoint in a is located in b under the condition that a and b view the same location, and p_f is the probability provided that a and b view the different locations. p_t and p_f can be estimated by

$$p_t = \delta + (1-\delta)\zeta m/M, \tag{14}$$

and

$$p_f = \zeta(1-\delta)(m/M), \tag{15}$$

where δ is the probability of a correspondence being correct, m is the amount of keypoints in region b, M is the size of $\mathcal{K}'$ in I', and ζ is a factor added to balance deviations caused by repeated structures. A quantitative score is next designed to evaluate the distinction between two distributions as

$$P = \frac{m_t - m_f}{s_t + s_f}, \tag{16}$$

where m is the mean value and s is the standard deviation. This equation can be simplified as

$$P \propto \sqrt{Kn}, \tag{17}$$

where the distinction is positive correlated to the number of correspondences.

In addition, to incorporate this approach into a real-time system, a fast gird-based score estimator is developed as follows. First, I and I' are divided into 20×20 non-overlapping cells. Second, for each cell in I, the cell containing the maximum amount of correspondences is grouped in I'. Third, in cell-pair (i, j) as well as its small neighborhoods (eight cell-pairs), S_{ij} is estimated as

$$S_{ij} = \sum_{k=1}^{K=9} \left|\chi_{i^k j^k}\right|, \tag{18}$$

where $|\chi|$ is the amount of correspondences in the cell-pair (i^k, j^k). All correspondences in (i, j) are judged as inliers if $S_{ij} > t_{gms}$, where t_{gms} is a threshold approximated by $\alpha\sqrt{n_i}$ with α being a given parameter and n_i being the average (of the nine cell-pairs) amount of correspondences.

H. Locality Preserving Matching (LPM) [19]

This algorithm removes mismatches by enforcing local geometric structure constraints. With the hypothesis that the local structure around a correspondence may not change freely, a cost function is defined as

$$L(\mathcal{W}, \lambda_{lpm}) = \sum_{i=1}^{N} w_i \left(l_i' - \lambda_{lpm}\right) + \lambda_{lpm} N, \tag{19}$$

where λ_{lpm} is a regularization parameter, $\mathcal{W}$ is a set of indicators that $w_i = 1$ indicates the inlier and $w_i = 0$ otherwise, N is the size of $\mathcal{K}$, and

$$l_i' = \sum_{j|\mathbf{x}_j \in \mathcal{K}_{\mathbf{x}_i}} d\left(\mathbf{x}_i', \mathbf{x}_j'\right) + \sum_{j|\mathbf{x}_j' \in \mathcal{K}_{\mathbf{x}_i'}'} d\left(\mathbf{x}_i, \mathbf{x}_j\right) \tag{20}$$

is a constraint term measuring the local geometric structure changes, where $\mathcal{K}_{\mathbf{x}_i}$ is the set of the k nearest neighbors of $\mathbf{x}_i$ ($\mathcal{K}_{\mathbf{x}_i'}'$ in the same way), and d indicates whether $\mathbf{x}_j$ (as an exemplar) is one of the k nearest neighbors of $\mathbf{x}_i$. With the objective of minimizing the cost function, a correspondence with the cost (i.e., $l_i > \lambda_{lpm}$) turns negative. Accordingly, the correct correspondence set is determined by

$$w_i = \begin{cases} 1, & \text{if } l_i' \leq \lambda_{lpm} \\ 0, & \text{otherwise,} \end{cases} \quad i = 1, \ \ldots \ , N. \tag{21}$$

To summarize, we note that the six non-parametric methods can be interpreted under a unified energy minimization framework. The key difference among them lies in the choice of mathematical formulation (e.g., graph theoretic vs Bayesian) and the definition of cost or objective functions.

IV. New Contribution: From Combination to Concatenation

In the literature, the idea of combining classifiers [50], [51] has been extensively studied. However, correspondence selection differs from traditional pattern classification in that the solution space is not characterized by a scalar (e.g., binary decision or matching score) but a collection of vectors. We have found it is easier to implement feature-level instead of decision-level combination. For instance, SUM-rule based combination accumulates the output indexes (i.e., 0 vs. 1) of all correspondence selection algorithms; the correspondence decision is made by comparing the accumulated result against the predefined threshold. A more flexible strategy is selectively combine a subset (usually the top-ranked ones) of the competing algorithms.

In this work, we propose a novel approach toward fusing different methods which we call *concatenation*. Unlike combination (conceptually similar to parallel processing), we suggest an alternative concatenation strategy (analogous to serial processing). The key idea is to concatenate a non-parametric algorithm with a parametric algorithm (so the former serves the purpose of preprocessing stage to the latter). Specifically, NNSR-RS (as an exemplar) combines NNSR and RANSAC, where NNSR plays the role of preselecting some candidates from the initial correspondences; and RANSAC is then performed in the candidate correspondences. In this manner, the generated parametric transformation is consequently utilized to extract the final inlier subset from the initial correspondence set. As demonstrated by our experimental results later, we have found that newly-proposed concatenation strategy often outperforms the conventional combination strategy for image feature correspondence selection.

TABLE II
Summary of Four Experimental Datasets

Dataset	Challenges	Matching pairs
VGG [28]	Zoom, rotation, blur, viewpoint change, light change and JPEG compression	40
Symbench [29]	Light change, different rendering styles	46
Heinly [24]	Zoom and rotation	29
AdelaideRMF [30]	Multiple structures, viewpoint change	38

V. Experimental Setup

A. Implementation Details

In our experiments, Hessian-affine detector [40] and SIFT descriptor [8] (a popular detector-descriptor combination [26]) are used as the default for image keypoint detection and description. Note that Hessian-affine detector is also required by the GTM method; other different combinations of detector-descriptor are considered in Sec. VI-E. The initial correspondence set $\mathcal{C}$ is generated by brute-force matching- i.e., greedy comparison of two feature sets. Detailed comparisons among parameter settings and implementation platforms for eight competing algorithms are listed in Table I. For NNSR and GTM methods, we set t_{nnsr} and t_{gtm} adaptively using the OTSU [45] algorithm to reduce the thresholding errors. All benchmark methods are implemented under OpenCV or MATLAB and tested on a standard desktop with a 3.2GHz processor and 8GB memory.

B. Benchmark Datasets

We conduct our experiments on four datasets- i.e., VGG [28], Symbench [29], Heinly [24], and AdelaideRMF [30]. A brief summary of dataset characteristics (e.g., challenging factors and database size) are shown in Table II. Exemplar images of datasets can be found in the supplementary material.

1) The VGG Dataset [28]: VGG is a hybrid dataset involving eight scenes. Each scene consists of six images with the first image being the reference one with respect to the others. Challenges including blur, viewpoint change, zoom, rotation, light change, and JPEG compression exist in this dataset. The ground-truth is the homography matrix $\mathbf{H}$, indicating that the transformation between two images on each scene satisfies the plane homographic constraint.

2) The Symbench Dataset [29]: The Symbench dataset is composed of 46 image pairs. Each pair includes the same object with light change or different rendering styles. The homographic transformation $\mathbf{H}$ of each image pair is given as the ground-truth.

3) The Heinly Dataset [24]: The Heinly dataset comprises images with dense or sparse viewpoint change, illumination, pure large-scale zoom or rotation. Considering that nuisances of viewpoint change and illumination have been covered in the other three datasets, we choose a subset of Heinly containing 29 pairs of image shot on 4 scenes with the specific challenges, i.e., pure zoom or rotation, to perform a more targeted test. The ground-truth is provided as the homographic transformation.

4) The AdelaideRMF Dataset [30]: AdelaideRMF includes 38 pairs of image with viewpoint change and multiple structures. The keypoint coordinates of initial correspondences are provided and the ground-truth correspondences are manually labeled in this dataset.

Motivations of employing these datasets can be summarized as: (i) The eight scenes in the VGG dataset cover a peculiar wide range of interferences such as the rigid/non-rigid transformation and image quality variation. Both the generality to diverse conditions and the robustness to a specific nuisance can be assessed on this dataset. (ii) The focus of Symbench is the image quality variation caused by light change and different rendering styles that give rise to potential errors of feature detection and description. The performance in the context of image quality variation can be specifically evaluated. (iii) The subset of Heinly is selected with the aim of testing the performance under the condition of a geometrical structure deformation (pure zoom or rotation). (iv) AdelaideRMF aims at evaluating the performance of those correspondence selection algorithms where plane homographic constraint fails and multiple consistent correspondence sets are involved due to multiple structures. All above peculiarities make the evaluation benchmarks complementary to each other and allow us to find the most appropriate algorithms for a specific nuisance.

C. Performance Evaluation

The performance of eight algorithms is evaluated by precision, recall and F-measure as in [19], [21], [32]. First, we denote the selected correspondence set, the ground-truth correspondence set and the correct subset in the selected correspondence set as $\mathcal{C}_{inlier}$, $\mathcal{C}_{inlier}^{GT}$ and $\mathcal{C}_{inlier}^{correct}$, respectively. Then, the precision, recall and F-measure are respectively defined as

$$\text{Precision} = \frac{|\mathcal{C}_{inlier}^{correct}|}{|\mathcal{C}_{inlier}|}, \tag{22}$$

$$\text{Recall} = \frac{|\mathcal{C}_{inlier}^{correct}|}{|\mathcal{C}_{inlier}^{GT}|}, \tag{23}$$

and

$$\text{F-measure} = \frac{2\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}, \tag{24}$$

where $|\cdot|$ denotes the cardinality of a set. A correspondence $c = \{\mathbf{x}, \mathbf{x}'\}$ belongs to $\mathcal{C}_{inlier}^{GT}$ if

$$\|\mathbf{x}_i' - \rho\left(\mathbf{H}_{gt}\begin{bmatrix}\mathbf{x}_i \\ 1\end{bmatrix}\right)\|_2 \leq t_{gt}, \tag{25}$$

where $\mathbf{H}_{gt}$ is the ground-truth homography matrix and t_{gt} is a threshold set to $10pix$ (pix being the unit of pixel) that controls the upper bound of the accuracy of a true inlier in our experiments.

Similarly, a correspondence belonging to $\mathcal{C}_{inlier}^{correct}$ is defined by

$$\|\mathbf{x}_i' - \rho\left(\mathbf{H}_{gt}\begin{bmatrix}\mathbf{x}_i \\ 1\end{bmatrix}\right)\|_2 \leq \tau \tag{26}$$

with τ being the matching tolerance. We vary τ from $1pix$ to t_{gt} with an interval of $1pix$ in order to generate the curves like previous works [21], [32].

D. Experimental Protocols

Our experimental protocols are designed as follows. In Sec. VI-A, the overall performance of the evaluated algorithms on four different datasets is tested. In Sec. VI-B, the robustness to different nuisances (i.e., blur, viewpoint change, zoom, rotation, light change, and JPEG compression) is independently examined on the VGG dataset. In Sec. VI-C, we address concerns about the efficiency in those algorithms by examining their overall time cost on different datasets paired with the speed comparison under different scales of initial matches. In Sec. VI-D, the performance with preselected correspondences by NNSR (i.e., commonly employed to improve the inlier ratio of initial matches [11], [27], [52], [53]) is tested on four datasets. In Sec. VI-E, different detector-descriptor combinations are considered to examine the performance variation of correspondence selection algorithms. Note that different combinations of detector and descriptor are desired in different application domains [26], [40] and will result in different distributions and inlier ratios. Finally, representative visual results of the evaluated algorithms and comparison results between combination and concatenation are shown in Sec. VI-F.

VI. Experimental Results

A. Performance on Different Datasets

In the following, we show the overall precision, recall and F-measure performance of our evaluated algorithms on four datasets (please refer to Fig. 2 for an overview).

1) Performance on the VGG Dataset: Fig. 2 (a) shows outcomes on the VGG dataset. It is interesting to see that NNSR achieves the best precision performance, being marginally better than USAC, RANSAC and GMS. This result is due to the fact that the feature distinctiveness cue is rather selective with rich-textured images, e.g., images in the VGG dataset. On the down side, feature distinctiveness is sometimes ambiguous and not a robust constraint as we can see that the recall of NNSR is just mediocre. It indicates that many correct correspondences have been filtered by NNSR. For ST and LPM, they are generally inferior to the others on this dataset in terms of the F-measure. That is because ST may fail to locate the main cluster in the spectral domain if the ourlier ratio is large, resulting in quite poor recall performance. LPM achieves much better recall performance than ST, while its precision performance is surpassed by most compared ones. It arises from the loose constraint employed in LPM. Overall, USAC is the best method on this dataset. Explanation behind is that USAC is a parametric method and the parametric model of each image pair existed in this dataset can be properly fitted.

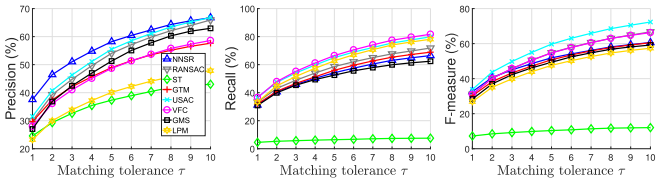
2) Performance on the Symbench Dataset: Fig. 2(b) presents results on the Symbench dataset. All methods suffer a clear drop in performance on this dataset when compared with that on the VGG dataset, which is attributed to light change and various rendering styles. More specifically, we observed

that the average inlier ratio of initial correspondences on this dataset is lower than 10%. As previously explained, the feature distinctiveness constraint strongly relies on the discriminative power of the local feature descriptor. However, the rendering style variation makes it fairly challenging to maintain descriptiveness in this case. As a result, NNSR delivers very poor precision performance. Another significant difference compared to that on the VGG dataset is USAC's performance. One can see that USAC returns the most and the second most inferior precision and recall performance, respectively. That is because USAC may find empty inlier sets in some cases when its average estimated scores decreases owing to the multiple constraints in this algorithm [14]. In general, GMS and VFC are the two most well-behaved methods after referring their F-measure rankings. A common trait of these two algorithms is that both of them are independent from the descriptor similarity.

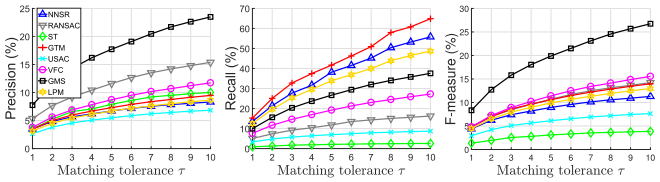
3) Performance on the Heinly Dataset: Fig. 2 (c) presents results on the Heinly dataset. Image pairs on this dataset only contain pure zoom or rotation, and we can observe that all methods obtain relatively decent performance on this dataset. In terms of precision, NNSR and RANSAC neatly outperform the others. Regarding recall, LPM and RANSAC are the two best ones. Note that the reason for the high recall of LPM is that most inliers are selected with the loose constraint designed by this algorithm. For NNSR and RANSAC, the former one is attributed to the high distinctiveness of SIFT (we will see its performance variation with less distinctive descriptors in Sect. VI-E), whereas the latter one is owing to the powerful homography fitting ability of RANSAC. GMS, due to its sensitivity to large degrees of rotation [21], shows worse results compared to its performance on the VGG and Symbench datasets.

4) Performance on the AdelaideRMF Dataset: Fig. 2(d) presents results on the AdelaideRMF dataset. Two explanations should be given on this dataset. First, as only manual labeled ground-truth correspondences are available, we present the exact scores rather than curves with respect to matching tolerance for each method. Second, the keypoints on this dataset are not located by image detectors. Rather, they were labeled manually. Thus, GTM requiring local affine information and NNSR based on auto-detected keypoints are not assessed on this dataset. Since each scene in this dataset contains multiple planes, the fundamental matrix based on the epipolar geometry constraint is employed to approximate the parametric model for RANSAC and USAC. By observing the scores in Fig. 2 (d), one can see that GMS, LPM and VFC achieve the best precision, recall and F-measure performance, respectively. All the three methods are non-parametric. This is reasonable since the AdelaideRMF contains multiple structures, and the parametric assumption for methods like RANSAC and USAC will fail in this case.

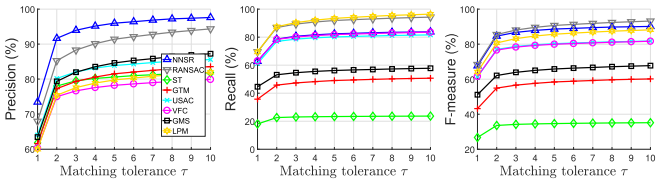
5) Overall Performance: By weighing up the results presented in Fig. 2, we can draw the following conclusions. First, the performance of all correspondences selection algorithms is affected by the initial inlier ratio. For instance, the performance of all algorithms deteriorates dramatically on the Symbench dataset with less than 10% inliers. Second, NNSR simply



(a) VGG



(b) Symbench



(c) Heinly

	RANSAC	ST	USAC	VFC	GMS	LPM
Precision (%)	93.84	77.32	96.28	97.59	**98.54**	81.21
Recall (%)	68.77	21.31	79.13	87.36	79.22	**99.62**
F-measure (%)	77.03	32.18	84.89	**91.63**	87.25	88.91

(d) AdelaideRMF

Fig. 2. Performance of the evaluated algorithms on four datasets, i.e., (a) VGG, (b) Symbench, (c) Heinly, and (d) AdelaideRMF, in terms of precision, recall and F-measure under different matching tolerance τ. The maximum values of precision, recall and F-measure are shown in bold face on the AdelaideRMF dataset.

relying on feature's distinctiveness produces pleasurable results if images are well-textured and clean. Third, parametric approaches, i.e., RANSAC and USAC, prefer the context that the transformation between two images can be well fitted by a parametric model. While non-parametric algorithms perform better in situations without large degrees of rigid/non-rigid transformation. Overall, VFC and RANSAC are the two best algorithms under across-dataset experiments. A more detailed view that illustrates F-measure scores for each image pair on the four datasets can be found in the supplementary material.

B. Robustness

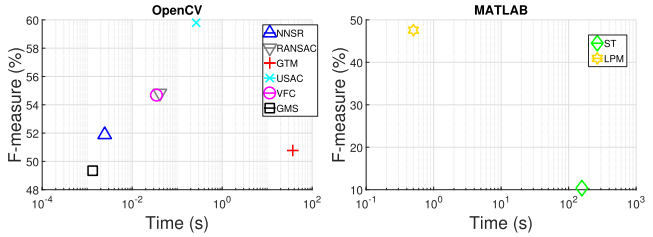
In this section, we independently evaluate the robustness of these algorithms to a specific nuisance, e.g., zoom, rotation, blur, viewpoint change, light change and JPEG compression on the VGG dataset. Results are listed in Table III. Some exemplar images with different nuisances are exhibited in the supplementary material.

Under zoom and rotation (case1 and case3), USAC and RANSAC, i.e., two parametric methods, behave the best (F-measure is referred) mainly attributed to that zoom and rotation are faint impact on homography fitting. Under blur (case2 and case6), GMS and NNSR outperform others. GMS is independent from feature similarity constraint, thus making it rational. For NNSR, it is still explicable as SIFT is very robust to blur. Regarding viewpoint change (case4 and case8),

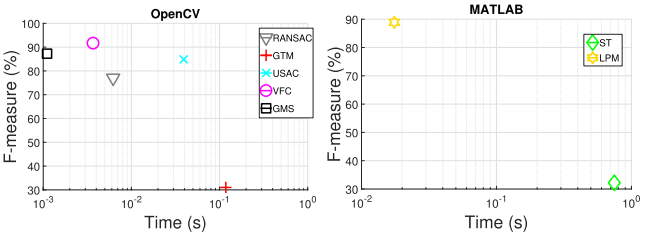
TABLE III

ROBUSTNESS RESULTS OF EVALUATED ALGORITHMS AGAINST DIFFERENT NUISANCES WITH $\tau = 5$. THE BEST RESULT IS EXPRESSED IN BOLD FACE

		NNSR	RANSAC	ST	GTM	UASC	VFC	GMS	LPM
Case1	P(%)	**81.16**	76.11	17.98	43.22	77.38	67.19	63.61	42.50
	R(%)	77.68	92.86	4.51	79.60	**99.05**	86.11	11.45	83.54
	F(%)	77.35	82.42	6.56	53.69	**84.48**	74.27	18.46	54.75
Case2	P(%)	**74.57**	36.87	44.23	67.00	49.66	29.44	41.71	27.73
	R(%)	**79.39**	41.85	8.23	56.74	60.00	51.41	50.45	54.75
	F(%)	**71.87**	38.71	13.46	61.12	54.30	35.27	45.54	35.86
Case3	P(%)	61.53	**70.54**	15.97	44.92	67.41	49.38	58.57	44.59
	R(%)	57.91	83.28	1.97	52.16	79.95	**99.22**	57.21	76.43
	F(%)	53.74	**74.81**	3.50	44.83	73.12	61.91	56.35	55.10
Case4	P(%)	51.77	55.58	37.21	50.94	**63.01**	57.86	57.05	45.08
	R(%)	61.63	66.38	3.52	68.55	79.73	**97.08**	75.52	83.97
	F(%)	51.75	58.56	6.41	55.69	70.23	**71.23**	64.55	56.56
Case5	P(%)	76.28	81.44	61.90	68.90	**83.76**	71.99	64.89	57.65
	R(%)	63.75	86.97	6.76	80.35	**100**	**100**	87.95	84.46
	F(%)	68.00	82.34	11.61	73.94	**91.11**	82.49	74.37	67.90
Case6	P(%)	31.90	45.33	24.95	33.45	32.23	31.18	**57.10**	26.72
	R(%)	**69.13**	27.06	2.57	39.10	40.00	40.00	47.00	66.81
	F(%)	31.49	28.86	4.34	34.29	35.68	35.02	**50.80**	35.82
Case7	P(%)	89.47	87.07	89.46	80.66	**89.59**	89.48	79.87	75.87
	R(%)	61.17	97.41	28.59	94.42	**100**	**100**	96.70	93.38
	F(%)	72.42	91.81	43.07	86.88	**94.43**	94.26	87.25	83.43
Case8	P(%)	67.27	74.42	52.34	72.05	73.03	72.40	**80.86**	62.67
	R(%)	61.08	79.03	4.02	80.12	80.00	79.51	73.10	**81.76**
	F(%)	58.39	76.23	7.36	73.42	**76.33**	75.74	76.08	69.64



(a) VGG



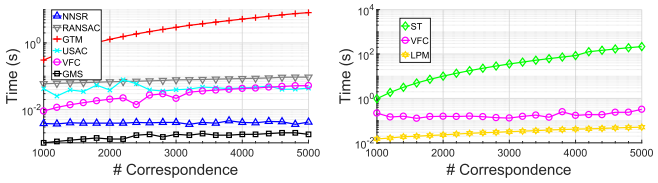
(b) AdelaideRMF

Fig. 3. Efficiency *v.s.* F-measure plots on the (a) VGG and (b) AdelaideRMF datasets. The efficiency-axis is shown logarithmically for clarity. The methods implemented in OpenCV and MATLAB are separately compared.

USAC and VFC are the best methods. Note that VFC generally delivers good performance under all kinds of nuisances, being benefited from the consensus search in the non-parametric field. USAC also achieves the best performance under light change (case5) and JPEG compression (case7), being the one that is robust to the broadest categories of nuisances.

C. Efficiency

To provide an overview of the evaluated methods by taking both selection performance and efficiency into consideration, we present the efficiency *v.s.* F-measure plots on the four experimental datasets in Fig. 3. Owing to fast execution speed and overall decent performance, USAC strikes a good balance between selection performance and efficiency.



(a) OpenCV (b) MATLAB

Fig. 4. Speed comparison of evaluated algorithms with respect to different numbers of initial correspondences. (a) and (b) present the results of methods implemented in OpenCV and MATLAB, respectively. To give a better comparison, VFC is implemented in both platforms. The time-axis is shown logarithmically for clarity.

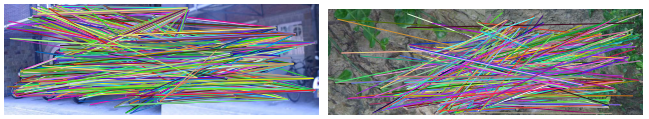
In order to further test an algorithms' efficiency regarding different numbers of initial correspondences, i.e., the number of initial correspondences may vary in different applications or with different feature detectors, we vary the amount of initial correspondences from 1000 to 5000 and record the average speed of the eight methods. This experiment has been repeated for 10 rounds and average statistics are retained. Because codes of these algorithms are implemented either in OpenCV (C++) or MATLAB, we assess methods within the same platform independently. In addition, the VFC method is evaluated on both platforms and can be a reference for comparing across-platform methods. Results are reported in Fig. 4.

For methods implemented in OpenCV, the efficiency of GMS is beyond all others. That is because GMS involves a grid framework for fast scoring. NNSR ranks the second, as only sort operation is needed to rank correspondences. RANSAC is slightly slower than USAC, and the core time consumption of both methods is dedicated to hypothesis generation-verification. GTM, with the computational complexity of $O(n^2)$ (n being the number of input correspondences), is significantly slower than the other five methods. The margin is rather significant as the number of correspondences increases. For methods implemented in MATLAB, LPM is very efficient as it relies on a simple yet efficient strategy by preserving local neighborhood structure. ST is the most inefficient method, being slower than others by tens of magnitude with dense correspondences. It is due to the fact that the time consumption for computing eigenvalues increases exponentially with the size of the affinity matrix.

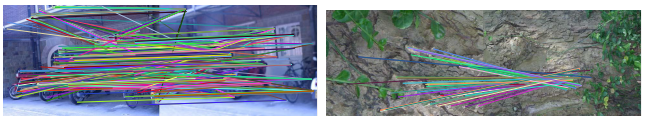
D. Performance on Selected Matches

Many existing works [11], [27], [52], [53] first prune false correspondences via NNSR and then use parametric or non-parametric methods to for further selection. This experiment then checks this scenario. Remarkably, since NNSR fails to work on the AdelaideRMF dataset, this dataset is not considered in this test. Fig. 5 shows the difference between correspondences before and after applying NNSR, and results using NNSR-selected correspondences for selection are shown in Fig. 6.

On the VGG dataset shown in Fig. 6(a), one can see that the performance of all methods has been improved using NNSR-selected matches compared to brute-force matches in Fig. 2(a). Particularly, USAC manages to be the best method regarding

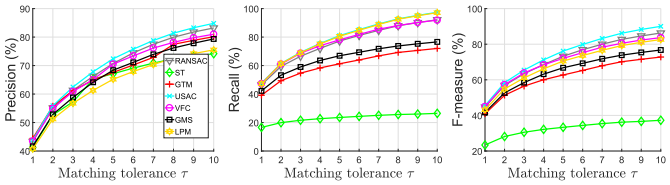


(a) Brute force matching



(b) NNSR-selected matching

Fig. 5. Examples of correspondence sets obtained via brute-force matching (a) and NNSR selection (b). Sample image pairs are taken from the VGG dataset.



(a)

	RANSAC	ST	GTM	USAC	VFC	GMS	LPM
Symbench	17.26	5.06	18.46	14.84	22.37	**26.59**	18.04
Heinly	97.12	51.20	58.41	87.45	96.16	70.45	**97.27**

(b)

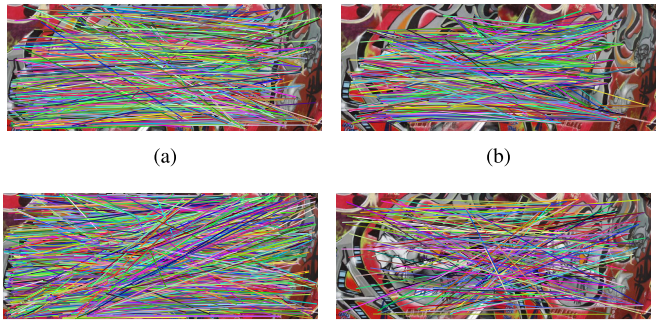
Fig. 6. Performance of evaluated algorithms (except NNSR) on selected matches on the (a) VGG, (b) Symbench and Heinly datasets. For aggregated view, precision, recall and F-measure curves are shown for the VGG dataset, and F-measure (%) performance under $\tau = 5$ is shown for the Symbench and Heinly datasets. The AdelaideRMF dataset is not tested as NNSR fails to work on this dataset.

precision, recall and F-measure. Also, gaps between most curves excluding that of ST are relatively small. On the Symbench and Heinly datasets, GMS and LPM respectively achieve the best overall performance, where LPM even produces an extremely high F-measure score, i.e., 97.27%, on the Heinly dataset. We can infer that LPM adapts well to initial correspondence sets with high inlier ratio.

E. Performance Under Different Detectors and Descriptors

In addition to Hessian-affine + SIFT, we also consider four other popular detector-descriptor combinations, i.e., SIFT + SIFT [8], ORB + ORB [9], ASIFT + ASIFT [54], and BLOB [55] + FREAK [56]. Fig. 7 shows the initial correspondences with these combinations on a sample image pair. Note that GTM is excluded in this test as it requires local affine information and these detectors do not provide this information. Also, the AdelaideRMF dataset is not considered due to human-labeled keypoints. The results are reported in Fig. 8 and Table IV.

A common characteristic of these results is that the best correspondence selection algorithm generally varies with combinations of detector and descriptor. While we can still find some consistencies, e.g., the VFC method achieves pleasurable performance on the VGG dataset in spite of the descriptor-detector combinations. The performance of some methods



(a) (b) (c) (d)

Fig. 7. Initial correspondences using (a) SIFT + SIFT, (b) ORB + ORB, (c) ASIFT + ASIFT and (d) BLOB + FREAK on an exemplar image pair taken from the VGG dataset.

TABLE IV
F-MEASURE (%) PERFORMANCE UNDER DIFFERENT GROUPS OF DETECTOR AND DESCRIPTOR ON THE SYMBENCH AND HEINLY DATASETS WITH $\tau = 5$

		NNSR	RANSAC	ST	USAC	VFC	GMS	LPM
SIFT + SIFT	Symbench	9.34	3.02	1.22	3.27	11.56	**11.64**	9.36
	Heinly	94.13	95.43	33.82	**98.75**	83.08	40.29	89.84
ORB + ORB	Symbench	4.70	5.27	2.13	3.33	3.00	**11.62**	6.31
	Heinly	57.62	58.98	17.57	56.45	56.30	50.24	**60.30**
ASIFT + ASIFT	Symbench	7.00	7.15	3.29	4.54	14.42	**17.48**	12.54
	Heinly	69.31	**92.31**	27.47	78.72	78.75	44.21	88.21
BLOB + FREAK	Symbench	4.62	2.35	0.95	1.97	**6.25**	0.50	2.19
	Heinly	68.63	**76.25**	20.32	74.45	68.71	4.30	66.64

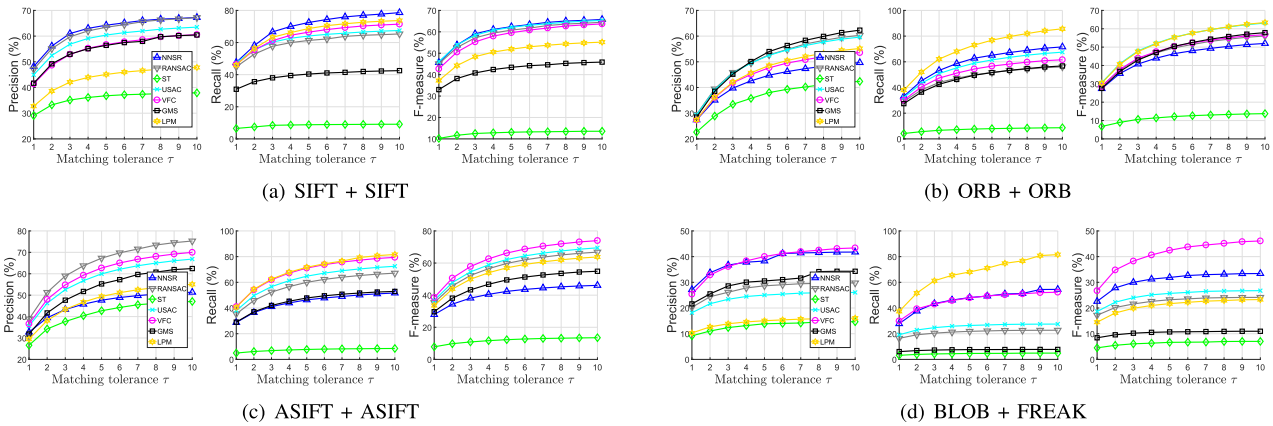
fluctuates dramatically. For example, NNSR ranks the first with SIFT + SIFT while performs poorly using ASIFT + ASIFT on the VGG dataset. On the Symbench and Heinly datasets, GMS and RANSAC are two prominent methods under different kinds of detector-descriptor combinations.

F. Visual Comparison and Fusion Results

To obtain a qualitative sense of outputs of evaluated algorithms, we present several visual results of these algorithms on the four experimental datasets in Fig. 9.

Two main observations can be made from the figure. First, distributions of selected correspondences by different algorithms are generally different from each other. For instance, few correspondences are found by GTM on the *bread* in Fig. 9(d). However, NNSR and LPM get plenty of correspondences on it. Second, the quantity of selected correspondences also varies with different methods. In particular, LPM manages to return dense correspondences on most datasets, while ST seeks out much less than others.

Finally, we report our experimental results with combination and concatenation strategies of fusing different correspondence selection algorithms. Fig. 10 illustrates the evaluation results of these fusion methods on VGG dataset. SUM denotes standard sum-based combination, and RANKED-SUM accumulates the indexes of USAC, VFC, RANSAC, and NNSR which achieve better F-measure. As shown in Fig. 10, the combination methods with proper τ and the concatenation methods achieve superior performance compared with the baselines, -i.e., the single selection approaches, which confirms the superiority of



(a) SIFT + SIFT (b) ORB + ORB

(c) ASIFT + ASIFT (d) BLOB + FREAK

Fig. 8. Performance of evaluated algorithms on the VGG dataset using four different detector-descriptor combinations, i.e., (a) SIFT + SIFT, (b) ORB + ORB, (c) ASIFT + ASIFT, and (d) BLOB + FREAK.

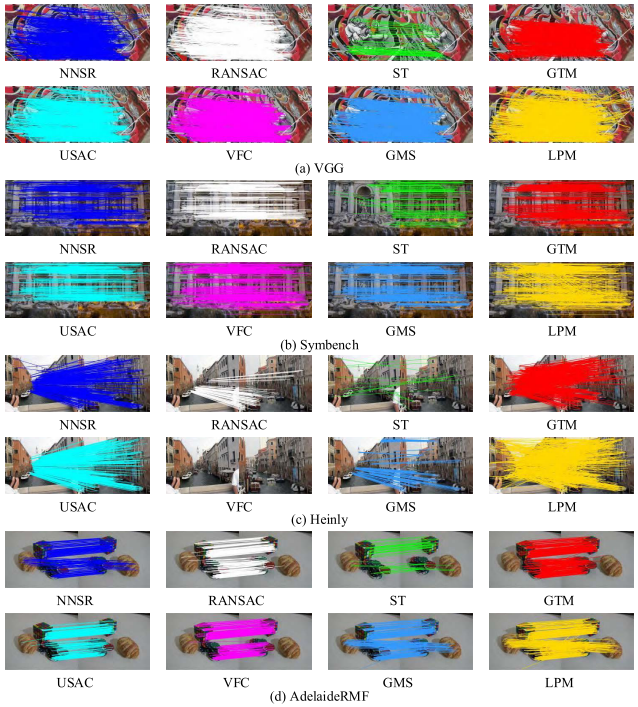


Fig. 9. Visual results of evaluated algorithms on examplar image pairs respectively taken from the (a) VGG, (b) Symbench, (c) Heinly and (d) AdelaideRMF datasets. For the best view, lines with different colors represent results of different algorithms.

fusion strategies. Moreover, the concatenation methods outperform the combination methods in most cases, verifying the strengths of individual approaches are more effectively leveraged. As aforementioned, parametric methods (e.g., RANSAC) are sensitive to the inlier ratio in the initial correspondence set, so the performance is able to be remarkably improved as a preselection is incorporated which significantly boosts the inlier ratio in a candidate correspondence set.

VII. Summary and Discussion

To facilitate the decision in practical applications, we provide a "user manual" for image feature correspondence selection in Table V. Some tips and lessons associated with each evaluated algorithm are presented as follows:

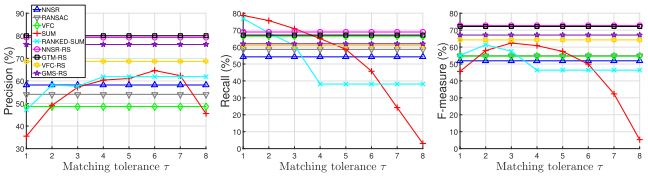


Fig. 10. Performance comparison of fusion methods and the baselines - i.e., single evaluated methods on VGG dataset, where SUM accumulates the output indexes (i.e., 0 vs. 1) of all evaluated approaches, and RANKED-SUM accumulates the indexes of USAC, VFC, RANSAC, and NNSR that achieve better F-measure. The matches are determined as inliers if the corresponding accumulated indexes are higher than τ. NNSR-RS (as an example) represents a fusion method that combines NNSR and RANSAC, where RANSAC is performed in a subset of candidate correspondences preselected by NNSR, and the estimated parametric transformation is consequently employed to select the final inlier subset from the initial correspondence set.

TABLE V

Summary of Correspondence Selection Algorithm Comparison in Different Scenarios Based on the Evaluation Results. Note That This Conclusion Is Drawn Upon the F-measure, i.e., the Aggregate Performance Regarding Both Precision and Recall. Keypoint Detector and Descriptor Are Abbreviated to Det and Des, Respectively

Scenarios		Superior methods	Inferior methods
Datasets	VGG	USAC, RANSAC, VFC	ST
	Symbench	GMS	ST, USAC
	Heinly	RANSAC, NNSR, LPM	ST, GTM, GMS
	AdelaideRMF	VFC, LPM	ST, RANSAC
Pre-selection	VGG	USAC, RANSAC, LPM	ST
	Symbench	GMS, VFC	ST
	Heinly	LPM, RANSAC, VFC	ST, GMS
Det/Des groups	SIFT+SIFT	USAC, NNSR, GMS	ST, RANSAC
	ORB+ORB	LPM, GMS, USAC	ST, VFC
	ASIFT+ASIFT	VFC, RANSAC, GMS	ST, USAC, NNSR
	BLOB+FREAK	VFC, NNSR, RANSAC	ST, GMS, USAC
Robustness	Zoom and rotation	USAC, RANSAC	ST, GTM, GMS
	Blur	NNSR, GMS	ST, RANSAC
	Viewpoint change	USAC, VFC	ST, NNSR
	Light change	USAC, VFC	ST
	JPEG compression	USAC, VFC	ST, NNSR
Efficiency		GMS, NNSR	ST, GTM

- **NNSR** is arguably the most straightforward strategy to select correspondences. Its key strength is that repeatable patterns can be removed reliably in certain circumstances, provided that its employed feature detectors can locate the keypoints accurately and descriptors possess strong discriminative power, e.g., SIFT. Also, the high execution

speed makes it suitable for real-time or near real-time systems. However, the limitation of NNSR is obvious because of the simple descriptor similarity constraint. It is vulnerable when image quality is low (e.g., facing with light change, blur, exposure, and style-transfer) and texture information is limited.

- **RANSAC** and **USAC**, i.e., two evaluated parametric approaches, can fit the parametric models including the homography and fundamental matrices between two images effectively, with the premise that the image pair has homography or epipolar geometry constraint. Thus, they are prior options in such circumstances. Nevertheless, such assumption also brings drawbacks, e.g., when non-rigid objects are captured in images with large scale of parallax or the pure rotation between two camera positions, resulting in the failure of RANSAC and USAC. Further, the reliable models may not be generated by limited iterations with high outlier ratios, which will give rise to expensive time cost. For RANSAC, the minimal-sample models sometimes fall into the local optimization. USAC optimizes over RANSAC, though, it does not guarantee convergence and may produce an empty inlier set due to strict constraints.
- **ST** and **GTM** are methods relying on the affinity matrix computed from initial matches. We can find that these two methods are relatively time-consuming, especially for the ST method. The performance of GTM is much better than ST, mainly because GTM employs local affine information to judge the compatibility of two correspondences. While ST is based on rigid constraint. ST, when inputted with high-quality correspondences, is able to achieve high precision performance (as verified in Sect. VI-D). These two methods are optional for off-line applications desiring high precision and with high-quality input.
- **LPM** rejects outliers by the local structure consistency. The constraint item in LPM is relatively loose, resulting in high recall yet relatively low precision. LPM prefers scenarios where the geometric structure information is well preserved between the same local pattern in the image pairs, e.g., small degrees of rigid transformations. Similar to NNSR, it relies strongly on the discriminative power of the feature descriptor. In other words, retrieving the local consistency can be problematic if the local region contains too few inliers. We therefore suggest to choose LPM in the context that has well preserved geometric structures and requires dense correspondences.
- **VFC**, as revealed by our experiment, is the most robust method under all tested scenarios. This is attributed to the fact that VFC is independent from the feature similarity and parametric models. Specifically, it performs inlier selection in a vector field. VFC generalizes well under different application contexts and can cope with various kinds of nuisances, especially for viewpoint change, light change and JPEG compression.
- **GMS**, similar to VFC, is also independent from the feature similarity and parametric models. However, it assumes that the motion between two images is smooth. Accordingly, it behaves unsatisfactory for image pairs undergoing large degrees of rotation. While if the motion smoothness assumption holds, its performance is superior even for correspondence set with very limited number of inlier, e.g., correspondences generated from the Sym-bench dataset. Another attractive merit of GMS is the ultra fast execution speed even under several thousands of initial correspondences, making it a prior selection for real-time applications.

VIII. Conclusions

This paper has comprehensively evaluated eight state-of-the-art image correspondence selection algorithms, covering both parametric and non-parametric families. The experiments addressed several critical issues regarding correspondence selection, e.g., different application scenarios (datasets), robustness under various challenging conditions including zoom, rotation, blur, viewpoint change, JPEG compression, light change, different rendering styles and multi-structures, efficiency, and inputs from different combinations of feature detector and descriptor. Advantages and limitations, in light of experimental outcomes, are summarized so as to guide developers to choose a proper algorithm given a specific scenario. According to the evaluation results that the best option varies in different circumstances, we suggest that fusing the selected results of different approaches is a promising solution, taking account of the generalization and robustness.

Remarkably, the performance of most existing algorithms changes dramatically in different scenarios and most methods fail to achieve satisfactory results when the inlier ratio of the initial correspondence set is low. Therefore we believe it is worth pursuing future research towards the development of correspondence selection algorithms with improved generality and robustness to low inlier rate. Concatenation-based fusion might be a promising strategy toward this direction; we will explore other ways of concatenation in future studies.

References

[1] 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, Nov. 2008.

[2] 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, Apr. 2005, pp. 943–948.

[3] 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.

[4] M. Brown and D. G. Lowe, "Automatic panoramic image stitching using invariant features," *Int. J. Comput. Vis.*, vol. 74, no. 1, pp. 59–73, Apr. 2007.

[5] J. Lu, J. Hu, and Y.-P. Tan, "Discriminative deep metric learning for face and kinship verification," *IEEE Trans. Image Process.*, vol. 26, no. 9, pp. 4269–4282, Sep. 2017.

[6] J. Lu, V. E. Liong, and J. Zhou, "Deep hashing for scalable image search," *IEEE Trans. Image Process.*, vol. 26, no. 5, pp. 2352–2367, May 2017.

[7] D. Lowe, "Object recognition from local scale-invariant features," in *Proc. 7th IEEE Int. Conf. Comput. Vis.*, vol. 2, Sep. 1999, pp. 1150–1157.

[8] D. G. Lowe, "Distinctive image features from scale–invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, Nov. 2004.

[9] 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, pp. 2564–2571.

[10] Y. Duan, J. Lu, Z. Wang, J. Feng, and J. Zhou, "Learning deep binary descriptor with multi-quantization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 1183–1192.

[11] 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.

[12] 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.

[13] O. Chum and J. Matas, "Matching with PROSAC—Progressive sample consensus," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2005, pp. 220–226.

[14] 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.

[15] M. Leordeanu and M. Hebert, "A spectral technique for correspondence problems using pairwise constraints," in *Proc. 10th IEEE Int. Conf. Comput. Vis. (ICCV)*, vol. 1, Oct. 2005, pp. 1482–1489.

[16] A. Albarelli, E. Rodolà, and A. Torsello, "Imposing semi–local geometric constraints for accurate correspondences selection in structure from motion: A game–theoretic perspective," *Int. J. Comput. Vis.*, vol. 97, no. 1, pp. 36–53, Mar. 2012.

[17] T. Collins, P. Mesejo, and A. Bartoli, "An analysis of errors in graph-based keypoint matching and proposed solutions," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2014, pp. 138–153.

[18] S. Khan, M. Nawaz, X. Guoxia, and H. Yan, "Image correspondence with CUR decomposition based graph completion and matching," *IEEE Trans. Circuits Syst. Video Technol.*, to be published.

[19] J. Ma, J. Zhao, H. Guo, J. Jiang, H. Zhou, and Y. Gao, "Locality preserving matching," in *Proc. Int. Joint Conf. Artif. Intell.*, 2017, pp. 4492–4498.

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

[21] 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.

[22] W.-Y. Lin *et al.*, "CODE: Coherence based decision boundaries for feature correspondence," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 1, pp. 34–47, Jan. 2018.

[23] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 10, pp. 1615–1630, Oct. 2005.

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

[25] H. Aanæs, A. Dahl, and K. S. Pedersen, "Interesting interest points: A comparative study of interest point performance on a unique data set," *Int. J. Comput. Vis.*, vol. 97, no. 1, pp. 18–35, 2012.

[26] P. Moreels and P. Perona, "Evaluation of features detectors and descriptors based on 3D objects," *Int. J. Comput. Vis.*, vol. 73, no. 3, pp. 263–284, Jul. 2007.

[27] R. Raguram, J.-M. Frahm, and M. Pollefeys, "A comparative analysis of RANSAC techniques leading to adaptive real–time random sample consensus," in *Proc. Eur. Conf. Comput. Vis.*, 2008, pp. 500–513.

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

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

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

[31] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 248–255.

[32] W.-Y.-D. Lin, M.-M. Cheng, J. Lu, H. Yang, M. N. Do, and P. Torr, "Bilateral functions for global motion modeling," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2014, pp. 341–356.

[33] T. Collins, P. Mesejo, and A. Bartoli, "An analysis of errors in graph–based keypoint matching and proposed solutions," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2014, pp. 138–153.

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

[35] P. 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.

[36] O. Chum, J. Matas, and J. Kittler, "Locally optimized RANSAC," in *Pattern Recognition*. Berlin, Germany: Springer, 2003, pp. 236–243.

[37] M. Cho, J. Lee, and K. Mu Lee, "Feature correspondence and deformable object matching via agglomerative correspondence clustering," in *Proc. IEEE 12th Int. Conf. Comput. Vis.*, Sep. 2009, pp. 1280–1287.

[38] T.-J. Chin, J. Yu, and D. Suter, "Accelerated hypothesis generation for multi-structure robust fitting," in *Proc. Eur. Conf. Comput. Vis.*, 2010, pp. 533–546.

[39] H.-Y. Chen, Y.-Y. Lin, and B.-Y. Chen, "Robust feature matching with alternate Hough and inverted Hough transforms," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 2762–2769.

[40] K. Mikolajczyk and K. Mikolajczyk, "Scale & affine invariant interest point detectors," *Int. J. Comput. Vis.*, vol. 60, no. 1, pp. 63–86, Oct. 2004.

[41] J. Kim and K. Grauman, "Boundary preserving dense local regions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2011, pp. 1553–1560.

[42] M. Cho, J. Lee, and K. M. Lee, "Reweighted random walks for graph matching," in *Proc. Eur. Conf. Comput. Vis.*, 2010, pp. 492–505.

[43] F. Tombari, S. Salti, and L. Di Stefano, "Performance evaluation of 3D keypoint detectors," *Int. J. Comput. Vis.*, vol. 102, nos. 1–3, pp. 198–220, Mar. 2013.

[44] Y. Guo, M. Bennamoun, F. Sohel, M. Lu, J. Wan, and N. M. Kwok, "A comprehensive performance evaluation of 3D local feature descriptors," *Int. J. Comput. Vis.*, vol. 116, no. 1, pp. 66–89, Jan. 2016.

[45] N. Otsu, "A threshold selection method from gray–level histograms," *IEEE Trans. Syst., Man, Cybern.*, vol. 9, no. 1, pp. 62–66, Jan. 1979.

[46] J. W. Weibull, *Evolutionary Game Theory*. Cambridge, MA, USA: MIT Press, 1997.

[47] J. Matas and O. Chum, "Randomized RANSAC with sequential probability ratio test," in *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 2, 2005, pp. 1727–1732.

[48] 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)*, vol. 1, Jul. 2005, pp. 772–779.

[49] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the em algorithm," *J. Roy. Stat. Soc.*, vol. 39, no. 1, pp. 1–38, 1977.

[50] J. Kittler, M. Hater, and R. P. W. Duin, "Combining classifiers," in *Proc. 13th Int. Conf. Pattern Recognit.*, vol. 2, Aug. 1996, pp. 897–901.

[51] L. I. Kuncheva, *Combining Pattern Classifiers: Methods and Algorithms*. Hoboken, NJ, USA: Wiley, 2014.

[52] J. Yang, Z. Cao, and Q. Zhang, "A fast and robust local descriptor for 3D point cloud registration," *Inf. Sci.*, vols. 346–347, pp. 163–179, Jun. 2016.

[53] J. Yang, Q. Zhang, and Z. Cao, "Multi-attribute statistics histograms for accurate and robust pairwise registration of range images," *Neurocomputing*, vol. 251, pp. 54–67, Aug. 2017.

[54] J.-M. Morel and G. Yu, "ASIFT: A new framework for fully affine invariant image comparison," *SIAM J. Imag. Sci.*, vol. 2, no. 2, pp. 438–469, Jan. 2009.

[55] T. Lindeberg, "Feature detection with automatic scale selection," *Int. J. Comput. Vis.*, vol. 30, no. 2, pp. 79–116, 1998.

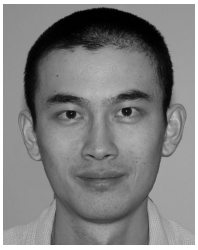
[56] A. Alahi, R. Ortiz, and P. Vandergheynst, "FREAK: Fast retina keypoint," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 510–517.

Chen Zhao received the B.S. degree from the Huazhong University of Science and Technology, Wuhan, China, in 2017, where he is currently pursuing the master's degree with the School of Artificial Intelligence and Automation. He visited the Lane Department of Computer Science and Electrical Engineering, West Virginia University, in 2019. His research focuses on local feature detection and description, feature matching, and point cloud analysis.

Zhiguo Cao received the B.S. and M.S. degrees in communication and information system from the University of Electronic Science and Technology of China and the Ph.D. degree in pattern recognition and intelligent system from the Huazhong University of Science and Technology. He is currently a Professor with the School of Automation, Huazhong University of Science and Technology. His research interests spread across image understanding and analysis, depth information extraction, 3-D video processing, motion detection, and human action analysis. His research results, which have over 50 articles at international journals and prominent conferences, have been applied to automatic observation system for crop growth in agricultural, for weather phenomenon in meteorology and for object recognition in video surveillance system based on computer vision.

Jiaqi Yang received the B.S. and Ph.D. degrees from the Huazhong University of Science and Technology, China, in 2014 and 2019, respectively. Sponsored by the China Scholarship Council, he visited the GRASP laboratory, University of Pennsylvania, from 2017 to 2018. He is currently an Associate Professor with the School of Computer Science, Northwestern Polytechnical University. His research interests include local geometric description, 3-D registration, 3-D feature matching, and 3-D object recognition.

Ke Xian received the B.S. degree from the Huazhong University of Science and Technology, Wuhan, China, in 2014, where he is currently pursuing the Ph.D. degree with the School of Artificial Intelligence and Automation. His research mainly centers on algorithms issues in dense prediction (e.g., depth estimation from single images and semantic image segmentation) and feature matching.



Xin Li (Fellow, IEEE) received the B.S. degree (Hons.) in electronic engineering and information science from the University of Science and Technology of China, Hefei, in 1996, and the Ph.D. degree in electrical engineering from Princeton University, Princeton, NJ, USA, in 2000. He was a Member of Technical Staff with Sharp Laboratories of America, Camas, WA, USA, from August 2000 to December 2002. Since January 2003, he has been a Faculty Member with the Lane Department of Computer Science and Electrical Engineering, West Virginia University. His research interests include image/video coding and processing. He received the Best Student Paper Award at the Conference of Visual Communications and Image Processing as the Junior Author in 2001, the Runner-Up Prize of the Best Student Paper Award at the IEEE Asilomar Conference on Signals, Systems and Computers as the Senior Author in 2006, and the Best Paper Award at the Conference of Visual Communications and Image Processing as the Single Author in 2010. He currently serves as an Associate Editor for the IEEE TRANSACTIONS ON IMAGE PROCESSING and a Senior Area Editor for the IEEE SIGNAL PROCESSING LETTERS.