This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

A Mixture Likelihood Model of the Anisotropic Gaussian and Uniform Distributions for Accurate Oblique Image Point Matching

Xunwei Xie, Yongjun Zhang, Xiang Wang, and Daifeng Peng

***Abstract*— In this letter, we propose a mixture likelihood model for accurate oblique image point matching. The basic prior assumption is that the noises are anisotropic with zero mean and different covariances in x- and y-directions for inliers, while the outliers have uniform distribution, which is more suitable for tilted scenes or viewpoint changes. Furthermore, the oblique image point matching problem is formulated as an improved maximum *a posteriori* (IMAP) estimation of a Bayesian model. In this model, based on the vector field interpolation framework, we combined the mixture likelihood model and our previous adaptive image mismatch removal method, where a two-order term of the regularization coefficient is introduced into the regularized risk function, and a parameter self-adaptive Gaussian kernel function is imposed to construct the regularization term. Subsequently, the expectation–maximization algorithm is utilized to solve the IMAP estimation, in which all the latent variances are able to obtain excellent estimation. Experimental results on real data sets verified that our method was superior to some similar methods in terms of precision and also had better self-adaptability characteristic than some hypothesis-and-verify methods. More experiments on viewpoint changes demonstrated our method's effectiveness without loss of precision–recall trade-offs, besides significant efficiency improvement.**

***Index Terms*— Mixture likelihood model, oblique image, parameter adaptation, point correspondence.**

I. Introduction

POINT matching is a critical prerequisite in applications including registration, camera self-calibration, bundle adjustment, and object recognition between images, but it continues to be a fundamental problem in photogrammetry and computer vision [1]–[4]. The common point matching problem can be regularized by a similarity constraint and a geometric constraint. Especially geometric constraint means that the matches satisfy some geometrical requirements, such as homography geometry, epipolar geometry, or nonrigid geometry [5]. However, the efficient ways are still required to obtain the best solution in the regularized method for rejecting mismatches. For the oblique image point matching problem, it becomes harder because of the presence of mismatches in two point sets due to the viewpoints.

The essence of point matching is to identify inliers and reject outliers and estimate the geometric parameters [6]. It commonly makes use of a general hypothesis-and-verify flowchart: estimate a parametric model from a minimum number of the randomly selected point sets; assess the quality of the model by some criteria; and choose the hypothesis with the highest score to identify inliers. Random sample consensus (RANSAC) [7] and maximum likelihood estimation sample consensus (MLESAC) [8] are the representatives of this flowchart in the literature. RANSAC evaluates the hypothesis with the count of inliers whose residuals are below a given threshold. However, MLESAC uses a weighted voting strategy and regards the solution that maximizes the likelihood as the final optimal estimation [5]. A similar vote-and-verify strategy also achieved verification accuracy similar to some hypothesis-and-verify methods [9]. These methods are successfully utilized in some situations with excellent ability of identifying inliers from the correspondences with large outlier percentages. However, the requirement of different thresholds for different scenes, which usually have to be manually selected, limits their wide application. For example, RANSAC and MLESAC are affected by the residual threshold while estimating the epipolar geometry.

Benefit from the developments in the minimal mapping theory and motion coherence theory [10], [11], the point matching problem can also be formulated as a vector field interpolation with a high-dimensional mapping constraint. The vector field can be solved by the regularization theory. Yuille and Grzywacz [11] introduced the motion coherence theory to compute the vector field using a quadratic regularization to impose nonparametric geometric constraints on the correspondences, and this was equivalent to formulating the problem in terms of a space of kernels [5]. The components of the vector fields can be directly encoded by a series of operations in a reproducing kernel Hilbert space (RKHS), associated with a certain regularization choice to obtain a meaningful solution [4], [12]. Based on Yuille's motion coherence theory, a general framework is established for correspondence problem [13]. On the basis of this framework, a robust vector field consensus method was proposed in [5]. It associates each correspondence with a latent variable that determines if it is an inlier and obtained good performance for mismatch removal problem. Ma *et al.* [14] also extended the vector field's nonparametric geometric constraints to a local linear

Manuscript received October 25, 2018; revised December 26, 2018 and February 2, 2019; accepted February 8, 2019. This work was supported in part by the National Natural Science Foundation of China under Grant 41871368 and Grant 41801386 and in part by the Natural Science Foundation of Jiangsu Province under Grant BK20180797. *(Corresponding author: Yongjun Zhang.)*

X. Xie, Y. Zhang, and X. Wang are with the School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, China (e-mail: xwxie@whu.edu.cn; zhangyj@whu.edu.cn; mld@whu.edu.cn).

D. Peng is with the School of Remote Sensing and Geomatics Engineering, Nanjing University of Information Science and Technology, Nanjing 210044, China (e-mail: daifeng@nuist.edu.cn).

Color versions of one or more of the figures in this letter are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LGRS.2019.2899042

transformation (LLT) that can preserve some local structures among neighboring correspondences, whose different variants obtained robust point matching results for remote sensing image registration. Other vector field interpolation-like methods for correspondence problem can also be found in [15] and [16]. However, most of these methods suffer from a suitable kernel function choice and the sensitive parameter settings, which limit their wide application in different situations. Therefore, in our previous work [i.e., adaptive image mismatch removal (AIMR)], a two-order term of regularization coefficient was introduced into the regularized risk function, and a parameter self-adaptive Gaussian kernel function was imposed to construct the regularization, so as to overcome this issue [4]. Most of these methods were built on a mixture likelihood model of the isotropic Gaussian and uniform distributions [4], [5], [14], in which they followed the assumption that the noises are Gaussian on each component for inliers. However, this assumption is not rigorous under the scene of viewpoint changes, because different parts of oblique images are different trapezoids with significant resolution differences on the ground due to the specific characteristic of tilted viewpoints [17].

Thus, on basis of AIMR, we propose a new method, accurate oblique image point matching (AOPM), which introduces a mixture likelihood model of the anisotropic Gaussian and uniform distributions in terms of the noises of correspondences considering the scenes with viewpoint changes. Experiments on oblique image pairs and a public benchmark with viewpoint changes verified that the proposed AOPM not only preserves the adaptive characteristics of AIMR but also obtains higher vector field interpolation accuracy than the compared vector field interpolation-like methods, demonstrating its good performance.

II. METHODOLOGY

A. Mixture Likelihood Model

Given a set of N putative correspondences $S = \{(u_n, v_n)\}_{n=1}^{N}$, where u_n and v_n are the normalized samples in the left and right images, respectively; our aim is to fit a mapping f interpolating the sample set, i.e., $\forall n$, $v_n = f(u_n)$. The purpose of data normalization is to control the influence of the point coordinate system on the method's performance [5]. According to the motion field theory, we convert the normalized correspondence (u_n, v_n) into a motion field sample by a transformation, i.e., $(u_n, v_n) \leftarrow (u_n, v_n - u_n)$, in order to be applied to the framework of the vector field interpolation method [4].

We make the assumption that the noises are anisotropic Gaussian with zero mean and uneven standard deviations (σ_x, σ_y) for inliers, and the distribution is uniform $1/a$ for outliers, where a is the volume of uniform distribution. A latent variable $z_n \subseteq \{0, 1\}$ is associated with the nth sample correspondence, where $z_n = 1$ indicates that the sample is an inlier, otherwise it is an outlier. Let $\mathcal{U}$ and $\mathcal{V}$ be the sets of inputs and outputs, in which the nth row represents a motion field correspondence (u_n, v_n). Thus, the mixture likelihood model of the anisotropic Gaussian and uniform distributions has the following form:

$$P(\mathcal{V}|\mathcal{U},\theta) = \prod_{n=1}^{N}\sum_{z_n} p(v_n, z_n|u_n,\theta) = \prod_{n=1}^{N}\left[\chi\frac{e^{-\frac{\left\|v_n^x-[f(u_n)]^x\right\|^2}{2\sigma_x^2}-\frac{\left\|v_n^y-[f(u_n)]^y\right\|^2}{2\sigma_y^2}}}{2\pi\sigma_x\sigma_y}+\frac{1-\chi}{a}\right] \quad (1)$$

where $\theta = (f, \sigma_x, \sigma_y, \chi)$ is the set of unknowns that should be determined; χ represents the mixing coefficient specifying the marginal distribution over the latent variable, i.e., $p(z_n = 1) = \chi$; σ_x and σ_y represent that the noises follow the different standard deviations in the x- and y-directions on an image; v_n^x and v_n^y represent the corresponding normalized x- and y-coordinates in the right image, while $[f(u_n)]^x$ and $[f(u_n)]^y$ represent the corresponding x- and y-coordinates interpolated by vector field f (the same expressions are used in Sections II-B–II-D).

In order to solve (1), we imposed a prior probability distribution of vector field f, denoted by $p(f)$, into the mixture likelihood model. The prior of f is modified on the basis of the slow-and-smooth model [4], as in

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

where ψ(f) is a smoothness term and λ is a positive number.

As the new prior probability distribution is imposed on the vector field, we estimated an improved maximum *a posteriori* (IMAP) solution of θ. The optimal solution of IMAP is $\theta^* = \arg\max_\theta p(\mathcal{V}\,|\mathcal{U},\theta)p(f)$, which is equivalent to minimizing the negative log-likelihood function

$$E(\theta) = -\sum_{n=1}^{N}\ln\sum_{z_n} p(v_n, z_n|u_n,\theta) - \ln p(f). \quad (3)$$

Then, the vector field f can be obtained from the optimal solution θ^*. We will demonstrate how to solve the IMAP estimation, i.e., (3), in Section II-B.

B. Solution of IMAP Estimation

To solve the IMAP estimation, an expectation–maximization (EM) algorithm that includes expectation step (E-step) and maximization step (M-step) is used. As the complete data log-likelihood is hardly computed 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 so as to update θ based on the current estimation [18]. The complete-data negative log posterior, which omitted some constant terms, is given by

$$\begin{aligned}\mathcal{Q}(\theta,\theta^{\text{old}}) &= -\frac{1}{2\sigma_x^2}\sum_{n=1}^{N} p_n\left\|v_n^x - [f(u_n)]^x\right\|^2 \\ &\quad -\frac{1}{2\sigma_y^2}\sum_{n=1}^{N} p_n\left\|v_n^y - [f(u_n)]^y\right\|^2 - \ln\sigma_x\sigma_y\sum_{n=1}^{N} p_n \\ &\quad + \ln\chi\sum_{n=1}^{N} p_n + \ln(1-\chi)\sum_{n=1}^{N}(1-p_n) - \lambda\psi(f) + \frac{\lambda^2}{2} \quad (4)\end{aligned}$$

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

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

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

$$p_n = 1 - \frac{2\pi\sigma_x\sigma_y(1-\chi)}{a\chi e^{-\frac{\left\|v_n^x - [f(u_n)]^x\right\|^2}{2\sigma_x^2} - \frac{\left\|v_n^y - [f(u_n)]^y\right\|^2}{2\sigma_y^2}} + 2\pi\sigma_x\sigma_y(1-\chi)}. \tag{5}$$

2) M-Step: The unknown θ is reestimated by using the current estimation: $\theta = \arg\max_\theta \mathcal{Q}(\theta, \theta^{\mathrm{old}})$. Taking first derivatives of $\mathcal{Q}$ with respect to σ_x^2, σ_y^2, χ, and λ, and setting them to zero, we can obtain the following expressions:

$$\begin{aligned}\left[\sigma_x^2, \sigma_y^2\right] &= (\mathcal{V} - F)^{\mathrm{T}} P (\mathcal{V} - F)/2\mathrm{trace}(P)\\ \chi &= \mathrm{trace}(P)/N\\ \lambda &= \psi(f)\end{aligned} \tag{6}$$

where $F = (f(u_1)^{\mathrm{T}}, \ldots, f(u_n)^{\mathrm{T}})^{\mathrm{T}}$. In order to complete M-step, the mapping f also should be estimated in (6). We will discuss it in Section II-C.

After the EM converges, given a present threshold τ, the inlier set S_{I} can be obtained by

$$S_{\mathrm{I}} = \{(u_n, v_n) \to p_n > \tau\}. \tag{7}$$

In our previous work [4], τ was verified that it was insensitive to its choice by experiments.

C. Regularization and Kernel Function Choice

In the regularized vector field interpolation, the mapping f is modeled by requiring it to lie within a specific functional space, i.e., RKHS [14]. Specifically, f can be expressed as

$$f(u) = \sum_{n=1}^{N} \Gamma(u, u_n) c_n \tag{8}$$

where Γ is the $N \times N$ Gram matrix with (i, j) block $\Gamma(u_i, u_j)$, and c_n is the coefficient.

When the terms of $\mathcal{Q}$ with respect to $p(f)$ is considered, a modified regularized risk functional can be obtained

$$\begin{cases}[\varepsilon(f)]^x = \dfrac{1}{2\sigma_x^2}\displaystyle\sum_{n=1}^{N} p_n \left\|v_n^x - [f(u_n)]^x\right\|^2 + \lambda[\psi(f)]^x - \lambda^2/2\\ [\varepsilon(f)]^y = \dfrac{1}{2\sigma_y^2}\displaystyle\sum_{n=1}^{N} p_n \|v_n^y - [f(u_n)]^y\|^2 + \lambda[\psi(f)]^y - \lambda^2/2\end{cases} \tag{9}$$

where the first term is a weighted empirical error; the second and third terms control the tradeoff with respect to the first term.

The solution of (9) is given by (8) with the coefficient set $\{C = \{c_n\}, c_n = [c_n^x, c_n^y], n \in N\}$ solved by the following equation (10):

$$\begin{cases}\left(\Gamma + \lambda\sigma_x^2 P^{-1}\right) C^x = \mathcal{V}^x\\ \left(\Gamma + \lambda\sigma_y^2 P^{-1}\right) C^y = \mathcal{V}^y\end{cases}. \tag{10}$$

To this end, λ in (6) can be computed by

$$\lambda = \|f_N\|_{\mathrm{RKHS}}^2 = C^{\mathrm{T}} \Gamma C \tag{11}$$

where $f_N \subseteq$ RKHS. The detailed proofs of (10) and (11) can be referenced in [4], [5], and [14].

In our previous work, AIMR [4], a Gaussian kernel function was chosen to construct the Gram matrix Γ

$$\Gamma(u_i, u_j) = \exp(-\|u_i - u_j\|^2/2\delta^2). \tag{12}$$

Consider the correspondence problem, the width of kernel-δ was computed by the diagonal of the maximal enveloping rectangle of the sampled motion correspondence set [4]

$$\delta^2 = \max \|u_i - u_j\|^2. \tag{13}$$

In AIMR, a simple training method was also proposed to calculate δ^2 for the extreme distribution situation of correspondence locations.

D. Optimization With Spare Approximation

As the putative point set may contain thousands of correspondences, a heavy computational burden exists when directly solving the linear system (10). Inspired by [5], a sparse approximation and a suboptimal solution searching in the RHKS space were adopted to overcome this issue. According the sparse approximation, the linear system (10) in this letter is equivalent to a linear system as follows:

$$\begin{cases}(\widehat{\mathcal{U}}^{\mathrm{T}} P \widehat{\mathcal{U}} + \lambda\sigma_x^2 \widehat{\Gamma}_{\mathrm{A}}) \widehat{\mathrm{C}}_{\mathrm{A}}^x = \widehat{\mathcal{U}}^{\mathrm{T}} P \widehat{\mathcal{V}}^x\\ (\widehat{\mathcal{U}}^{\mathrm{T}} P \widehat{\mathcal{U}} + \lambda\sigma_y^2 \widehat{\Gamma}_{\mathrm{A}}) \widehat{\mathrm{C}}_{\mathrm{A}}^y = \widehat{\mathcal{U}}^{\mathrm{T}} P \widehat{\mathcal{V}}^y\end{cases} \tag{14}$$

where $\widehat{C}_{\mathrm{A}} = (c_1^{\mathrm{T}}, \ldots, c_M^{\mathrm{T}})^{\mathrm{T}}$ is the coefficient set with $M \ll N$; $\widehat{\mathcal{U}}$ is an $N \times M$ block matrix with (i, j) block $\Gamma(u_i, \widehat{u}_{\mathrm{j}})$; and $\widehat{\Gamma}_{\mathrm{A}}$ is an $M \times M$ block Gram matrix with (i, j) block $\Gamma(\widehat{u}_i, \widehat{u}_j)$. $\widehat{\bullet}$ means the computational component with respective to the randomly selected correspondence set.

The sparse approximation can obtain a significant increase both in time and space complexities with negligible decrease in accuracy. M is fixed to 16 in our later experiments.

III. Experiments and Discussion

We use precision and recall as evaluation criteria, where precision is the ratio of the number of preserved correct matches and preserved total correspondences, and recall is the ratio of the number of preserved correct matches and the ground truth inliers. We compared AOPM with RANSAC [7], MLESAC [8], SparseVFC [5], the nonrigid version of Local Linear Transformation (LLT) [14], and AIMR [4]. For SparseVFC and LLTV, we implemented them with the publicly codes using default parameters. All the experiments were performed on a laptop with 2.5Gz Inter(R) Core(TM) i5-3210M CPU, 8-GB memory, and MATLAB Codes.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

TABLE I
PRECISION RECALL OF RANSAC, MLESAC, SPARSEVFC, LLTV, AIMR, AND AOPM (%)

Image pair	RANSAC (average)	MLESAC (average)	SparseVFC	LLTV	AIMR	AOPM
N-F	(96.90, 83.19)	(96.97, 77.99)	(95.98, 89.32)	(94.18, 99.33)	(95.98, 89.32)	(96.05, 87.72)
N-B	(94.51, 80.12)	(94.59, 81.24)	(93.26, 90.22)	(92.05, 99.31)	(93.28, 88.16)	(93.32, 86.28)
B-F	(96.82, 80.44)	(97.03, 82.74)	(95.60, 91.84)	(95.30, 85.80)	(95.61, 92.14)	(95.90, 91.84)
Total average	(96.08, 81.25)	(96.20, 80.66)	(94.95, 90.46)	(93.84, 94.81)	(94.96, 89.88)	(95.09, 88.61)

TABLE II
MEAN AND STANDARD DEVIATION OF THE RESIDUALS OF INLIERS IDENTIFIED BY SPARSEVFC, LLTV, AIMR, AND AOPM (PIXELS)

Image pair	SparseVFC	LLTV	AIMR	AOPM
N-F	(4.46, 3.34)	(281, 141)	(4.28, 3.42)	**(3.82, 2.60)**
N-B	(4.81, 4.66)	(239, 134)	(3.68, 3.47)	**(3.34, 2.52)**
B-F	(4.77, 3.91)	(985, 634)	(5.57, 4.19)	**(4.73, 3.57)**

Fig. 1. Airborne oblique image pairs in Yangjiang City, Guangdong Province, China. Left: Forward. Middle: Nadir. Right: Backward.

A. Test on Airborne Oblique Image Pairs

We tested AOPM on three airborne oblique image pairs, i.e., Nadir–Forward (N–F), Nadir–Backward (N–B), and Backward–Forward (B–F) as shown in Fig. 1, to assess our method. We use scale invariant feature transform (SIFT) [19] to generate putative point set in these cases, and the SIFT distance ratio threshold was set as 0.8. The ground truth inliers were manually selected from the correspondences visually. There are 922, 731 and 460 putative correspondences with 749, 583 and 331 inliers, respectively; the inlier percentages are 81.24%, 79.75%, and 71.96% with respective to N–F, N–B, and B–F, respectively. Based on the epipolar geometry constraint, we performed RANSAC and MLESAC with the residual threshold as [1:1:8] pixels.

The results of six methods are presented in Table I, where each digital pair in parentheses represents a precision–recall pair. For RANSAC and MLESAC, we calculated their average precision–recall of the eight experiments with different parameter settings for each image pair. The total average precision–recall of these three image pairs is (96.08%, 81.25%), (96.20%, 80.66%), (94.95%, 90.46%), (93.84%, 94.81%), (94.96%, 89.88%), and (95.09%, 88.61%) for RANSAC, MLESAC, SparseVFC, LLTV, AIMR, and AOPM, respectively. RANSAC and MLESAC owned the worst average precision–recall tradeoff, as they will have different performances when different parameters are adopted. However, if an optimal parameter is known in advance, both RANSAC and MLESAC can yield quite satisfactory performance. Their limitations are that different parameters are required when encountering different scene structures, different imaging conditions, and different image distortions, which limit their wide application. AIMR with good parameter self-adaptability had almost the same performance in terms of precision–recall tradeoff as SparseVFC. LLTV showed a little lower precision and higher recall, because its identified inliers were corrupted by considerable outliers whose number is 46, 50, and 14 with respective to N–F, N–B, and B–F, respectively. AOPM inherits the self-adaptive characteristic of AIMR and shows a better precision for these oblique image pairs.

To illustrate the accuracy of AOPM, we compared the residual distribution of the inliers identified by these four vector field interpolation-like methods. For SparseVFC, AIMR, and AOPM, the posterior vector field f was used to calculate the fitting residuals solved by the form $\{res_n = \|v_n - f(u_n)\|^2\}$, and for LLTV, the posterior transformation model $\mathcal{T}$ [14] was used to calculate the transformed residuals solved by the form $\{res_n = \|v_n - \mathcal{T}(u_n)\|^2\}$. The mean and standard deviation of residuals of their identified inliers are presented in Table II, where each parenthesis represents a mean–standard deviation pair. As shown, AOPM has the smallest mean and standard deviation when comparing with the other three methods. This may be explained by two reasons: 1) adaptive parameter design makes the empirical error closeness to the data term and prevents overfitting or underfitting that may be caused by a fixed regularization coefficient in (9) and 2) the mixture likelihood model of the anisotropic Gaussian and uniform distributions is much more suitable for the scenes with viewpoint changes. The first argument can be verified that the fitting accuracy of AIMR was much better than that of SparseVFC, due to its self-adaptability. LLTV showed the worst fitting residual results for these tilted scenes due to the improper regularization, although it performed well in terms of recall. This can be explained that the difference between the disparities of the point set in local areas is not quite small for airborne oblique image pairs, unlike that for remote sensing satellite image pairs. Thus, the nonoptimal local geometrical constraint naturally has a negative influence on the regularization with respective to its transformation $\mathcal{T}$, which leads to an abnormal mean and standard deviation.

In summary, AOPM is superior to some hypothesis-and-verify methods in terms of parameter self-adaptability, such as RANSAC and MLESAC, and it is also superior to some vector field interpolation-like methods in terms of precision–recall tradeoffs and accuracy of vector field fitting, such as SparseVFC, LLTV, and AIMR.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

TABLE III
AVERAGE PRECISION RECALL AND EFFICIENCY OF THE SIX METHODS ON VGG BENCHMARK WITH VIEWPOINT CHANGES

Criterion	RANSAC	MLESAC	SparseVFC	LLTV	AIMR	AOPM
Precision-Recall (%)	(97.54, 95.71)	(98.56, 98.01)	(97.42, 99.67)	(93.31, 97.02)	(97.53, 99.65)	(98.57, 97.16)
Time (s)	13.249	13.672	0.164	1.277	0.032	0.068

B. Test on VGG Affine Benchmark With Viewpoint Changes

We also tested AOPM on VGG affine benchmark [20] with viewpoint changes, compared with other five methods in terms of average precision–recall and efficiency. The test data contain 10 image pairs. In these cases, the camera position is fixed during acquisition, thus they obey homographies, and the ground truth homographies are supplied by the benchmark. We use the Hessian-affine method [21] in these cases to generate the putative point set, because it is more robust against big viewpoint variations than SIFT. The distance ratio threshold was set as 0.9, and there are four image pairs with inlier percentage below 50%. Based on the homography geometric constraint, the residual threshold for RANSAC and MLESAC was set as 6.0 pixels.

The experimental results of average precision–recall and efficiency are given in Table III. As shown, AOPM is superior to RANSAC both in precision and recall and slightly inferior to MLESAC in terms of recall. However, this is based on the prerequisite that the approximate optimal parameter threshold of RANSAC and MLESAC was well set in advance. Compared with other vector field interpolation-like methods, AOPM also yields obvious advantages in precision due to the fact that the proposed mixture likelihood model is more adapted to the image pairs with viewpoint changes for correspondence problem. In these cases, LLTV showed the worst precision–recall tradeoff for its much lower ability of rejecting outliers. AOPM showed a little lower recall than SparseVFC and AIMR for the reason that it is more likely to eliminate the correspondence with relatively large gross error, which directly demonstrated the good accuracy of AOPM. In addition, AOPM achieves a significant speedup with respect to RANSAC and MLESAC without degrading any performance. Meanwhile, AOPM also is slightly more efficient than SparseVFC due to its self-adaptive design as AIMR, further reducing the iteration number of EM algorithm so as to improve efficiency.

IV. CONCLUSION

In this letter, we propose a method for AOPM. Based on the vector field interpolation framework for correspondence problem, AOPM combines the mixture likelihood model of the anisotropic Gaussian and uniform distributions and our previous AIMR method. It guarantees that AOPM not only has the same good self-adaptive characteristic as AIMR but also is more adapted to the image pairs with viewpoint changes than the latter for mismatch removal problem. Experiments on real data sets demonstrated that AOPM outperformed some hypothesis-and-verify methods in terms of self-adaptability and efficiency, such as RANAC and MLESAC, and also won out on precision–recall tradeoff and yielded more accurate vector field fitting results without loss of efficiency when comparing with other state-of-the-art vector field interpolation-like methods, such as SparseVFC, LLTV, and AIMR. In future work, more quantitative experiments comparing with other advanced methods should be taken into consideration.

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.*, Jun. 2013, pp. 2147–2154.
[4] Y. Zhang, X. Xie, X. Wang, Y. Li, and X. Ling, "Adaptive image mismatch removal with vector field interpolation based on improved regularization and Gaussian kernel function," *IEEE Access*, vol. 6, pp. 55599–55613, 2018. doi: 10.1109/ACCESS.2018.2871743.
[5] 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.
[6] X. Li and Z. Hu, "Rejecting mismatches by correspondence function," *Int. J. Comput. Vis.*, vol. 89, no. 1, pp. 1–17, 2010.
[7] 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, Jul. 1981.
[8] 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.
[9] 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," in *Proc. Asian Conf. Comput. Vis.*, Taipei, Taiwan, Nov. 2016, pp. 321–337.
[10] S. Ullman, *The Interpretation of Visual Motion*, vol. 28. Cambridge, MA, USA: MIT Press, 1979.
[11] 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.
[12] C. A. Micchelli and M. A. Pontil, "On learning vector-valued functions," *Neural Comput.*, vol. 17, no. 1, pp. 177–204, 2005.
[13] H. Chui and A. Rangarajan, "A new point matching algorithm for non-rigid registration," *Comput. Vis. Image Understand.*, vol. 89, nos. 2–3, pp. 114–141, Feb./Mar. 2003.
[14] 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.
[15] 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.
[16] 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.
[17] L. Xie, H. Hu, J. Wang, Q. Zhu, and M. Chen, "An asymmetric re-weighting method for the precision combined bundle adjustment of aerial oblique images," *ISPRS J. Photogramm. Remote Sens.*, vol. 117, pp. 92–107, Jul. 2016.
[18] C. M. Bishop, *Pattern Recognition and Machine Learning*, New York, NY, USA: Springer, 2006.
[19] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.
[20] *Affine Covariant Regions Datasets*. Accessed: Feb. 23, 2019. [Online]. Available: http://www.robots.ox.ac.uk/~vgg/data/data-aff.html
[21] K. Mikolajczyk and C. Schmid, "An affine invariant interest point detector," in *Proc. 7th Eur. Conf. Comput. Vis. (ECCV)*, Copenhagen, Denmark, May 2002, pp. 128–142.