IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received August 16, 2018, accepted September 18, 2018, date of publication September 24, 2018, date of current version October 19, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2871743

Adaptive Image Mismatch Removal With Vector Field Interpolation Based on Improved Regularization and Gaussian Kernel Function

YONGJUN ZHANG[1], XUNWEI XIE[1], XIANG WANG[1], YANSHENG LI[1], AND XIAO LING[2]
[1]School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, China
[2]Future Cities Laboratory, Singapore-ETH Centre, Singapore 138602

Corresponding author: Yongjun Zhang (zhangyj@whu.edu.cn)

This work was supported in part by the National Key Research and Development Program of China under Grant 2018YFB0505003 and in part by the National Natural Science Foundation of China under Grant 41571434.

ABSTRACT When the regularized kernel methods are utilized in the mismatch removal problem, the regularization coefficient and the choice of kernel function will seriously affect the performance of the methods. In this paper, we propose a method that combines an improved regularization and an adaptive Gaussian kernel function to interpolate the vector fields so as to overcome the issue. We formulated the problem as a modified maximum *a posterior* estimation of a Bayesian model. In this model, a two-order term of the regularization coefficient is introduced into the regularized risk function in order that the coefficient can be adaptively estimated in the expectation–maximization algorithm. In addition, an adaptive Gaussian kernel function also is imposed to construct the regularization, in which the width of the kernel function is adaptively determined by the diagonal length of the maximum enveloping rectangle of the sample set. Our experimental results verified that our method was robust to large outlier percentages and was slightly superior to some state-of-the-art methods in precision-recall tradeoff and efficiency. The evidence that the performance of our method was insensitive to the remaining inner parameters verified its good self-adaptability. Finally, airborne image pairs were used to demonstrate that our method can establish the feature correspondences even under a discontinuous vector field scene. In addition, we found that our method can obtain higher precision given a residual threshold for special applications such as robust epipolar geometry estimation in computer vision and photogrammetry.

INDEX TERMS Point correspondence, mismatch removal, regularization, Gaussian kernel function.

I. INTRODUCTION

Image mismatch removal is a prerequisite in applications including motion analysis, camera self-calibration, registration and object recognition between two images, but it continues to be a fundamental problem in photogrammetry and computer vision [1]–[5]. The initial correspondences are usually established by image matching methods (e.g., SIFT [6] and SURF [7]) that require the corresponding points can only possess similar descriptors and satisfy some underlying geometric constraints such as homography geometry, epipolar geometry and some non-parametric geometry relationships [8]. However, there are many mismatches in the initial correspondences due to occlusion, viewpoint change, etc., which may ruin the results of traditional estimation methods. A great deal of research effort therefore continues to be put forth by the photogrammetry and computer vision community to eliminate its influence.

A traditional common strategy for solving the matching problem is to use a two-stage process: 1) putative corresponding points computing, which includes many true matches but also a large number of mismatches and 2) mismatch removal, which is removal of the mismatches and estimation of the inliers and geometric constraint model [5], [8]–[11]. The RANSAC method [11] and its variants [12], [13] are the classical representatives of this strategy. Although these methods are very successfully applied to many situations, their results suffer from the following difficulties. First, the efficiency of these methods decreases dramatically with

an increase in the mismatch percentage [5]. Second, setting the adaptive parameters is difficult, which is detrimental to the application of these methods. In addition, Yuille and Grzywacz formulated the visual motion task in terms of finding those matches which give rise to the best vector field interpolation [14], [15]. On this basis, this theory is applied most often to shape matching, image registration, and machine learning research [16]–[19]. In terms of the correspondence problem, the results of the related works were sensitive to the parameter settings, such as VFC [8] and LLT [19].

In this paper, the mismatch removal problem is formulated as a modified maximum a posteriori estimation of a Bayesian model which can be solved by an iterative expectation-maximization (EM) [20] algorithm, whereby given a large initial value for the EM algorithm, all the latent variables reach the optimal solution with the iteration increases, which is similar to deterministic annealing [21]. Our method is adaptively and computationally attractive especially for its performance in dealing with a large mismatch percentage.

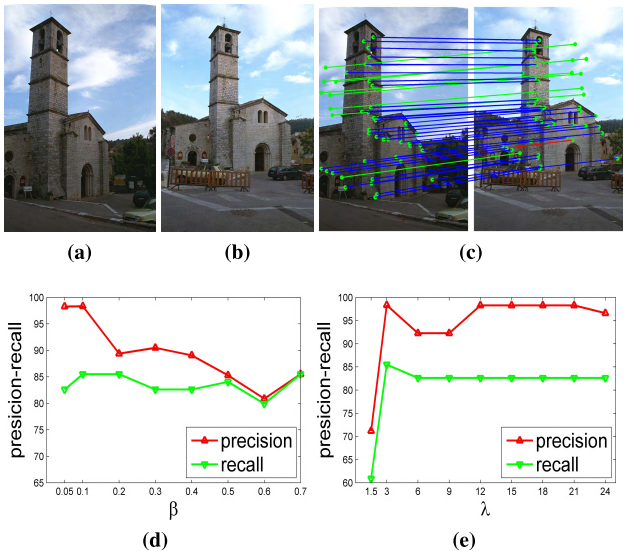


FIGURE 1. Influence of parameters on precision-recall of VFC. (a), (b) *Valbonne* dataset. (c) Mismatch removal result of our method. (d), (e) Influence of β and λ on the precision-recall using VFC. β is the parameter of Gaussian kernel function, and λ is the regularization coefficient in VFC. In this dataset, there is a large depth discontinuity, and in this case it may be problematic for the vector field interpolation-based mismatch removal methods as a smoothness constraint is imposed as a prior.

The mismatch removal methods based on Yuille's motion field theory often suffer from the regularization coefficient and the parameter choice of a kernel function as shown in Fig. 1. Given a set of putative correspondences matched from *Valbonne* dataset used in [8], as shown in Fig. 1(a) and Fig. 1(b), the VFC with different parameter settings was implemented with the results as shown in Fig. 1(d)-(e); and our method's mismatch removal result shown in Fig.1(c). Note that our result in Fig. 1(c) is the same as the optimal result of VFC with parameter settings - ($\beta = 0.1$, $\lambda = 3$). Different regularization coefficient and the kernel function can greatly influence the performance of VFC, which makes it less attractive for wide application. Interestingly, our experimental result demonstrated that our method exhibited the same best result that VFC can obtain with some specific (maybe optimal) parameter settings without loss of precision and recall. This leads to the core concerns of our method that how to determine regularization coefficient and the parameter of kernel function.

To make it easier for the readers to follow this paper, we firstly grouped the basic terminologies appeared in this paper as follows.

- ***Inliers*** *are the correct matches.*
- ***Outliers*** *are the mismatches.*
- ***Precision*** *is the ratio of the preserved correct match number and preserved correspondence number.*
- ***Recall*** *is the ratio of the preserved correct match number and correct match number in the original putative correspondences or ground truth inliers.*

The contributions in this paper include the following. First, a two order term of regularization coefficient is introduced in the regularized risk function, which can guarantee that the coefficient can be adaptively estimated in an EM algorithm. As the smooth term of our method is similar to that of regularization theory [22], the regularization coefficient can be considered as a regularization variable at each iteration process of the EM algorithm, which will be discussed later. Second, an adaptive Gaussian kernel function is imposed to construct regularization, where the width of the kernel function is adaptively determined by the diagonal length of the maximum enveloping rectangle of the sample set for the correspondence problem, avoiding the influence of the width of the Gaussian kernel on the performance of our method. In addition, we make explicit the calculation of the volume of uniform distribution in our method, whereas it is fuzzy and set as 10 by default in some similar methods. It must be emphasized that the adaptability of the method is the primary factor in our consideration under the premise of ensuring considerable precision-recall trade-off and efficiency.

The remainder of this paper is organized as follows. Section II describes the related literature and background materials. Section III describes the details of our method, which is not only and most importantly robust to large mismatches percentage, but also insensitive to the parameter settings, called ''inner parameters'' in this paper. Section IV illustrates the performance of our method on datasets as well as makes comparisons to the following state-of-the-art methods: RANSAC [11], MLESAC [13], LMedS [23], VFC, and LLT, which are the most widely used methods in photogrammetry and computer vision. Finally, our conclusions are discussed in Section V.

II. RELATED WORK

The background material on which our method is based is discussed in this section. Since methods for putative point

correspondences computing are included, some of the commonly used methods in photogrammetry and computer vision for mismatch removal are discussed as well. The section also concludes with a brief overview of vector field interpolation and regularization.

A. PUTATIVE POINT CORRESPONDENCES COMPUTING

Putative point correspondences are mostly obtained by image matching using feature-based methods that are the most popular with the automatic processing of photogrammetry and computer vision. The feature-based methods basically extract the salient features, calculate the descriptors, and consider the one with the most similar descriptor as a correspondence. The common feature-based methods include SIFT [6], SURF [7] and shape contexts [24] in the practice software. However, while the putative point correspondences obtained by these methods include not only most of the true matches (inliers), but they also have a large number of mismatches (outliers) due to the ambiguities in the similarity constraint [19].

B. COMMON METHODS FOR MISMATCH REMOVAL

The mismatches can be removed using geometrical constraints such as epipolar geometry and homography geometry. When combined with the above constraints, the mismatch removal methods can be divided into three types: statistical robust regression methods, resampling methods, and case diagnostic methods [5].

The statistical robust regression methods, such as LMedS [23] and M-estimators [25], try to alleviate the passive influence of mismatches by replacing the sum of the squared error criterion with ones those are less influenced by the outliers [5]. The hypothesis of LMedS was evaluated with the median residual of putative correspondences. In 1994, Deriche combined LMedS and an epipolar geometry constraint to reject mismatches [26]. In theory, LMedS only deals with the situation where the outlier percentage is less than 50%, which restricts its applicability. M-estimators, classical robust methods, minimize the sum of a symmetric and positive define function of the residuals, which suffer from the lack of a good initial estimation in order for the parameters to be estimated.

The main idea of the resampling methods, like RANSAC, as well as its variants, such as PROSAC [12] and MLESAC [13], make use of the following general hypothesis-and-verify flowchart: a parametric model is estimated from a minimum number of randomly selected putative correspondences; the quality of the model is estimated by some method; and the hypothesis with the highest score to reject the mismatches is chosen. RANSAC evaluates the hypothesis with the inliers whose residuals are below some given threshold. PROSAC assumes prior beliefs about the probability of a correspondence being an inlier to modify the sampling step of the original RANSAC. For a certain geometric estimation problem, a graph-cut algorithm in the local optimization step is applied when the so-far-the-best-model is found in RANSAC flowchart, which can be more geometrically accurate [27]. While MLESAC utilizes an M-estimation-based weighted voting strategy and accepts the solution that maximizes the likelihood rather than the inlier count as the final optimal estimation [13]. A similar vote-and-verify strategy also can be found in [28], and it achieved a verification accuracy similar to some hypothesis-and-verify approaches. In theory, most of these methods are capable of rejecting mismatches from the correspondences with large outlier percentages. However, when the outlier percentage increases, the computational efficiency dramatically decreases. In addition to being strongly influenced by the inlier percentage, such methods are also affected by the number of model parameters to be estimated as well as by some given threshold such as that used in RANSAC.

The case diagnostic methods are also classical robust mismatch removal methods [25], [29]. The applications of these methods can be found in [30] and [31]. Such methods are very successful for situations where a small number of outliers are present. However, when there are many mismatches, these methods often suffer from the masking effect [32], which considers some mismatches invisible by others, and the swamping effect [33], which considers the correct matches as mismatches. In addition, when the mismatch percentage increases, these methods also suffer from a very heavy computational load because of their greedy search scheme. Therefore, initial mismatch removal is needed when many outliers are present.

Thanks to developments in minimal mapping theory [34] and motion coherence theory [14], [15], a new strategy for mismatch removal is to formulate it as a correspondence matrix between point correspondences together with a parametric constraint, or a non-parametric constraint [16], [35], [36] in terms of vector field interpolation and the choice of kernel function. Building on Yuille and Grzywacz 's theory [14], [15], Rangarajan established a general framework for the correspondence estimation problem. Inspired by this, J. Ma solved for correspondences by interpolating a motion field between two point sets whose inliers follow a nonparametric geometric constraint [8]. They then extended the nonparametric geometric constraint to a local geometric parametric constraint that can preserve local structures among neighboring feature points, which not only improved the method but also made it robust to a large number of mismatches [19]. However, setting the parameters for these two methods in order to adaptively obtain the best results is not an easy task.

C. VECTOR FIELD INTERPOLATION AND REGULARIZATION

The vector field interpolation-based mismatch methods were built on Yuille's computational theory for the perception of coherent visual motion [14], [15], [19], [42]. The theory is comprised of two stages: measuring and smoothing [15]. On this basis, many methods have been developed in this context [37] combining with regularization theory. Yuille and Grzywacz introduced the motion coherence theory for computing the velocity field defined in an image by using a

quadratic regularizer to impose geometric constraints on the correspondences [14]. They also demonstrated that this was equivalent to formulating the problem in terms of a space of kernels. The relationships between the components of the vector fields can be directly encoded by a series of operations in the Reproducing Kernel Hilbert Space (RKHS) [38], [39] associated with a certain choice of regularization so as to obtain a meaningful solution. The two most well-known regularization methods are regularized least-squares [40] and support vector machines [41] which were developed for rejecting mismatches in [5]. The regularized least-squares, which require that the number of functions is equivalent to the training size, has large time and space complexities, leading to a significant computational burden on large dataset. On this basis, J. Ma also proposed a spare approximation to the vector field learning for the mismatch removal problem [42]. The robust vector field interpolation and regularization technique were adopted in the Gaussian process by utilizing the t-progresses [43]. In [8], a new robust vector field interpolation, which associates each correspondence with a latent variable that determines if it is an inlier, obtained excellent performance for mismatch removal in rigid and non-rigid situations. Other field consensus and regularization methods for rejecting mismatches also can be found in [44]–[47]. However, almost all these methods suffer from the choice of a suitable kernel function and sensitive parameters, which can seriously affect the precision of vector field interpolation. In this paper, we introduce an improved regularized risk function and impose an adaptive Gaussian kernel function to address the adaptive issues.

III. METHODOLOGY

This section describes our adaptive algorithm for image mismatch removal. We start by briefly introducing the vector field interpolation problem via an improved regularization and then discuss a mixture likelihood model for mismatch removal in which a modified slow-and-smooth constraint is introduced on the prior of the vector fields. We use an iterative EM algorithm next to solve a modified maximum a posteriori to determine the latent variables in the mixture likelihood model. Thereafter, we introduce how to design an adaptive Gaussian kernel function to construct regularization. Finally, the parameter settings and computational complexity are described to conclude this section.

A. VECTOR FIELD INTERPOLATION VIA AN IMPROVED REGULARIZATION

The problem of vector field interpolation in an image is to fit a mapping vector f which interpolates a given sample set $S = \{(u_n, v_n) : n \subseteq N_N\}$, i.e. $\forall n \subseteq N_N$, $v_n = f(N_N)$, where u_n is each normalized position in the left image; v_n is the normalized motion field sample by a transformation $v_n = \hat{v}_n - u_n$ ($\hat{v}_n$ is each normalized position in the right image). The purpose of data normalization is to control the influence of the point coordinate system on the performance of vector field interpolation. Generally, the vector field interpolation problem is ill-posed because of an infinite number of solutions [8]. As indicated above in the related literature review, the vector field interpolation problem can be formulated into a problem with some regularization which operates in a RKHS. Specifically, we introduce a modified regularized risk function inspired by the Tikhonov regularization in a RKHS as follows

$$\varepsilon(f) = \min\{\sum_{n=1}^{N} \|v_n - f(u_n)\|^2 + \lambda\|f\|^2_{RKHS} + \xi\}$$
$$\xi = -\lambda^2, \quad f \subseteq RKHS \tag{1}$$

where the first term is the empirical risk which enforces closeness to the data; the second term is a stabilizer which enforces smoothness to the vector field f; λ is a regularization coefficient which controls the trade-off between the first two terms; ξ is a two order term of the coefficient which also enforces smoothness to the vector field; and $\|\cdot\|_{RKHS}$ denotes the norm of RKHS.

In our method, parameter λ is considered as a constant in each iteration process of the EM algorithm, which is discussed later. It is clear that parameter ξ will not affect the solution of the modified regularized risk functional (1). Therefore, according to [39], the solution of Eq. (1) is given by

$$f(u) = \sum_{n=1}^{N} \Gamma(u, u_n)c_n \tag{2}$$

with the set $\{c_n\}$ determined by the following system

$$(\widetilde{\Gamma} + \lambda I)\,\widetilde{C} = \widetilde{V} \tag{3}$$

where $\widetilde{\Gamma}$ is the Gram matrix with the (i, j) block $\Gamma(u_i, u_j)$, I is an identity matrix, and $\widetilde{V} = (v_1, \ldots, v_n)^T$ and $\widetilde{C} = (c_1, \ldots, c_n)^T$ are column vectors.

In this paper, we chose a Gaussian radial basis function kernel in the regularization for the correspondence problem, whose performance was verified by [19] and [45]. We will discuss how to adaptively determine the width of the Gaussian kernel later in this paper.

B. MIXTURE LIKELIHOOD MODEL

In order to robustly estimate vector field f, we followed the common assumption that the noise is isotropic Gaussian with zero mean and covariance σ^2 for the inliers; and the outliers distribution is uniform with the volume a. A latent variable $z_n \subseteq \{1, 0\}$ is associated with the n^{th} sample correspondence, where $z_n = 1$ indicates the sample is an inlier, otherwise it is an outlier. Let U and V be the inputs and outputs, in which the n^{th} rows represent u_n and v_n. Thus, the mixture likelihood model of the Gaussian and uniform distributions [13] has the

following form:

$$P(V|U,\theta) = \prod_{n=1}^{N}\sum_{z_n} p(v_n, z_n|u_n,\theta) = \prod_{n=1}^{N}[\frac{\gamma e^{-\frac{\|v_n - f(u_n)\|^2}{2\sigma^2}}}{2\pi\sigma^2} + \frac{1-\gamma}{a}] \quad (4)$$

where $\theta = (f, \sigma, \gamma)$ includes a set of unknown parameters that should be determined and γ represents the mixing coefficient specifying the marginal distribution over the latent variable (i.e., $p(z_n = 1) = \gamma$).

In order to solve Eq. (4), we assume vector field f to be with a prior probability distribution $p(f)$ and impose it on f. On the basis of a slow-and-smooth model [48] which is mostly applied to the motion phenomena, the prior of f is modified as the following form:

$$p(f) = e^{-[\frac{\lambda}{2}\psi(f)+\xi]}, \xi = -\lambda^2 \quad (5)$$

where $\psi(f)$ is a smoothness term and λ is a positive real number. Note that Eq. (5) is much like Eq. (1) in terms of form. (The reason for which will be discussed later.)

Unlike the estimation of the maximum a posterior solution of θ in [8] and [42], we estimated a modified maximum a posterior solution of θ, as a new prior probability distribution is imposed on the vector field. Its optimal solution is $\theta^* = \arg\max_\theta p(V|U,\theta)p(f)$, which is equivalent to minimizing the negative log-likelihood function:

$$E(\theta) = -\sum_{n=1}^{N}\ln\sum_{z_n} p(v_n, z_n|u_n,\theta) - \ln p(f) \quad (6)$$

Then, the vector field f is determined from the optimal solution θ^*. We demonstrate how to solve Eq. (6) in the next subsection.

C. ITERATIVE EM ALGORITHM

In this paper, we use an iterative EM algorithm that is a common means to deal with the latent variables to solve Eq. (6), including expectation step (E-step) and maximization step (M-step).

Because we could not use the complete-data log likelihood directly, we considered its expectation under the posterior distribution of the latent variable in the E-step and maximized the expectation in the M-step [49]. This expectation is denoted as $\Omega(\theta, \theta^{old})$. The complete-data negative log posterior, which omitted some constant terms, is given by

$$\begin{aligned}\Omega(\theta,\theta^{old}) &= -\frac{1}{2\sigma^2}\sum_{n=1}^{N} p_n\|v_n - f(u_n)\|^2 - \ln\sigma^2\sum_{n=1}^{N} p_n \\ &+ \ln\gamma\sum_{n=1}^{N} p_n + \ln(1-\gamma)\sum_{n=1}^{N}(1-p_n) - \frac{\lambda}{2}\psi(f) + \lambda^2\end{aligned} \quad (7)$$

where $p_n = P(z_n = 1|u_n, v_n, \theta^{old})$ is a posterior probability which determines how the sample (u_n, v_n) fits an inlier.

E-step: Denote a diagonal matrix $P = diag(p_1, \ldots, p_n)$, where p_n can be computed by using Bayes rule, as in

$$p_n = \frac{\gamma}{2\pi\sigma^2}e^{-\frac{\|v_n - f(u_n)\|^2}{2\sigma^2}} / (\frac{\gamma}{2\pi\sigma^2}e^{-\frac{\|v_n - f(u_n)\|^2}{2\sigma^2}} + \frac{1-\gamma}{a}) \quad (8)$$

M-step: We re-estimated the parameter set θ^{new} by maximizing the function [49]: $\theta^{new} = \arg\max_\theta \Omega(\theta, \theta^{old})$. Taking the first-order derivatives of Ω with respect to σ^2, λ, and γ, and setting them to zero, we obtained the following expressions,

$$\begin{cases}\sigma^2 = \dfrac{(V-F)^T P(V-F)}{2\cdot trace(P)} \\ \gamma = \dfrac{trace(P)}{N} \\ \lambda = \dfrac{1}{4}\psi(f)\end{cases} \quad (9)$$

where $F = (f(u_1)^T, \ldots, f(u_n)^T)^T$. In order to estimate Eq. (9), the mapping of f should be estimated firstly in this step.

We considered the terms of $\Omega(\theta)$ with respect to the prior probability distribution $p(f)$, and the modified regularized risk functional (i.e. Eq. (1)) was obtained as follows

$$\varepsilon(f) = \frac{1}{2\sigma^2}\sum_{n=1}^{N} p_n\|v_n - f(u_n)\|^2 + \frac{\lambda}{2}\psi(f) - \lambda^2 \quad (10)$$

where the first term is a weighted empirical error; the second and third term together control the trade-off with respect to the first term, and they are together the denoted smoothness term.

As λ is a constant in every iterative M step, it does not affect the extreme of Eq. (10), so we were able to obtain the coefficient set $\{c_n\}$ by solving the linear system (The proof is given in [19]):

$$(\widetilde{\Gamma} + \lambda\sigma^2 P^{-1})\,\widetilde{C} = \widetilde{V} \quad (11)$$

We solved λ in Eq. (9) as follows. The regularized risk functional (i.e. Eq. (10)) satisfies

$$\begin{aligned}\varepsilon(f) &= \frac{1}{2\sigma^2}\sum_{n=1}^{N} p_n\|v_n - f(u_n)\|^2 + \frac{\lambda}{2}\|f_N + f_{1/N}\|^2 \\ &\geq \frac{1}{2\sigma^2}\sum_{n=1}^{N} p_n\|v_n - f(u_n)\|^2 + \frac{\lambda}{2}\|f_N\|^2\end{aligned} \quad (12)$$

where $f(u_n) = f_N(u_N)$, $\|f_N + f_{1/N}\|^2_{RKHS} = \|f_N\|^2_{RKHS} + \|f_{1/N}\|^2_{RKHS}$, $\|f_N\|^2_{RKHS} = \widetilde{C}^T\widetilde{\Gamma}\widetilde{C}$, which means that $\psi(f) = \|f_N\|^2_{RKHS} + \|f_{1/N}\|^2_{RKHS}$. The reader is referred to [8] for the details of these equations, i.e. $f_N, f_{1/N}$.

We imposed $\psi(f)$ into Eq. (7), and then obtained

$$\Omega(\theta, \theta^{old}) \leq -\frac{1}{2\sigma^2}\sum_{n=1}^{N} p_n \|v_n - f(u_n)\|^2 - \ln\sigma^2 \sum_{n=1}^{N} p_n + \ln\gamma \sum_{n=1}^{N} p_n + \ln(1-\gamma)\sum_{n=1}^{N}(1-p_n) - \frac{\lambda}{2}\|f_N\|^2_{RKHS} + \lambda^2 \quad (13)$$

Eq. (13) indicates the maximum expectation of Eq. (7) with respect to the prior $p(f)$ coming from f_N and λ. Therefore, Eq. (13) is equivalent to Eq. (7) in terms of maximizing the expectation. Taking the derivative of Eq. (13) with respect to λ, and setting it to zero, we obtained

$$\lambda = \frac{1}{4}\|f_N\|^2_{RKHS} = \frac{1}{4}\tilde{C}^T \overset{\sim}{\Gamma}\tilde{C} \quad (14)$$

Eq. (14) is our reason for introducing the improved regularized risk function as Eq. (1) and our motivation for imposing the modified slow-and-smooth model constraint as Eq. (5).

After the iterative EM algorithm converges, the inlier set S_I can be obtained according to the following criterion:

$$S_I = \{(u_n, v_n) : p_n > \tau\} \quad (15)$$

where τ is a present threshold, and is insensitive to its choice on the later experiments.

From the above process, parameters σ^2, λ, γ were adaptively estimated in the iterative EM algorithm. However, mapping f was highly correlated with a kernel function Γ, which was very detrimental to adaptively realizing mismatch removal. In the next subsection, we discuss how to determine an adaptive kernel function.

D. ADAPTIVE GAUSSIAN KERNEL FUNCTION

The adaptive estimation of Eq. (14) depends on the choice of a kernel function. We chose a Gaussian radial basis function kernel (RBF) for the correspondence problem because of its simple form and good performance, as in

$$\Gamma(x, x') = e^{\left(-\frac{\|x-x'\|_2^2}{2\bar{\sigma}^2}\right)} \quad (16)$$

where $\bar{\sigma}$ is the width of the kernel function, which controls the radial scope of the function; x is a multidimensional vector; and x' is the center vector of RBF.

According to the representor theorem [39], the optimal solution in a RKHS is a linear combination of a number of basis functions as the form (2), which means that the kernel function is determined by the distributed points in the left image for our method. Fig. 2 below shows how we subsequently used an adaptive calculation method of $\bar{\sigma}$. In Fig. 2(a), the red double arrow represents the diagonal of the image, while the blue one represents the diagonal of the maximal enveloping rectangle of all the normalized point set. Therefore, the maximal scope of the kernel function is the diagonal length of the left image intuitively (e.g., the length of the red double arrow in Fig.2(a)). Considering its square form, $\bar{\sigma}^2$ was calculated as:

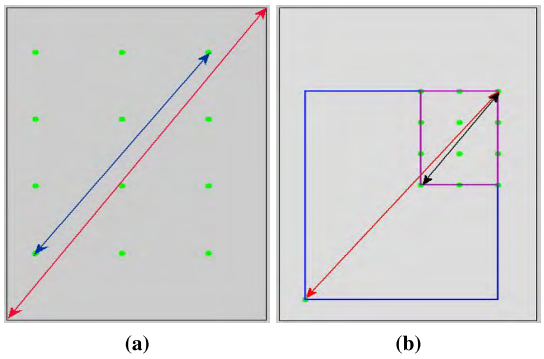


FIGURE 2. Schematic diagram for determining the width-$\bar{\sigma}$ of RBF. (a) Determine $\bar{\sigma}$ by maximal enveloping rectangle of all the normalized point set. (b) Determine $\bar{\sigma}$ by maximal enveloping rectangle of the normalized sample point set. The gray rectangle represents the image range, and the green points represent the normalized point set in the left image.

$$\bar{\sigma}^2 = w_x^2 + h_y^2 \quad (17)$$

where w_x, h_y represent the normalized width and height of the left image, respectively. Considering the correspondence problem, we utilized the diagonal length of the maximal enveloping rectangle of the normalized point set in the left image as the width of the kernel (e.g., the length of the blue double arrow in Fig.2(a)) as:

$$\bar{\sigma}^2 = \max \|u_i - u_j\|^2 \quad (18)$$

Since solving vector field f suffers from heavy computational burdens, it was overcome by a spare approximation and a suboptimal solution searching with much less basis functions in the RKHS. We found that random sampling of the training inputs (the normalized points in the left image) did not reduce the performance of the vector field interpolation-based methods [19], [42]. In order to avoid the impact of the isolated points on the width of RBF, we utilized a simple training method to determine optimal $\bar{\sigma}$ based on the fact that parameter $\bar{\sigma}$ should be uniquely determined by the maximal enveloping rectangle of the sampling points excluding the isolated points. Likewise, as shown in Fig. 2(b), the blue rectangle represents maximal enveloping rectangle of the normalized sample point set, from which we can see that it is more reasonable to determine $\bar{\sigma}$ according to the orange rectangle because an isolated sample point in left corner pulls up its value intuitively. The concrete implementation means was as follows:

First, we randomly selected M sample points in the left image (M was fixed as 16 for the correspondence problem in this paper).

Second, we calculated the current maximum square distance of each pair of sample points, as in

$$\bar{\sigma}_k^2 = \max \|u_i - u_j\|^2, k < maxT \quad i, j < M \quad (19)$$

where, k represents the current training; and $maxT$ represents the max training number.

Third, if $k < maxT$, return to the first step; otherwise, a set of $\bar{\sigma}^2$ can be obtained, as in

$$S_{\bar{\sigma}} = \left\{\bar{\sigma}_i^2, \ldots, \bar{\sigma}_{maxT}^2\right\} \tag{20}$$

Finally, we excluded the 5% maximum of set $S_{\bar{\sigma}}$ and chose the maximum of the remaining set to determine $\bar{\sigma}$, as in

$$\bar{\sigma}^2 = \max S_{\bar{\sigma}}^{0.95} \tag{21}$$

where, $S_{\bar{\sigma}}^{0.95}$ represents the remaining set after excluding the 5% maximum sample.

Next, we conducted the adaptive Gaussian kernel function selection through Eq. (18) or Eq. (21). We recommend Eq. (21) that was used in our later experiments to calculate the width of the kernel function.

The volume of uniform, parameter-a, used in Eq. (4) also was determined by the maximal enveloping rectangle of the normalized point set. Since residual term $v_n - f(u_n)$ used in Eq. (4) is either positive or negative, the volume of uniform was calculated as

$$a = 2\bar{\sigma} \tag{22}$$

which means that constant a was two times the diagonal of the maximal enveloping rectangle of all the normalized point sets in an image. It is worth noting that the calculation of constant a in this paper was different from the similar vector field interpolation-based methods, such as VFC and LLT in which it was set as a default of 10.

E. PARAMETER SETTINGS AND INITIALIZATION

A good adaptive mismatch removal method should not contain the parameters that are known to affect its performance. There are two fixed parameters in our method: τ, which is the posterior probability of a sample being an inlier, and $maxT$, which is the maximum number used in the simple training method to determine the width of RBF. We found that, in general, our method was insensitive to these two parameters and thus they were selected for $\tau = 0.7$ and $maxT = 100$ throughout this paper.

In the iterative EM algorithm, there are two free parameters that should be initialized: σ^2, which is the covariance of the Gaussian distribution for inliers, and γ, which is the mixing coefficient specifying the marginal distribution over the latent variable. σ^2 is given a large initial value-$\bar{\sigma}^2$ which was adaptively determined in the previous steps, and γ was initialized as $\gamma = 0.5$. Our method also was not influenced by the initialization of parameter γ.

Our adaptive mismatch removal method is summarized in **Algorithm 1** to facilitate understanding and programming for the readers.

F. COMPLEXITY ANALYSIS

We also use a spare approximation matrix and a suboptimal solution searching with much less basis functions to solve vector field f so as to overcome heavy computational burdens [19], [44]. The corresponding Γ matrix of the proposed method is $M \times M$ in size, so the time complexity of estimating the mapping f by solving a linear equation (11) is $O(NM^2)$. And the total time complexity of the iterative EM algorithm is $O(mNM^2)$, where m is the iterative number of the EM algorithm. In addition, as the width of Gaussian kernel function should be determined in the pretreatment, the time complexity of this process is $maxT \cdot M^2$. As a result, the total time complexity is $O(mNM^2 + maxT \cdot M^2)$. As storing the Γ matrix in the pretreatment, the space complexity is $O(M^2)$. For the matrix $\widetilde{\Gamma}$ with the sparse approximation, the space scales is reduced to $O(NM)$ from (N^2). Therefore, the total space complexity is $O(M^2 + NM)$. For a large sample set, as $M(16) \ll N$, the space complexity is reduced to $O(NM)$.

Algorithm 1 Adaptive Image Mismatch Removal

Input:
Normalized correspondences $S = \{(u_n, v_n) : n \subseteq N_N\}$, parameter-$\tau$, parameter-$maxT$ (optional);

Output:
Inliers S_I;

Pretreatment:
Calculate $\bar{\sigma}$ using Eq. (21) or Eq. (17) or Eq. (18) (We recommend Eq. (21).);
Calculate RBF by Eq. (16);
Calculate a by (22);

Initialization:
$F = 0_{2N}$, $diag(P_{N\times N})$, $\lambda = \bar{\sigma}^2$; Initialize σ and γ using Eq. (9), and construct $\widetilde{\Gamma}$;
repeat
E-Step
Update P by Eq. (8)
M-Step
Update $\widetilde{\Gamma}$ by solving Eq. (11);
Update F by using Eq. (12);
Update σ^2, γ and λ by Eq. (9) and Eq. (14)
until Eq. (13) converges
return Inliers S_I calcuate by Eq. (15)

IV. EXPERIMENTS AND DISCUSSIONS

The performance of our method was assessed from the following aspects. First, the robustness of our method was tested on datasets with different percentages of mismatches. Subsequently, the precision-recall and efficiency of our method were evaluated in a comparison with the following state-of-the-art methods: RANSAC, MLESAC, LMedS, VFC and LLT. Finally, the influence of the inner parameters on the performance of our method also was analyzed in experiments, followed by further experiments on high resolution airborne image pairs to conclude this section. All the putative point correspondences were computed by SIFT matching.

A. EXPERIMENTS ON ROBUSTNESS

The robustness of our method was tested on the putative point correspondences with different percentage of mismatches. The test data come from the *VGG* affine benchmark that contains five different changes in imaging conditions: image blur, illumination, viewpiont changes, rotation-scale changes and JPEG compression[1]. The images are either of planar scenes or the camera position is fixed during acquisition, thus they obey homographies or plane projective transformation, and the ground truth homographies are supplied by the benchmark. From this benchmark, we chose three image pairs with different imaging conditions that contained image blur (*Bikes*), illumination (*Leuven*), and viewpoint change(*Graffiti*) respectively. In addition, we also chose an image pair, i.e. *Valbonne*, in which a large depth discontinuity exists as shown in Fig. 1(a)-(b). The ground truth inliers of *Valbonne* were provided by Ma *et al.* [8].

In order to determine the correct matches to be regarded as ground truth, we first carried out a robust standard deviation estimation of the residuals [29], as in

$$\breve{\sigma} = 1.486\left(1 + \frac{5}{n - DoF}\right) median|r_i| \tag{23}$$

where n is the number of putative correspondences; DoF is the degree of freedom of the homography matrix; r_i is the residual of the ith correspondence according to the ground truth homography, which was calculated by

$$r_i = \frac{1}{2}\left(\sqrt{\|x_2 - Hx_1\|} + \sqrt{\|x_1 - H^{-1}x_2\|}\right) \tag{24}$$

A correspondence (x_1, x_2) was regarded as an inlier if $r_i \leq 3\breve{\sigma}$, where $\breve{\sigma}$ is a threshold. In general, $\breve{\sigma}$, which is set as 2 pixels in this paper, is necessary to prevent excessive outliers from corrupting the correct matches. Once $\breve{\sigma} > 2$, the correspondences whose residuals were larger than $3\breve{\sigma}$ were considered as outliers and eliminated; and then, $\breve{\sigma}$ was recalculated with the remaining putative correspondences.

The selection results of correct matches (ground truth) were shown in Fig. 3(b), Fig. 4(b), and Fig. 5(b) for the datasets, i.e., *Bikes*, *Leuven*, and *Graffiti*, respectively. The residual distribution shown in Fig. 3(d), Fig. 4(d) and Fig. 5(d) demonstrate that the preserved correspondences obtained good accuracy and were good candidates for ground truth. In addition, the image pair of *Valbonne* provided the ground truth by [8], so there was no need to calculate the correct matches using the above criterion.

Subsequently, the putative correspondences with a different percentage of correct matches were obtained by eliminating correct matches or eliminating mismatches. Considering a posteriori percentage of correct matches defined as

$$P_{poster} = \frac{cM - eC}{iM - eC - eM} \tag{25}$$

where iM is the initial matches; cM is the correct matches; eC is the possibly eliminated correct matches; eM is the possibly eliminated mismatches; and P_{poster} is the percentage of correct matches.

[1]http://www.robots.ox.ac.uk/∼vgg/data/data-aff.html.

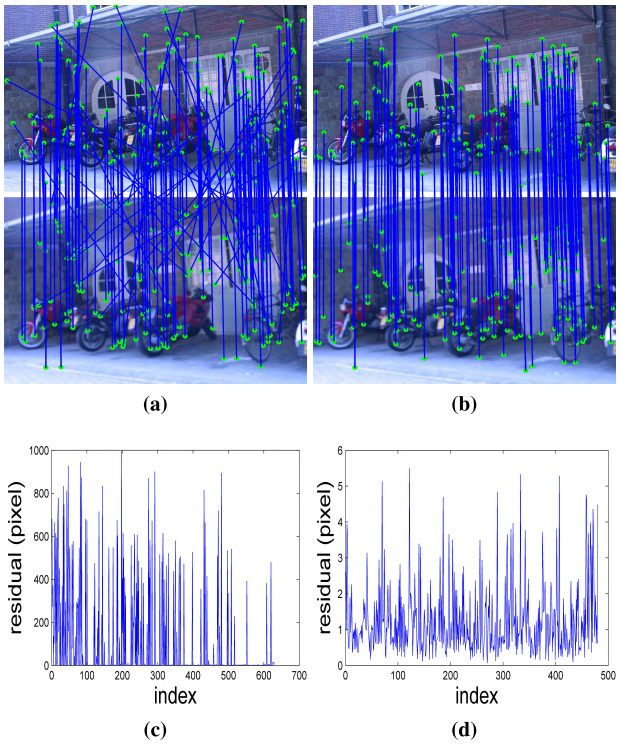


FIGURE 3. Determining inliers with *Bikes*. (a) Putative correspondence with 480 inliers and 150 outliers. (b) Inliers, which are regarded as ground truth. (c) The residual distribution of (a). (d) The residual distribution of (b). The standard deviation of the residuals was reduced from 202.26 pixels in (c) to 0.96 pixels in (d). Only 150 randomly selected correspondences are present in (a) and (b) for visualization (blue color: a correspondence, green dot: location) (The same visualization method was used in *Leuven* and *Graffiti*).

A criterion was defined to eliminate the correct matches or the mismatches, as in

$$\begin{cases} eC = 0, eM = iM - \dfrac{cM}{P_{poster}}, & if\ cM < iM \cdot P_{poster} \\ eM = 0, eC = \dfrac{cM - iM \cdot P_{poster}}{1 - P_{poster}}, & else \end{cases} \tag{26}$$

where the first expression determines how to eliminate the mismatches while the second one determines how to eliminate the correct matches.

In this series of experiments, P_{poster} was reduced from 90% to 20%. We considered two situations: 1) inlier percentage larger than 50% and 2) inlier percentage less than 50%. The mismatch removal results for the above four datasets are presented in Table 1, where each digital pair in parentheses represented a precision-recall pair (The same representation also was used in the proceeding experiments). For the *Bikes*, *Leuven*, and *Graffiti*, almost all the correct matches were identified and a high precision level was remained when the inlier percentage was more than 50%. However, when the inlier percentage was less than 50%, the precision slowed

TABLE 1. Robustness when inlier percentage decreased. The inlier percentage preserved integers.

Inlier Pct.	90	80	70	60	50	40	30	20
Bikes	(100.0,100.0)	(99.79,100.0)	(99.43,100.0)	(99.12,100.0)	(98.04,100.0)	(97.09,100.0)	(91.43,100.0)	(75.51,100.0)
Leuven	(99.69,100.0)	(100.0,100.0)	(99.69,100.0)	(99.13,100.0)	(95.57,99.34)	(90.99,100.0)	(84.42,100.0)	(73.08,100.0)
Graffiti	(99.91,99.56)	(99.57,100.0)	(99.07,100.0)	(95.80,99.71)	(93.85,100.0)	(91.07,100.0)	(85.22,100.0)	(77.03,99.71)
Valbonne	(100.0,85.51)	(100.0,85.51)	(100.0,94.92)	(100.0,85.51)	(70.37,66.67)	(61.36,71.05)	(81.82,75.00)	(78.57,78.57)

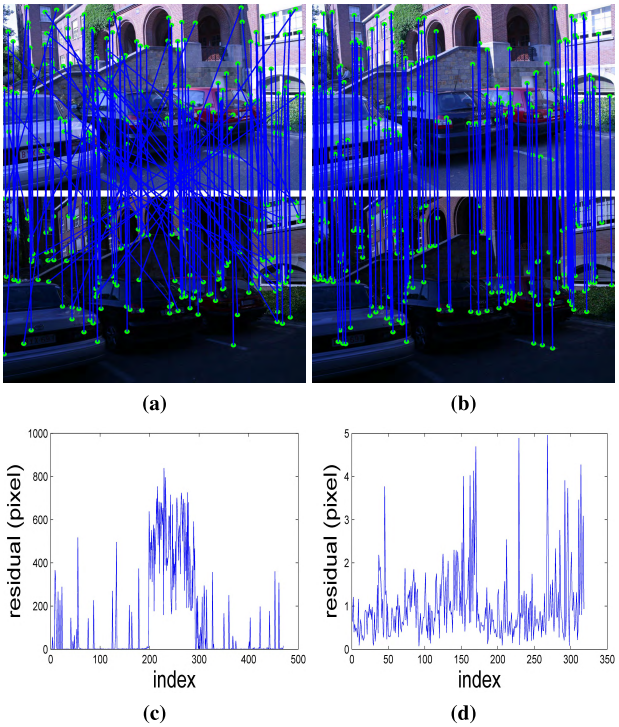


FIGURE 4. Determining the inliers with *Leuven*. (a) Putative correspondence with 318 inliers and 152 outliers. (b) Inliers, which are regarded as ground truth. (c) The residual distribution of (a). (d) The residual distribution of (b). The standard deviation of the residuals was reduced from 209.30 pixels in (c) to 0.83 pixels in (d).

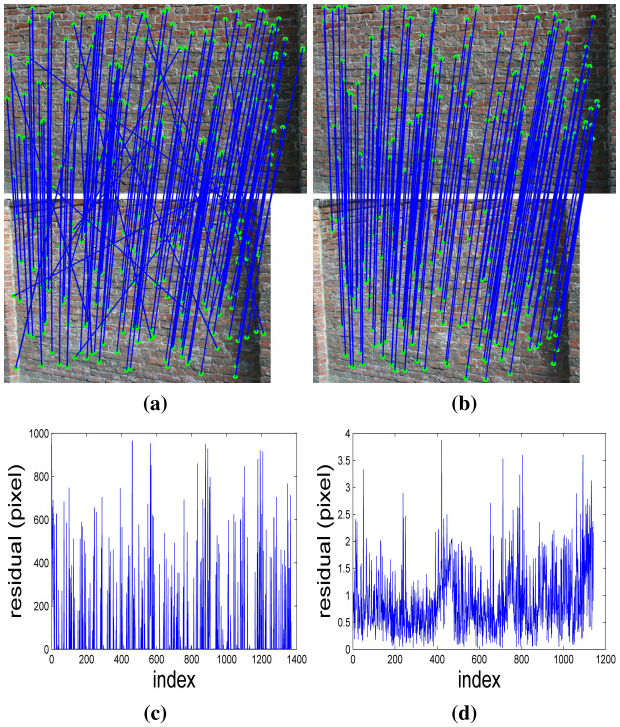


FIGURE 5. Determining the inliers with *Graffiti*. (a) Putative correspondence with 1141 inliers and 229 outliers. (b) Inliers, which are regarded as ground truth. (c) The residual distribution of (a). (d) The residual distribution of (b). The standard deviation of the residuals was reduced from 174.91 pixels in (c) to 0.60 pixels in (d).

down as the inlier percentage continued to decline. Dramatically, the recall more than 99%, which meant that it was rarely possible for our method to reject the correct matches. The mismatch removal results of *Valbonne* were unsatisfactory because of the following reasons. The putative correspondences and the inliers contained in the putative correspondences were 126 and 69, respectively. However, the numbers decreased to (77, 69), (86, 69), (99, 69), (116, 69), (114, 57), (95, 38), (81, 24), (71, 14) for the inlier percentages from 90% to 20%. In addition, comparing to the ground truth, there were ten inliers on the sky that were removed by our method as shown in Fig. 1(c). In reality, it was hard to determine whether the correspondences on the sky were inliers or outliers, and these correspondences were not what we desired in the application because of their ambiguities. Therefore, the ground truth was corrupted by the so-called inliers on the sky when a different percentage of correct matches were generated using Eq. (26). In addition, a small number of inliers would make our method ineffective in statistics. For example, the recall 0.75 (18/24) was larger than 0.67 (16/24), which is significant in statistics, but there was almost no difference in terms of the preserved inlier number (18 vs. 16).

B. EXPERIMENTS ON PRECISION-RECALL AND EFFICIENCY

In this subsection, we compared our method with the following state-of-the-art methods: RANSAC, MLESAC, LMedS, VFC, and LLT in terms of precision-recall and efficiency. Our first aim was to illustrate the ability of our method to eliminate large outlier percentage mismatches but not to demonstrate it as a replacement for the existing methods, utilizing once again the *Bikes*, *Leuven* and *Graffiti* datasets. Our second aim was to show the overall performance comparisons between our method and the five compared methods in terms of average precision-recall and efficiency on the whole *VGG* affine benchmark. The maximum number of iterations for all the above methods were set as 1000. We also provide the results of SparseVFC, which is VFC's fast

implementation version and those of LLT's non-rigid version and LLT's affine version. For VFC and LLT, we implemented them based on the publicly available codes, and their parameter settings were kept at the default. For RANSAC, the distance threshold was set as 6.0 pixels because the max error of the ground truth was below 6.0 pixels as shown in Fig. 3(d). All the experiments were performed on a laptop with 2.50 GHz Inter(R) Core(TM) i5-3210M CPU, 8GB memory and Matlab Codes.

1) RESULTS ON *BIKES*, *LEUVEN* and *GRAFFITI* WITH LARGE OUTLIER PERCENTAGE

a: PRECISION-RECALL

The experimental results in Table 2, Table 3 and Table 4 show that the precision of our method was inferior to that of RANSAC and the recall of our method was superior to that of RANSAC. Actually, RANSAC had the highest precision because it was able to set up the best distance threshold according to the ground truth calculated by Eq. (26) in these experiments. In general, the distance threshold of RANSAC is difficult to be determined in advance, which is not conducive to its universal application, and this is the reason we develop an adaptive mismatch removal method. MLESAC had high precision and low recall, which meant that a considerable number of inliers were regarded as outliers and eliminated. MLESAC also suffers from a prior sigma value in normal distribution in its maximizing the likelihood process. The precision of LMedS declined sharply when the outlier percentage was more than 50%. Generally speaking, LMedS is only robust to the situation where high inlier percentages exist. VFC, LLT and our method, which are all based on vector field interpolation theory, had similar precision-recall trade-offs. However, our method performed slightly better than VFC and LLT due to its adaptive design where the most sensitive parameters were adaptively estimated in the iterative EM algorithm process.

b: EFFICIENCY

The experimental results of efficiency shown in Table 5, Table 6 and Table 7 demonstrated that the RANSAC-like methods (i.e., RANSAC, MLESAC and LMedS) were almost ten times less efficient than the vector field interpolation-like methods (i.e., VFC, LLT and our method). The efficiency of our method has the same magnitude as those of SparseVFC and LLT-affine, and it is slightly better than the latter, but it is also obviously superior to those of VFC

TABLE 2. Precision-recall of our method comparing with RANSAC, MLESAC, LMedS, VFC and LLT on *Bikes*.

Outlier Pct.	40.00	60.00	70.09	80.21
RANSAC	(99.12,100.0)	(98.02,99.00)	(98.39,95.31)	(100.0,75.68)
MLESAC	(100.0,83.56)	(100.0,76.00)	(100.0,43.75)	(100.0,75.68)
LMedS	(100.0,83.56)	(80.00,100.0)	(59.81,100.0)	(34.04,86.49)
VFC	(99.56,100.0)	(96.15,100.0)	(87.67,100.0)	(74.00,100.0)
SparseVFC	(99.56,100.0)	(96.15,100.0)	(87.67,100.0)	(74.00,100.0)
LLT-affine	(99.56,100.0)	(98.04,100.0)	(86.49,100.0)	(74.00,100.0)
LLT-nongrid	(92.59,100.0)	(82.64,100.0)	(77.38,100.0)	(61.67,100.0)
Our's	(99.12,100.0)	(97.09,100.0)	(91.43,100.0)	(75.51,100.0)

TABLE 3. Precision-recall of our method comparing with RANSAC, MLESAC, LMedS, VFC and LLT on *Leuven*.

Outlier Pct.	40.00	60.08	70.05	80.00
RANSAC	(98.70,100.0)	(98.04,99.01)	(95.52,98.46)	(92.68,100.0)
MLESAC	(100.0,83.56)	(98.73,77.23)	(97.30,55.38)	(100.0,44.74)
LMedS	(100.0,83.33)	(79.53,100.0)	(59.63,100.0)	(35.79,89.47)
VFC	(99.13,100.0)	(89.38,100.0)	(84.42,100.0)	(76.00,100.0)
SpareVFC	(99.13,100.0)	(89.38,100.0)	(84.42,100.0)	(76.00,100.0)
LLT-affine	(98.28,100.0)	(90.18,100.0)	(84.42,100.0)	(58.33,55.26)
LLT-nongrid	(99.28,100.0)	(91.82,100.0)	(86.67,100.0)	(77.55,100.0)
Our's	(99.13,100.0)	(90.99,100.0)	(84.42,100.0)	(73.08,100.0)

TABLE 4. Precision-recall of our method comparing with RANSAC, MLESAC, LMedS, VFC and LLT on *Graffiti*.

Outlier Pct.	40.03	59.95	70.03	80.07
RANSAC	(98.00,100.0)	(95.63,100.0)	(94.74,91.84)	(96.00,84.21)
MLESAC	(100.0,89.50)	(99.07,69.28)	(95.56,75.44)	(100.0,45.61)
LMedS	(100.0,83.38)	(80.10,100.0)	(59.76,100.0)	(39.86,100.0)
VFC	(98.00,100.0)	(90.63,100.0)	(85.22,100.0)	(75.00,100.0)
SpareVFC	(98.00,100.0)	(90.63,100.0)	(85.22,100.0)	(75.00,100.0)
LLT-affine	(93.21,100.0)	(86.29,98.69)	(80.00,97.96)	(70.00,98.25)
LLT-nongrid	(91.91,99.42)	(85.71,94.12)	(77.69,95.92)	(62.92,98.25)
Our's	(98.00,100.0)	(91.07,100.0)	(85.22,100.0)	(77.03,100.0)

TABLE 5. Efficiency of our method comparing with RANSAC, MLESAC, LMedS, VFC and LLT on *Bikes* (seconds).

Outlier Pct.	40.00	60.00	70.09	80.21
RANSAC	3.524	2.482	2.272	2.111
MLESAC	3.802	2.909	2.591	2.188
LMedS	3.810	2.760	2.353	2.004
VFC	0.320	0.230	0.250	0.127
SparseVFC	0.110	0.131	0.174	0.117
LLT-affine	0.096	0.073	0.071	0.394
LLT-nongrid	0.284	0.226	0.151	0.657
Our's	0.076	0.073	0.099	0.080

TABLE 6. Efficiency of our method comparing with RANSAC, MLESAC, LMedS, VFC and LLT on *Leuven* (seconds).

Outlier Pct.	40.00	60.08	70.05	80.00
RANSAC	3.245	2.255	1.901	2.996
MLESAC	3.172	2.281	2.017	1.908
LMedS	3.155	2.156	2.062	1.827
VFC	0.344	0.198	0.125	0.672
SparseVFC	0.110	0.094	0.109	0.438
LLT-affine	0.094	0.047	0.079	0.047
LLT-nongrid	0.078	0.250	0.186	0.370
Our's	0.053	0.056	0.049	0.051

TABLE 7. Efficiency of our method comparing with RANSAC, MLESAC, LMedS, VFC and LLT on *Graffiti* (seconds).

Outlier Pct.	40.03	59.95	70.03	80.07
RANSAC	4.655	3.168	2.974	2.729
MLESAC	4.457	3.423	3.015	2.652
LMedS	4.622	3.300	2.845	2.567
VFC	0.444	0.365	0.250	0.235
SparseVFC	0.101	0.095	0.091	0.125
LLT-affine	0.140	0.125	0.109	0.098
LLT-nongrid	0.203	0.294	0.188	0.230
Our's	0.056	0.059	0.051	0.054

TABLE 8. Average precision-recall of our method comparing with RANSAC, MLESAC, LMedS, VFC and LLT on *VGG* affine benchmark.

datasets	*Bark*	*Bikes*	*Boat*	*Graffiti*	*Leuven*	*Trees*	*Ubc*	*Wall*	Total
RANSAC	(99.67,96.12)	(98.14,100.0)	(98.05,97.57)	(93.91,90.67)	(97.10,100.0)	(96.40,99.09)	(96.89,99.97)	(99.17,98.96)	(97.61,98.69)
MLESAC	(99.87,95.59)	(98.20,99.97)	(97.93,99.34)	(96.56,96.75)	(97.53,100.0)	(96.68,98.36)	(97.06,99.94)	(99.20,98.96)	(97.87,99.00)
LMedS	(66.75,84.30)	(100.0,61.99)	(86.02,71.17)	(60.98,74.34)	(100.0,64.41)	(84.74,87.21)	(100.0,61.03)	(88.02,65.47)	(86.67,70.08)
VFC	(98.60,95.49)	(98.69,100.0)	(97.89,99.89)	(92.13,99.70)	(98.34,100.0)	(95.09,99.94)	(97.77,100.0)	(99.28,99.65)	(97.57,99.56)
SparseVFC	(98.60,95.49)	(98.69,100.0)	(97.94,99.89)	(92.21,99.59)	(98.30,100.0)	(95.09,99.94)	(97.77,100.0)	(99.27,99.65)	(97.57,99.55)
LLT-affine	(99.93,99.87)	(98.39,99.97)	(97.72,99.83)	(89.19,98.78)	(98.14,100.0)	(95.19,99.97)	(97.74,99.97)	(97.53,99.04)	(97.13,99.73)
LLT-nongrid	(92.51,94.92)	(98,42.99.97)	(92.69,72.27)	(89.42,82.25)	(98.14,100.0)	(95.38,99.97)	(97.74,99.97)	(97.30,98.03)	(96.40,92.29)
Our's	(98.74,96.25)	(98.72,100.0)	(97.73,99.94)	(91.02,99.70)	(98.30,100.0)	(94.90,99.94)	(97.77,100.0)	(99.26,99.63)	(97.47,99.62)

TABLE 9. Average efficiency of our method comparing with RANSAC, MLESAC, LMedS, VFC and LLT on *VGG* affine benchmark (seconds).

datasets	*Bark*	*Bikes*	*Boat*	*Graffiti*	*Leuven*	*Trees*	*Ubc*	*Wall*	Total
RANSAC	6.48	6.95	5.40	4.21	5.22	11.40	6.72	9.03	6.93
MLESAC	6.68	6.89	5.24	4.32	5.30	11.50	6.95	9.43	7.04
LMedS	7.44	7.05	5.43	4.36	5.36	11.73	6.95	9.38	7.21
VFC	1.41	0.75	0.58	0.85	0.42	4.38	0.81	2.18	1.42
SparseVFC	0.09	0.26	0.34	0.11	0.07	0.24	0.06	0.05	0.15
LLT-affine	0.30	0.21	0.15	0.20	0.13	0.49	0.21	0.35	0.25
LLT-nongrid	0.66	0.22	0.22	0.63	0.41	0.50	0.22	0.42	0.41
Our's	0.36	0.04	0.04	0.06	0.03	0.05	0.04	0.04	0.08

and LLT-nongrid. For LLT-affine, vector field f is an affine matrix so it does not need to be interpolated in the complex matrix-value RKHS as in Eq. (1) and can yield a significant increase in speed. For SpareVFC, a spare approximation was proposed by [8] and [42] to increase the efficiency without losing its performance. Our method, similar to SpareVFC, applies an improved regularization and adaptive RBF to increase its robustness and wide application. However, as shown in Table 5, Table 6 and Table 7, our method performed slightly better in efficiency than SparseVFC. We found that solving the linear system in SpareVFC was easy to singular with a fixed regularization coefficient-λ, which would be avoided by our improved adaptive regularization term as shown in Eq. (1), Eq. (5) and Eq. (14). As it may change the iterative number of EM algorithm and further influence the efficiency of the proposed method, it was no surprise that it improved the efficiency of our method.

2) RESULTS ON THE WHOLE *VGG* AFFINE BENCHMARK

The *VGG* affine benchmark contains eight different types of datasets, i.e. *Bark*, *Bikes*, *Boat*, *Graffiti*, *Leuven*, *Trees*, *Ubc*, *Wall*, and each one contains five image pairs. We set the SIFT distance ratio threshold as 1.0, and tested our method compared with the other five methods in terms of average precision-recall and efficiency.

The experimental results in terms of average precision-recall were shown in Table 8, and the average precision-recall pairs on the whole benchmark were (97.61, 98.69), (97.87, 99.00), (86.67, 70.08), (97.57, 99.56), (97.57, 99.55), (97.13, 99.73), (96.40, 92.29) and (97.47, 99.62) for RANSAC, MLESAC, LMeds, VFC, SpareVFC, LLT-affine, LLT-nongrid and our method, respectively. As shown in Table 8, the average precision of our method was basically the same as those of RANSAC-like methods (RANSAC and MLESAC), but the average recall was slightly superior to those of the latter. However, our method was superior to RANSAC-like methods in terms of both precision and recall in some cases, such as *Leuven*, *Ubc* and *Wall*. MLESAC performed slightly better than RANSAC, and they both achieved excellent performance. LMeds showed the worst precision-recall trade-offs because of the fact that the verify progress in LMeds' hypothesis-and-verify flowchart is blind without knowing the prior inlier percentage. VFC, LLT and our method had the best precision-recall trade-offs, and our method was a little better than LLT. In addition, the number of the cumulative inliers identified by our method was 20299, compared to 20110, 20173, 14280, 20287, 20286, 20321 and 18805 for RANSAC, MLESAC, LMeds, VFC, SpareVFC, LLT-affine, and LLT-nongrid respectively, which only was less than that identified by LLT-affine. These experiments also indirectly demonstrated that the mismatch removal capability of our method is not affected by image blur, illumination, viewpoint changes and image rotation-scaling as these cases are all contained in *VGG* benchmark.

The experimental results of average efficiency were shown in Table 9, and the average elapsed time (seconds) on the whole benchmark were 6.93, 7.04, 7.21, 1.42, 0.15, 0.25, 0.41 and 0.08 for RANSAC, MLESAC, LMeds, VFC, SpareVFC, LLT-affine, LLT-nongrid and our method, respectively. As shown in Table 9, the vector field interpolation-like methods have obvious advantages in terms of efficiency, compared to the RANSAC-like methods. However, our method is most time efficient, compared to the other vector field interpolation-like methods. This may be explained by the adaptive design of regularization coefficient that reduces the iteration numbers of EM algorithms, because the fixed improper regularization coefficients may increase the number of iterations, or even affect the solution to the mixture likelihood model.

C. SENSITIVENESS OF INNER PARAMETERS

There are two insensitive fixed parameters (called inner parameters) in our method: τ, which is the posterior probability of a sample being an inlier, and *maxT*, which is the maximum number used in the simple training method to determine the width of RBF according to Eq. (19), Eq. (20) and Eq. (21). When we determined the width of RBF via Eq. (17) or Eq. (18), there was an inner parameter τ left. Few parameters also indirectly determine the adaptability and robustness of our method, and illustrated our method is easier to implement software applications.

To validate the sensitiveness of these two parameters, using the *Valbonne* dataset again, we fixed each one of them individually and evaluated the influence of the remained inner parameter on the precision-recall. The putative correspondence and ground truth with 69 inliers determined that we neither eliminated the outliers nor the inliers as was done in subsection IV-A.

Other factors remained unchanged, and parameter τ was reduced from 95% to 55%. The experimental results shown in Table 10 demonstrated that our method was slightly sensitive to parameter τ. Actually, the posterior probability of a sample being an inlier was either larger than 90% or less than 5% as shown in Fig. 6. Therefore, we utilized 70% throughout this paper.

TABLE 10. Influence of τ on identifying inliers.

τ	0.95	0.85	0.75	0.65	0.55
inliers	55	59	59	59	59

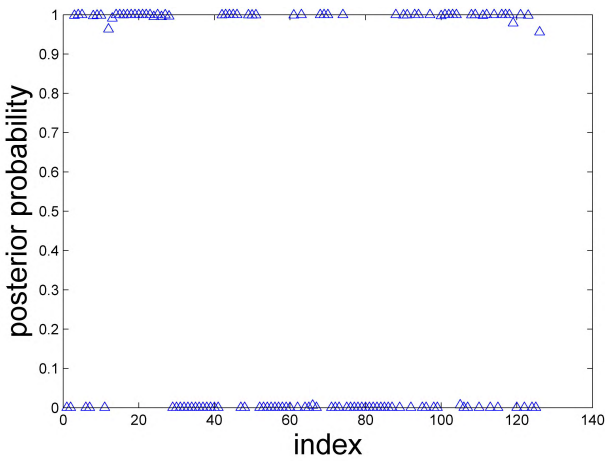


FIGURE 6. Posterior probability of a correspondence being an inlier. The value of blue triangle on y-axis is the probability of the i^{th} correspondence being an inlier.

Likewise, we fixed parameter-τ as 70% and reduced *maxT* from 500 to 100. Intuitively, the larger *maxT*, the closer the width of RBF to the optimal. The experimental results shown in Table 11 demonstrated that *maxT* had no obvious influence on the performance of our method and also indirectly verified that the width of RBF calculated according to Eq. (19), Eq. (20) and Eq. (21) could be approximated to the optimal. Considering the efficiency needed, therefore, we recommend *maxT* at 100 throughout this paper.

TABLE 11. Influence of *maxT* on identifying inliers.

$maxT$	500	400	300	200	100
$\bar{\sigma}$	4.352	4.327	4.336	4.342	4.364
β^*	0.106	0.109	0.106	0.106	0.105
inliers	59	59	59	59	59

β^*: the parameters used in VFC and LLT, which is 0.1 by default. We presented its equivalent value calculated by $\bar{\sigma}$, for comparison.

D. EXPERIMENTS ON HIGH RESOLUTION AIRBORNE DATASETS

We also tested our method on two high resolution airborne image pairs that were structured scenes where the vector fields were discontinuities, which meant that the the priori assumption about the vector field f (Eq. (5)) was likely untenable. The experiments aimed to show the ability of our method to reject mismatches and identify correct matches under these scenarios.

FIGURE 7. Results on an airborne image pair with 60% overlap along the track. (a) 4711 putative correspondences. (b) 2447 preserved correspondences. They are both 7500×11500 in size, downscaled two times in the experiment. There are no obvious outliers from visualization.

The SIFT distance ratio threshold was set at 1.5 to establish the putative correspondences as shown in Fig. 7(a) and Fig. 8(a) and most of the correct correspondences were established as shown in Fig. 7(b) and Fig. 8(b). However, for the wide baseline image pairs, such as Fig. 8, where the vector fields were likely discontinuous, partial outliers corrupted the preserved inlier correspondences.

In order to evaluate the precision-recall, we utilized the preserved inlier correspondences to re-estimate the epipolar geometry using RANSAC and obtained the fundamental matrix that was subsequently used to measure the residual distribution of the preserved correspondences and rejected correspondences. The experimental results are presented in Table 12 and Table 13 for the two experiments, which were referred to as "Test-along" and "Test-across," respectively. As shown in Table 12 and Table 13, considering the correspondences with the residuals below 2 pixels, 4 pixels, or 6 pixels as inliers, the mismatch percentages in the putative correspondences were 59.97%, 53.45%, 51.12% for "Test-along" and 93.42%, 91.10%, 89.73% for "Test-across," respectively. Correspondingly, the

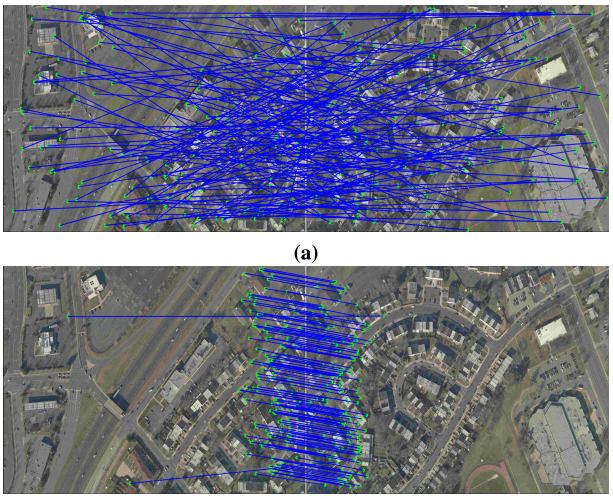


FIGURE 8. Results on an airborne image pair with 15% overlap across the track. Both the images are rotated 90 degrees for visualization. (a) 3010 putative correspondences. (b) 360 preserved correspondences. They are both 7500×11500 in size, downscaled two times in the experiment. From (b), we can see that a very small number of outliers are mixed in the preserved inliers.

TABLE 12. Residual distribution of "Test-along."

residual	$\leq$ 2pix.	2-4pix.	4-6pix.	>6pix.
2447 Preserved correspondences	1907	309	112	119
2264 Rejected correspondences	3	2	1	2258

TABLE 13. Residual distribution of "Test-across."

residual	$\leq$ 2pix.	2-4pix.	4-6pix.	>6pix.
360 Preserved correspondences	196	68	41	55
2650 Rejected correspondences	2	2	0	2646

precision-recall pairs were (77.93, 99.84), (90.56, 99.77), (95.14, 99.74) for "Test-along" and (54.44, 98.99), (73.33, 97.78), (84.74, 98.71) for "Test-across," respectively.

These experiments demonstrated that when our method established good correspondences even with a large mismatch percentage (e.g. 90%), a small number of outliers also may be introduced. However, these outliers actually can be eliminated given a certain residual threshold. For example, providing a residual threshold of 2.0 pixels and thereby considering the correspondences with residuals below 2.0 pixels as inliers, the precision-recall is raised from (54.44, 98.99) to (100.0, 98.99) for "Test-across."

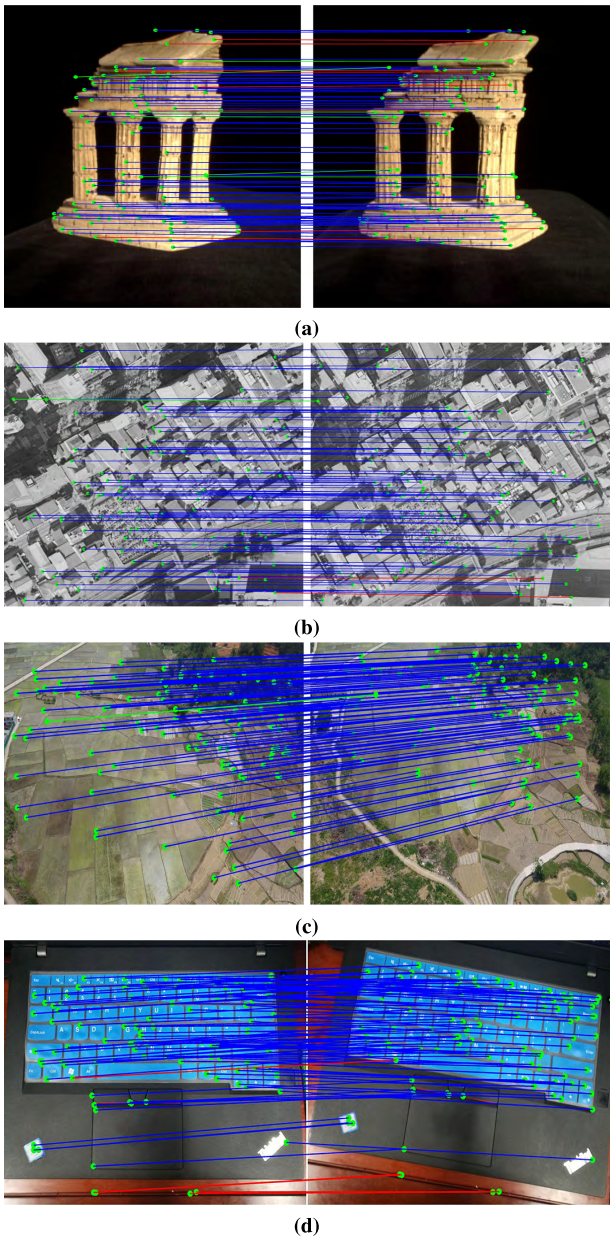


FIGURE 9. Results on other image pairs. (a) Result on *Temple*,[2] and precision-recall is (92.13, 93.60). (b) Result on *Unkown* [42], and precision-recall is (98.92, 99.84). (c) Result on an oblique image pair (obtained by unmmanned aerial vehicle), and precision-recall is (100.0, 99.75). (d) Result on *Computer* (obtained by our smart mobilephone), and precision-recall is (95.53, 100.0) (blue = true positive, green = true negative, red = false positive, greed dot = location). And at most 100 randomly chosen matches are shown for visibility.

V. CONCLUSION

In this paper, we introduced a modified regularized risk function and an adaptive Gaussian kernel function to interpolate vector fields to achieve our adaptive mismatch removal method. Experiments on real datasets demonstrated our method was robust to a large outlier percentage and slightly outperformed the vector field interpolation-like methods such as VFC and LLT and some RANSAC-like methods such as RANSAC, MLESAC and LMedS in precision-recall trade-off and efficiency. Our method has a significant advantage over state-of-the-art methods such as RANSAC in self-adaptability, in which the inner parameters were proved to have no obvious influence on its performance. In addition, we also tested our method on high resolution airborne image pairs where our modified slow-and-smooth assumption of the vector fields (Eq. (5)) may not hold because of the

[2]This dataset is available from: http://vision.middlebury.edu/mview/data.

vector fields' discontinuities. However, the experimental results showed that our method was still very effective for these cases. In summary, our method exhibited good performance and adaptability, and demonstrated that it is suitable for wide use to solve the image mismatch removal problem in computer vision and photogrammetry. However, more quantitative experiments comparing with more advanced methods also should taken into consideration in future work.

APPENDIX
VISUALIZATION RESULTS ON OTHER DATASETS

At the end of this paper, we also provided some visualization results on other datasets as supplements to the experiment section shown in Fig. (9)(a)-(d). In these cases, the ground truth was obtained by manually selecting inliers. However, the ground truth may have some ambiguities (i.g. some inliers were regarded as mismatches in the ground truth), because the manual inspection was affected by the image local texture, which played a negative impact on visualization. For example, some correct matches (i.g. some red ones) on the top of Fig. 9(a) were regarded as mismatches in the ground truth, however, they were identified as inliers in our method and they were exactly correct matches from their actual locations on the image pair. The other three image pairs have higher resolution than *Temple*, so there are no such ambiguities.

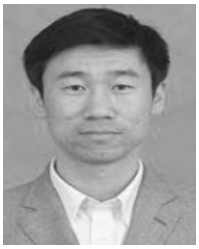
ACKNOWLEDGMENT

The authors would like to thank J. Ma for the public MATLAB codes for implementing VFC and LLT online.

REFERENCES

[1] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 2003.
[2] C. Liu, J. Yuen, and A. Torralba, "SIFT flow: Dense correspondence across scenes and its applications," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 5, pp. 978–994, May 2011.
[3] 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.*, Jul. 2013, pp. 2147–2154.
[4] M. Sonka, V. Hlavac, and R. Boyle, *Image Processing, Analysis, and Machine Vision*, 2nd ed. Pacific Grove, CA, USA: Brooks/Cole, 1999.
[5] X. Li and Z. Hu, "Rejecting mismatches by correspondence function," *Int. J. Comput. Vis.*, vol. 89, no. 1, pp. 1–17, 2010.
[6] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.
[7] R. Bouchiha and K. Besbes, "Automatic remote-sensing image registration using surf," *Int. J. Comput. Theory Eng.*, vol. 5, no. 1, pp. 88–92, Feb. 2013.
[8] 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.
[9] J. Ma, J. Zhao, Y. Zhou, and J. Tian, "Mismatch removal via coherent spatial mapping," in *Proc. Int. Conf. Image Process.*, 2012, pp. 1–4.
[10] 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.
[11] M. A. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with application to image analysis and automated cartography," *Commun. ACM*, vol. 24, no. 6, pp. 381–395, 1981.
[12] O. Chum and J. Matas, "Matching with PROSAC—Progressive sample consensus," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2005, pp. 220–226.
[13] P. H. S. Torr and A. Zisserman, "MLESAC: A new robust estimator with application to estimating image geometry," *Comput. Vis. Image Understand.*, vol. 78, no. 1, pp. 138–156, 2000.
[14] A. L. Yuille and N. M. Grzywacz, "A mathematical analysis of the motion coherence theory," *Int. J. Comput. Vis.*, vol. 3, no. 2, pp. 155–175, 1989.
[15] A. L. Yuille and N. M. Grzywacz, "A computational theory for the perception of coherent visual motion," *Nature*, vol. 333, no. 6168, pp. 71–74, 1988.
[16] H. Chui and A. Rangarajan, "A new point matching algorithm for nonrigid registration," *Comput. Vis. Image Understand.*, vol. 89, nos. 2–3, pp. 114–141, 2003.
[17] C. A. Micchelli and M. A. Pontil, "On learning vector-valued functions," *Neural Comput.*, vol. 17, no. 1, pp. 177–204, 2005.
[18] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*. Cambridge, MA, USA: MIT Press, 2006.
[19] J. Ma, H. Zhou, J. Zhao, Y. Gao, J. Jiang, and J. Tian, "Robust feature matching for remote sensing image registration via locally linear transforming," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 12, pp. 6469–6481, Dec. 2015.
[20] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Roy. Statist. Soc. B, Methodol.*, vol. 39, no. 1, pp. 1–38, 1977.
[21] A. L. Yuille, "Generalized deformable models, statistical physics, and matching problems," *Neural Comput.*, vol. 2, no. 1, pp. 1–24, 1990.
[22] G. Wahba, *Spline Models for Observational Data*. Philadelphia, PA, USA: SIAM, 1990.
[23] P. J. Rousseeuw, "Least median of squares regression," *J. Amer. Stat. Assoc.*, vol. 79, no. 388, pp. 871–880, 1984.
[24] S. Belongie, J. Malik, and J. Puzicha, "Shape matching and object recognition using shape contexts," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 4, pp. 509–522, Apr. 2002.
[25] P. J. Huber, *Robust Statistics*. New York, NY, USA: Wiley, 1981.
[26] R. Deriche, Z. Zhang, Q.-T. Luong, and O. Faugeras, "Robust recovery of the epipolar geometry for an uncalibrated stereo rig," in *Proc. Eur. Conf. Comput. Vis.* Berlin, Germany: Springer, 1994, pp. 567–576.
[27] D. Barath and J. Matas, "Graph-cut RANSAC," presented at the IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2018, pp. 6733–6741.
[28] J. L. Schönberger, T. Price, T. Sattler, J.-M. Frahm, and M. Pollefeys, "A vote-and-verify strategy for fast spatial verification in image retrieval," presented at 13th Asian Conf. Comput. Vis., Taipei, Taiwan, Nov. 2016, pp. 321–337.
[29] P. J. Rousseeuw and A. Leroy, *Robust Regression and Outlier Detection*. New York, NY, USA: Wiley, 1987.
[30] P. H. S. Torr and D. W. Murray, *Outlier Detection and Motion Segmentation*, vol. 2059, P. Schenker, Ed. Bellingham, WA, USA: SPIE, 1993, pp. 432–443.
[31] H. Li, "A practical algorithm for L-infinity triangulation with outliers," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2007, pp. 1–8.
[32] A. H. M. R. Imon, "Identifying multiple influential observations in linear regression," *J. Appl. Statist.*, vol. 32, no. 9, pp. 929–946, 2005.
[33] V. Barnett and T. Lewis, *Outliers in Statistical Data*, 3rd ed. New York, NY, USA: Wiley, 1994.
[34] S. Ullman, *The Interpretation of Visual Motion*, vol. 28. Cambridge, MA, USA: MIT Press, 1979.
[35] S. Gold, A. Rangarajan, C. P. Lu, S. Pappu, and E. Mjolsness, "New algorithms for 2-D and 3-D point matching: Pose estimation and correspondence," *Pattern Recognit.*, vol. 31, no. 8, pp. 1019–1031, 1998.
[36] A. Myronenko and X. Song, "Point set registration: Coherent point drift," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 12, pp. 2262–2275, Dec. 2010.
[37] T. Corpetti, E. Memin, and P. Perez, "Dense estimation of fluid flows," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 3, pp. 365–380, Mar. 2002.
[38] N. Aronszajn, "Theory of reproducing kernels," *Trans. Amer. Math. Soc.*, vol. 68, no. 3, pp. 337–404, 1950.
[39] C. A. Micchelli and M. Pontil, "On learning vector-valued functions," *Neural Comput.*, vol. 17, no. 1, pp. 177–204, 2005.
[40] R. Rifkin, G. Yeo, and T. Poggio, "Regularized least-squares classification," *Nato Sci. Ser. Sub Ser. III Comput. Syst. Sci.*, vol. 190, pp. 131–154, 2003.
[41] T. Evgeniou, M. Pontil, and T. Poggio, "Regularization networks and support vector machines," *Comput. Math.*, vol. 13, no. 1, pp. 1–50, 2000.
[42] J. Ma, J. Zhao, J. Tian, X. Bai, and Z. Tu, "Regularized vector field learning with spare approximation for mismatch removal," *Pattern Recognit.*, vol. 46, no. 12, pp. 3519–3532, 2013.

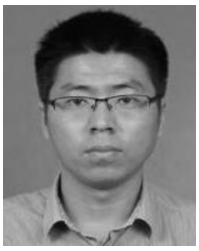
[43] S. Yu, V. Tresp, and K. Yu, "Robust multi-task learning with t-processes," in *Proc. Int. Conf. Mach. Learn.*, 2007, pp. 1103–1110.
[44] Y. Ma, J. Wang, H. Xu, S. Zhang, X. Mei, and J. Ma, "Robust image feature matching via progressive sparse spatial consensus," *IEEE Access*, vol. 5, pp. 24568–24579, 2017.
[45] G. Wang, Y. Chen, and X. Zheng, "Gaussian field consensus: A robust nonparametric matching method for outlier rejection," *Pattern Recognit.*, vol. 74, pp. 305–316, Feb. 2018.
[46] G. Wang and Y. Chen, "Fuzzy correspondences guided Gaussian mixture model for point set registration," *Knowl. Based Syst.*, vol. 136, pp. 200–209, Nov. 2017.
[47] G. Wang, Z. Wang, Y. Chen, X. Liu, Y. Chen, and L. Peng, "Learning coherent vector fields for robust point matching under manifold regularization," *Neurocomputing*, vol. 216, pp. 393–401, Dec. 2016.
[48] Y. Weiss and E. H. Adelson, "Slow and smooth: A Bayesian theory for the combination of local motion signals in human vision," Cambridge, MA, USA, Massachusetts Inst. Technol., Tech. Rep. 1624, 1998.
[49] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York, NY, USA: Springer-Verlag, 2006.



YONGJUN ZHANG was born in 1975. He received the B.S. degree in geodesy, the M.S. in geodesy, and the Ph.D. degree in geomatics from Wuhan University, Wuhan, China, in 1997, 2000, and 2002, respectively. He is currently a Professor of photogrammetry and remote sensing with the School of Remote Sensing and Information Engineering, Wuhan University. His research interests include computer vision, space, aerial, and low-attitude photogrammetry, image matching, combined bundle adjustment with multi-source data sets, 3-D city reconstruction, and industrial inspection.

XUNWEI XIE was born in 1991. He received the B.S. and M.S. degrees in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2013, and 2015, respectively, where he is currently pursuing the Ph.D. degree with the School of Remote Sensing and Information Engineering. His research interests include computer vision, image matching, and oblique photogrammetry.

XIANG WANG was born in 1990. He received the B.S. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2013, where he is currently pursuing the Ph.D. degree with the School of Remote Sensing and Information Engineering. His research interests include computer vision, visual navigation, saliency modeling, and space/low-attitude photogrammetry.



YANSHENG LI received the B.S. degree from the School of Mathematics and Statistics, Shandong University, Weihai, China, in 2010, and the Ph.D. degree from the School of Automation, Huazhong University of Science and Technology, Wuhan, China, in 2015. Since 2015, he has been an Assistant Professor with the School of Remote Sensing and Information Engineering, Wuhan University, Wuhan, China. From 2017 to 2018, he was a Visiting Assistant Professor with the Department of Computer Science, Johns Hopkins University, Baltimore, USA, where he is hosted by the distinguished Bloomberg professor A. L. Yuille. He has authored and co-authored more than 20 scientific articles. His research interests include computer vision, deep learning, saliency modeling, and remote sensing.

XIAO LING was born in 1989. He received the B.S., M.S., and Ph.D. degrees from Wuhan University, Wuhan, China, in 2012, 2014, and 2017, respectively. He is currently a Post-Doctoral Researcher with the Future Cities Laboratory, Singapore-ETH Centre, Singapore. His research interests include photogrammetry, computer vision, camera calibration, and image matching.