IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received December 3, 2018, accepted December 14, 2018, date of publication December 20, 2018,
date of current version January 16, 2019.

Digital Object Identifier 10.1109/ACCESS.2018.2888856

Local Feature Descriptor for Image Matching: A Survey

CHENGCAI LENG[1,2,3], HAI ZHANG[1,4], BO LI[2], GUORONG CAI[5], ZHAO PEI[6], AND LI HE[7]

[1]School of Mathematics, Northwest University, Xi'an 710127, China
[2]School of Mathematics and Information Sciences, Nanchang Hangkong University, Nanchang 330063, China
[3]Department of Computing Science, University of Alberta, Edmonton, AB T6G 2E8, Canada
[4]Faculty of Information Technology and State Key Laboratory of Quality Research in Chinese Medicines, Macau University of Science and Technology, Macau 999078, China
[5]College of Computer Engineering, Jimei University, Xiamen 361021, China
[6]School of Computer Science, Shaanxi Normal University, Xi'an 710119, China
[7]Department of Electromechanical Engineering, Guangdong University of Technology, Guangzhou 510006, China

Corresponding authors: Hai Zhang (zhanghai@nwu.edu.cn) and Li He (heli@gdut.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61702251, Grant 61363049, Grant 61703115, Grant 61673125, and Grant 61501286, in part by the State Scholarship Fund of China Scholarship Council (CSC) under Grant 201708360040, in part by the Natural Science Basic Research Plan in Shaanxi Province of China under Grant 2018JM6030, in part by the Key Research and Development Program in Shaanxi Province of China under Grant 2018GY-008, in part by the Fundamental Research Funds for the Central Universities under Grant GK201702015, in part by the Doctor Scientific Research Starting Foundation of Northwest University under Grant 338050050, in part by the Youth Academic Talent Support Program of Northwest University, and in part by the Natural Science Foundation of Jiangxi Province under Grant 20161BAB212033.

ABSTRACT Image registration is an important technique in many computer vision applications such as image fusion, image retrieval, object tracking, face recognition, change detection and so on. Local feature descriptors, i.e., how to detect features and how to describe them, play a fundamental and important role in image registration process, which directly influence the accuracy and robustness of image registration. This paper mainly focuses on the variety of local feature descriptors including some theoretical research, mathematical models, and methods or algorithms along with their applications in the context of image registration. The existing local feature descriptors are roughly classified into six categories to demonstrate and analyze comprehensively their own advantages. The current and future challenges of local feature descriptors are discussed. The major goal of the paper is to present a unique survey of the state-of-the-art image matching methods based on feature descriptor, from which future research may benefit.

INDEX TERMS Local feature descriptor, image matching, point pattern matching, pattern recognition.

I. INTRODUCTION

Image matching is used to determine to geometrically align two or more images of the same scene taken from different viewpoints at the same or different times by the same or different sensors; this is an important task in image processing and has been widely applied in the computer vision and pattern recognition. Image matching methods are classified into two different types i.e., area-based method and feature-based method [1]. Area-based methods deal with the images and used for checking the similarity of the pixels between the reference image and the sensed image without detecting salient features by using the optimization algorithms [2]–[4]. This method has some intrinsic limitations due to the intensity distribution, varying illumination, and geometric deformations that are caused by noise. Feature-based methods directly use the salient features that is extracted from two images instead of image intensity values that is more suitable for illuminated change and complicated geometric deformation [5], [6] and has also been widely used in image matching.

Feature-based image registration methods consists four main steps [7]: First feature extraction (feature descriptor), second feature correspondence, third transformation estimation and the final one is resampling. Feature extraction and feature correspondence require lots of manipulation techniques [8], image registration is the difficult part and is the most important step for accurate feature correspondence. Any problem in the feature extraction will result in incorrect correspondences and incorrect transformation function that will give wrong registration results. Feature extraction or feature descriptor is the main problem for inaccurate image registration.

Image matching was developed in 1980 by Moravec [9], which is repeatable under small variations and near edges and was applied for stereo matching; But the Moravec detector was not rotation invariance and sensitive to noise. Harris and Stephens [10] developed the Harris corner detector in 1988, by improving the Moravec detector, that consist the gradient information and the eigenvalues of symmetric positive definite 2×2 matrix to make it more repeatable, that is widely applied for image matching tasks. Harris corner detector is sensitive to scale, that does not result in a good basis for image matching of different sizes [11]. Smith and Brady [12] presented SUSAN operator in 1997; that is was not sensitive to local noise and has high anti-interference ability. SIFT is the one of the mostly widely used descriptor that was developed by Lowe [11], Scale invariant feature transform (SIFT) has best performance in the context of matching and recognition due to its invariance to rotation, scale and translation [13]. Ke and Sukthankar [14] presented the PCA-SIFT descriptor that represents the local appearance by principal components of the normalized gradient field. Mikolajczyk and Schmid proposed the Harris-Laplace and Harris-Affine detectors for scale and affine invariant, that deal with larger scale changes and also provide for reliable matching even for images with significant perspective deformations [15]. Mikolajczyk and Schmid proposed a new descriptor that is named Gradient Location and Orientation Histogram (GLOH) by exploring SIFT by changing the location grid and using PCA to reduce the size. Lazebnik *et al.* [16] suggested a sparse texture representation descriptor by using local affine regions called the Rotation Invariant Feature Transform (RIFT). Bay *et al.* [17] revealed that Speed-Up Robust Features (SURF) is an efficient implementation of SIFT by applying the integral image to compute image derivations, and quantifying the gradient orientations in a small number of histogram bins [18]. Wu *et al.* [19], [20] suggested a new practices for learning method that is based on fisher vectors (FV) and vectors of locally aggregated descriptors (VLAD), that achieve high accuracy for good practices for video encoding or action recognition in videos. Lin *et al.* [21] suggested a tube-and-droplet-based representation approach to indicate the global motion pattern in practical applications such as trajectory clustering, trajectory classification, abnormality detection and 3D action recognition.

Recently, different important local feature descriptors are proposed for matching and recognition from binary descriptors that is based on pixel intensity comparisons, such as BRISK [22], ORB [23], BRIEF [24], LDAHash [25] and pooling configuration methods based on location and shape of the regions [26]–[28], which are fast for both descriptor construction and matching. Binary descriptors are used for comparing the intensities of pixels sampled at different locations or mapping the local descriptor into the Hamming space that is more efficient to compute [29], and in order to make pooling configuration more tractable, the pooling configuration needs to be considered and restricted to circular, and symmetrically arranged pooling regions centered about the patch to improve the performance by the convex optimization [28]. Rublee *et al.* [23] stated that the ORB descriptor is invariant to rotation changes and robust to noise that is significantly faster than SIFT in many situations. ORB uses binary tests by learning method that decrease the correlation among the binary tests by improving the performance and scalability [30]. Trzcinski *et al.* [28] suggested a new and general framework to learn highly discriminative and compact local feature descriptors with boosting that leverages the boosting-trick to simultaneously optimize for both weighting and sampling strategy respectively for nonlinear feature responses. Chen *et al.* [18] studied a simple and powerful local descriptor called Weber Local Descriptor (WLD) which is based on the Weber's Law according to the perception of human beings. WLD contains two components [18]: differential excitation and orientation. The differential excitation component is a function of the ratio between two terms: one is the relative intensity differences of a current pixel against its neighbors; the other one is the intensity of the current pixel. The orientation component is the gradient orientation of the current pixel. Liu *et al.* [31] developed the local descriptor called Weber Local Binary Pattern (WLBP) by combining the advantages of WLD and Local Binary Pattern (LBP), that includes two components: differential excitation and LBP. The 2D histogram of WLBP is computed by encoding patterns of differential excitation and LBP, that have high performance as compared to LBP and WLD. Liu *et al.* [32] proposed a new depth descriptor called Geodesic Invariant Feature (GIF) that is multilevel feature representation framework that is based on the nature of depth images and can encode the local structures in the depth data.

The local feature descriptors play an important role in computer vision and pattern recognition problems including graph matching [33]–[35], object recognition or clustering [36]–[38], image retrieval [39]–[41], object tracking [42], [43], face recognition [44]–[46], and change detection [47]. The local feature descriptors describe keypoints with distinctiveness, repeatability, compactness, accuracy and efficient representations which are invariant and robust to scale, rotation, affine transformation, occlusion, and illumination [48], [49].

Recently, a large number of researchers have significantly contributed on local feature descriptors for image matching and recognition and so on. To our knowledge, there is no review paper in the literature that comprehensively analyzes these local feature descriptors for summarization. The main objective of this paper was to present an insight analysis framework on the latest findings of local feature descriptors and promote further research on these aspects in computer vision and pattern recognition, especially in the context of image registration, and also provide the most recent and advanced innovations from which the researchers can benefit the state-of-the-art local feature descriptors to present a number of improvements. In this paper, the existing local feature descriptors are classified into six categories:

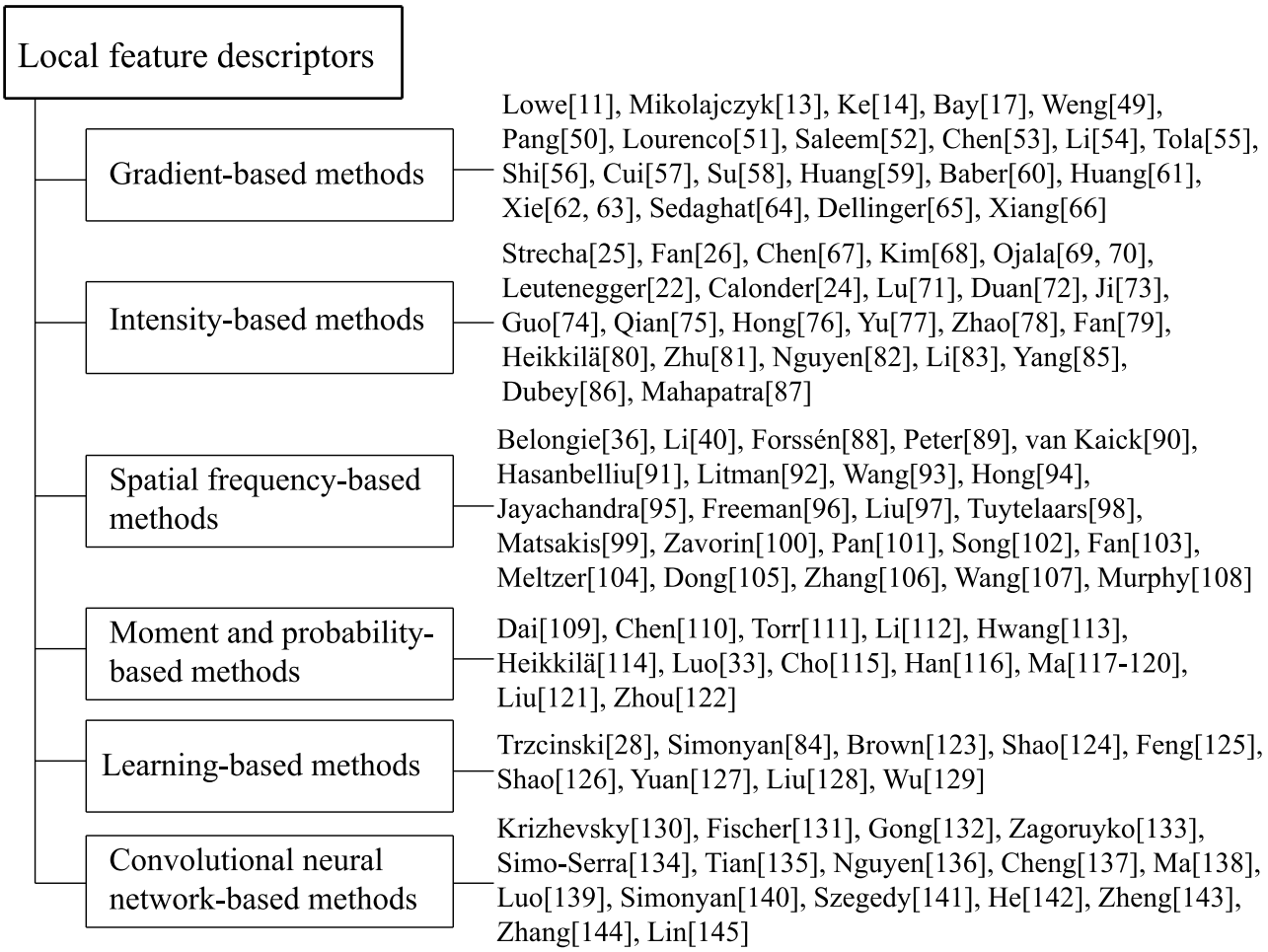


FIGURE 1. The categorization of local feature descriptors and algorithms.

gradient-based methods, intensity-based methods, spatial frequency-based methods, moment and probability-based methods, learning-based methods and convolutional neural network-based methods as shown in Fig. 1.

The rest of this paper is organized as follows: first we provide a comprehensive survey of the existing local feature descriptors for matching and recognition in detail. Finally we present the conclusion with some future recommendations for future research.

II. LOCAL FEATURE DESCRIPTORS

Feature detection is the first main step of the local feature descriptors that is based for image matching system. The simplest and the most commonly used keypoint detection method is Harris corner detector without using local descriptor. Harris corner detector does not result in qualified keypoints in term of repeatability and informativeness for big rotation and scale change images because the Harris corner detector takes little consideration to the richness of discriminative information of these keypoints detected. It is necessary to encode the keypoints extracted into a representative feature descriptor that is based on the local geometric information to improve their distinctiveness.

A. GRADIENT-BASED METHODS

Image gradient-based methods have received intense attention of the researchers due to its promising performance in a variety of applications. Lowe [11] stated that the scale invariant feature transform (SIFT) is based on the gradient distribution in the detected regions, which is the mostly classical and is invariant to scale, rotation and viewpoint change. There are main four steps to generate the SIFT algorithm: (1) Scale-space extreme detection. The first stage is to identify potential interest points that are invariant to scale and orientation by applying Difference of Gaussian function. (2) Keypoints localization. The location and the scale of each candidate point is based on measures of stability that is determined. (3) Orientation assignment is the stage in which one or more orientations are assigned to each keypoint location that is based on the local image gradient directions. (4) Keypoint descriptor is the local feature descriptor; that is created by computing the gradient magnitude and orientation in the region around each keypoint. However, it is the computationally demanding. Gradient location and orientation histogram (GLOH) is developed by Mikolajczyk and Schmid [13], that is an extension of the SIFT descriptor by changing the location of the grid and using PCA to decrease the size in order

to increase its robustness and distinctiveness. Ke and Sukthankar [14] stated that the PCA-SIFT descriptor is similar to the SIFT descriptor, that represents the local appearance by applying the principal components analysis (PCA) to the normalized image gradient patch. Speeded-up robust features (SURF) was proposed by Bay *et al.* [17] that has fast speed by describing keypoints with the response of a few Haar-like filter, but the SURF does away with SIFT's spatial weighting scheme which produces damaging artifacts [49].

Weng *et al.* [49] developed a new local image descriptor that is named distinctive efficient robust features (DERF); that is used for modeling the response and distribution properties of the parvocellular-projecting ganglion cells in the primate retina. Pang *et al.* [50] suggested a fully affine invariant SURF algorithm, which makes full use of the affine invariant advantage of affine SIFT (ASIFT) and the efficient merit of SURF, that avoid their drawbacks, and the experimental results on optical image matching demonstrate the proposed algorithm was robust. Several modifications in the SIFT model have proposed [51]–[54] in order to improve the repeatability of detection and effectiveness of matching under nonlinear intensity changes and big distortion. Tola *et al.* [55] introduced a DAISY descriptor by switching the weighted sums of gradient norms by orientation maps that are convolutions of the gradients in specific directions with several Gaussian filters, to make DAISY faster and more robust to affine transformation and brightness change between images. Because convolutions with a large Gaussian kernel can be obtained from several consecutive convolutions by using smaller kernels, and thus coordination maps for different sizes can be computed at a low cost [56]. Cui and Ngan [57] proposed the scale and the affine invariant Fan feature that described by Fan-SIFT which is based on the affine shape diagnosis of the mirror predicted surface patch and is made by using the automatic sale selection method based on the Fan Laplacian of Gaussian (FLOG). Su *et al.* [58] suggested the histogram of oriented gradient (HOG) descriptor, that considered weight for every bin of gradient orientation histogram according to the significance of the gradient information. Huang *et al.* [59] suggested the local main gradient and the tracklets-based features method by applying the statistics of tracklets to describe head-and-shoulder shapes to improve the detection performance. Baber *et al.* [60] suggested the BIG-OH, a simple method for binary quantization of any descriptor based on the gradient orientation histogram by computing a bit vector representing the relative magnitudes of local gradients associated with neighboring orientation bins. BIG-OH needs small memory requirements and only requires 16 bytes per descriptor for the construction of the gradient orientation histogram by applying the SIFT default parameters. Huang *et al.* [61] suggested a novel and powerful local image descriptor for making use of the histograms of the second order gradients (HSOGs) to capture curvature related local geometric properties. In order to reduce the time and improve the recognition accuracy, Xie *et al.* [62], [63] suggested the reversal invariant descriptor for local patterns to obtain the identical representation for an image and its left-right reversed copy, which has high performance for object recognition and image classification.

Sedaghat and Ebadi [64] suggested a new local feature descriptor that is named adaptive binning scale-invariant feature transform (AB-SIFT) by exploiting an adaptive binning strategy to describe the image content around a local feature and applying an adaptive histogram quantization strategy for both the location and gradient orientations respectively, that significantly increases the discriminability and robustness of the final descriptor. The local descriptors are based on the gradient information, that is widely used for image matching and registration including optical images, multispectral images [52], remote sensing images [64] and the synthetic aperture radar (SAR) images [65], [66]. These methods are more suitable for big geometric distortion and rotation and can achieve robust image registration. The feature descriptor is based on gradient methods, that leads to the slow speed and poor real-time performance.

B. INTENSITY-BASED METHODS

The image intensity-based methods are applied to compare the intensities of pixels sampled at different location or mapping the local descriptor into the Hamming space. Strecha *et al.* [25] suggested a new and simple approach to produce a binary string from a SIFT descriptor, named LDAHash, that aligns the SIFT descriptors according to the problem specific to covariance structure. The reliable thresholds can be estimated to perform the binarization according to an appropriate cost function. Fan *et al.* [26] proposed a new method for constructing interest region descriptor for the key idea to pool the sample local features into several groups based on the intensity orders in multiple support regions. A distinctive local feature descriptor named partial intensity invariant feature descriptor (PIIFD) [67] was proposed by Chen *et al.* In their framework, the corner points are detected instead of bifurcations; PIIFD are extracted for all corner points by following a bilateral matching technique to identify corresponding PIIFD matches; any unsuitable matches are removed and these unsuitable matches are refined.

In order to deal with illumination changes, many local features based on the intensity order have been proposed as compared to the raw intensity, because the intensity order of pixels in an image are invariant to monotonic changes of intensity [68]. Ojala *et al.* [69] firstly suggested the local binary pattern (LBP) operator, that creates an order based feature for pixel by comparing each pixel's intensity value with that of its neighboring pixels. The BRISK descriptor was presented by Leutenegger *et al.* [22]. BRISK descriptor is based on binary string by assigning location, scale and orientation clues for each sample points by concatenating the results of the simple brightness comparison tests, which is invariant to scale and rotation to obtain more compact and robust performance. A simple descriptor based on binary descriptor is called BRIEF that is proposed by Calonder *et al.* [24], in which each bit is computed by comparing the intensity

difference between a pair of sample points from the image patches. Despite the clear advantage in computation and storage, BRIEF still has weakness in terms of reliability and robustness.

The LBP is among the most widely used intensity based feature due to its computational simplicity [70], with applications in face recognition [71], [72], texture recognition or classification [73]–[77], video or receptive detection [78], [79], interest region description [80]–[82], and information retrieval [83]–[85]. LBP need to compare the gray-level intensity of a pixel with that of k of its neighbors at a pixel distance of r according to the LBP features for a given image patch, and LBP can obtain a binary vector expressing the relationship between the gray level intensity at the point of interest to each of its neighbors from the comparisons [60]. In order to decrease the dimension of the descriptors while considering a large number of local neighbors, Dubey *et al.* [86] suggested the interleaved order based local descriptor (IOLD) that merges the patters extracted for constructing the descriptor over each set separately to produce a single pattern based on the local neighbors of a pixel as a set of interleaved neighbors.

These methods can be widely used for multimodal retinal image registration [67], [87], optical image matching with different geometric and photometric transformations such as scale, rotation, blur, illumination, and JPEG compression, and textured scenes images [68], [80]. These methods have the following properties such as short histogram, tolerance to illumination changes and computational simplicity [80]. These methods also have some limitations due to the intensity distribution, varying illumination, and geometric deformations which are caused, for instance, by noise.

C. SPATIAL FREQUENCY-BASED METHODS

Spatial frequency-based methods are very important descriptors methods for image registration mainly including shape contexts [36], [40], [88]–[94], directionlets [95], steerable filters [96], [97], affine invariant [98], [99] and fractional fourier or wavelet transform [100]–[102]. Mathematicians typically define shape as an equivalence class under a group of transformations, that tells us when two shapes are exactly the same [36]. A survey on shape correspondence in computer vision, pattern recognition, medical image processing, and many other fields can be found in [90]. Tasks such as content-based image retrieval, face recognition, and image registration all require matching of features such as points, lines, and contours extracted from the reference and sensed images [103]–[107], i.e., finding the correspondence between two shapes is then equivalent to finding the point in each object with a similar shape context [91]. Litman and Bronstein [92] suggested a parametric spectral descriptors that takes into account the statistics of the corpus of shape to which it is applied and the class of transformations to which it is made insensitive. Wang *et al.* [93] developed a new shape descriptor method for leaf shape identification by extracting multiscale shape features using arch height features instead of curvatures for leaf shape description for discriminatively representing the shape of the leaf. Hong and Soatto [94] stated that a shape descriptor provides desirable invariance properties that are based on a series of isotropic integral kernels that characterize the local shape geometry by enabling the shape signature to be robust with respect to undesirable perturbation while retaining discriminative power. In addition, the shape signature is designed to be invariant with respect to group transformations that include rotation, scaling, translation, and reflection.

The steerable pyramid filters introduced by Freeman and Adelson [96], which are multi-scale, multi-orientation image decomposition methods, rotation and translation invariant to image distort, that can be widely applied for image registration. These characteristics make it useful for image registration such as remote sensing images and the registration results that are more robust under big distort based on the steerable pyramid filters [97]. Zavorin and Moigne [100] suggested a wavelet feature pyramids method for automatic registration exploiting the inherent multi-resolution character of the wavelet transformation, that can achieve a fast computational speeds and accurate registration. Pan *et al.* [101] proposed an adaptable-multilayer fractional Fourier transform approach by combining the polar Fast Fourier Transform (FFT) and the log-polar FFT, that has a lower interpolation error in both polar and log-polar Fourier transform and can reach better accuracy with the nearly same computing complexity as the pseudopolar FFT. Song and Li [102] proposed a new feature descriptor called the Local Polar DCT Feature (LPDF), which is robust to a variety of image transformation by directly extracting the DCT features from the local image patch quantized in the polar geometric structure.

These methods are more accurate and significantly faster, that is more suitable for point sets matching including deformation, noise, outliers, rotation and occlusion [91], optical image registration without any interpolation and iteration [101]. In addition, these methods are also applied to remote sensing image registration with geometrically warped, noisy and radio metrically warped [108]. Two major techniques in Fourier-based methods for image registration are the phase correlation and the log-polar transform, but both have poor performance and limited applications such as large scales with arbitrary rotations.

D. MOMENT AND PROBABILITY-BASED METHODS

Feature representations are invariant to rotation, scale, and translation in the matching process, general feature representations are chain code [109], moment invariants [110], and probability descriptors [111], [112]. Dai and Khorram [109] stated that the automated image registration method was establish to correspondences between the protentially matched regions detected by combining an invariant moment shape descriptor with improved chain-code matching. The Zernike moments are the extension of the geometric monents by replacing the conventional transform kernel with orthogonal Zernike polynomials [110]. Zernike moments are used

in the image registration and object recognition regardless of variations in size, position, and orientation [113], [114]. Chen and Sun [110] proposed a new descriptor called the Zernike moment phase-based descriptor by applying a common set of elliptical interest regions that are further normalized to circular with a fixed size. In addition, the normalized circular regions become affine invariant to a rotational ambiguity.

The likelihood function has a mixture-structure such as expectation-maximization (EM) algorithm that provide a principled way for recovering maximum likelihood solutions to problems posed in terms of missing or hidden data [33]. Luo and Hancock [33] stated from a probability distribution for errors and show that the problem of graph matching can be started as maximum likelihood estimation by applying the EM algorithm, and the correspondence matches between the graph models that can be found in a matrix framework which is based on singular value decomposition to improve the matching. Cho and Lee [115] proposed a progressive method to update the candidate matches by applying a move of graphs based probabilistic voting, that greatly boost the objective function in an integer quadratic programming problem. A High-order statistics of Weber local descriptors was proposed by Han *et al.* [116], that explore the local patch, called micro-Texton, transformed domain and employed a parametric probability process and extract the higher-order statistics to model the Weber local descriptors. Ma *et al.* [117]–[120] investigated about point matching methods based on the local features under the assumption that the point matches undergo a coherent transformation that can be iteratively estimated by the expectation maximization (EM) algorithm [121]. They presented a unified framework for non-rigid feature matching based on the transformation, and the underlying transformation between the point pairs is represented by vector field [117], [118], robust L_2E estimator [119], or using a Gaussian mixture model [120], that exploit both global and local structure to find the better correspondences.

Moment and probability-based methods are applied to point sets, textured scenes and structured scenes images for matching or registration including photometric and geometric transformations [110]. Moment and probability-based methods are suitable for remote sensing images, that are efficient and robust and are able to handle outliers [122]. The probability model depends on hidden variables, has a low convergence rate. As a consequence, it is not suitable for large-scale data sets and high dimensional data; but the computational structure of the algorithm is stable and accurate.

E. LEARNING-BASED METHODS

Learning-based methods are applied for higher level visual tasks that can be classified into two categories [49]: learning low-level features [28], [84], [123], [124] and the deep learning neural networks which provide more invariance to various distortions by learning multiple levels of feature including low-level feature to obtain higher-level features [125]–[127]. Trzcinski *et al.* [28] suggested a new supervised learning low-dimensional but highly discriminative descriptors, that is applied for boosting to obtain a non-linear mapping of the input to a high-dimensional feature space. Simonyan *et al.* [84] suggested a learning local feature descriptors as a convex optimization problem by applying sparsity. The proposed method can decrease the dimensionality as well as to improve discrimination of the descriptors by applying the Mahalanobis matrix nuclear norm regularization.

Leaning local image descriptors method was proposed by Brown *et al.* [123]. Leaning local image descriptors method is based on building blocks for constructing descriptors, which considers both linear and non-linear transforms with dimensionality reduction, and to make use of discriminant learning techniques and Powell minimization to minimize the error of a nearest neighbor classifier. Most existing image classification methods use hand-craft features, that are not adaptive for different image domains. In order to solve this problem, Shao *et al.* [124] suggested an evolutionary learning method to automatically generate domain adaptive global feature descriptor; that is based on multi-objective genetic programming (MOGP); that is applied to evolve robust and discriminative feature descriptors with a set of domain specific images and random constants as terminals, a number of primitive operators as a function, both the classification error rate and tree complexity as the fitness criterion. Yuan *et al.* [127] proposed a manifold standardized deep architecture method for recognition that is used to learn the high-level features in an unsupervised fashion by exploiting the structural information of data and making mapping between visible layer and hidden layer. Liu *et al.* [128] developed a motion feature descriptor based on genetic programming (GP) to evolve discriminative spatiotemporal representations, that simultaneously fuse the color and optical flow sequences for high-level action recognition.

Learning-based methods doesn't need manually labeled ground-truth data, for low-level and high-level features that are more flexible than conventional handcrafted features. In addition, the trained deep learning network selected features capture accurately the complex morphological patterns in the image patches, that improve the image registration performance on new image modalities or new imaging applications [129]. Deep learning can only provide limited amount of data in application scenarios, and cannot give unbiased estimation of data.

F. CONVOLUTIONAL NEURAL NETWORK (CNN)-BASED METHODS

End-to-end learning of patch descriptors based on deep learning was introduced by [130]– [135], the convolutional neural networks (CNN) have recently led to breakthroughs in computer vision and pattern recognition [136] such as objective detection [137], feature matching [138], and image classification [139] and so on. In 2012, Krizhesky *et al.* [130] suggested a CNN called AlexNet that exceeded the previous

results by a large margin in the ImageNet Large-Scale Visual Recognition Challenge (ILSVRC). CNN belong to specific category of deep learning methods, that become exceptionally popular method because CNN have not only been able to automatically learn image feature representations, but have also simply a scaled version of the LeNet with a deeper structure [136]. CNN models serve as good choices for extracting features such as VGGNet [140], GoogleNet [141] and ResNet [142], that are viewed as a set of non-linear functions that are composed of a number of layers including convolution, pooling, non-linearity [143].

Ma and Zhao [138] proposed a new feature matching method that is based on CNN feature as the holistic image representation. The CNN feature is applied to retrieve key-frames that have similar appearance from a topological map. Finally, the vector field consensus was used for geometric verification and to retrieve the most similar key-frame to improve the matching performance. Zhang *et al.* [144] proposed fine-grained recognition method, includes two steps of picking neural activations computed from the CNN, one for localization, and the other one for description. Zheng *et al.* [143] presented a local descriptor comparisons called SIFT meets CNN: This survey provides insights connection between SIFT and CNN-based methods for modern instance retrieval, reviews a broad selection in different categories. Lin *et al.* [145] suggested visual data matching method based on generalized similarity measure and feature learning. The similarity measure is unified with feature representation learning via deep CNN that incorporates the similarity measure matrix into the deep architecture, enabling an end-to-end way of model optimization. Wen *et al.* [146] introduced a new deep color guided coarse-to-fine convolutional neural network (CNN) framework to alleviate texture copying artifacts and preserve edge details effectively for depth image super-resolution.

CNN-based methods are widely applied to computer vision and patter recognition, CNN features can be extracted in an end-to-end manner through a single pass to the CNN model [143]. The feature descriptor representations exhibit improved discriminative ability for image matching or image registration. CNN-based methods have some limitations such as adjusting parameters, large samples and using GPUs to train samples models. The physical meaning is not clear and neural network itself is an inexplicable ''black box mode''.

There are some special local features, i.e., feature encoding methods including Bag-of-Words (BoW) [147], Fisher Vector (FV) [148], Vector of Locally Aggregated Descriptors (VLAD) [149] and so on, that are widely applied to image classification [150]–[154], object localization [155], face recognition [156], crowd counting [157] and video analysis [158].

Yuan and Hu [150] applied the bag-of-words (BoW) model to construct the compact feature vectors from densely extracted local features for automatic cloud extraction and image classification. Liu *et al.* [153] used the pre-trained CNN activations as local features that is proposed for particular compositional model based on the fisher vector (FV) coding for image classification. Li *et al.* [154] proposed a multiple vector for locally aggregated descriptors (VLAD) encoding method with CNN features for image classification. Wang *et al.* [155] proposed a scheme for instance annotation inspired by the successful application of bag-of-words (BoW) to feature representation to incorporate the BoW learning and instance labeling in a single optimization formulation for object localization. Wang *et al.* [156] developed the Compact FV (CFV) descriptor that is obtained by zeroing out small posteriors, calculating first-order statistics and reweighting its elements properly to apply CFV to encode convolutional activations of CNN for face recognition. By taking into consideration diverse coefficient weights, Sheng *et al.* [157] proposed a generalized form of weighted-VLAD (W-VLAD) with the help of a CNN for accurate crowd counting. Xu *et al.* [158] developed a new sequential vector of locally aggregated descriptor (VLAD) layer to combine the recurrent convolution networks (RCNs) architecture into a whole framework to improve feature extraction and motion analysis.

III. CONCLUSION

This paper aims to investigate the insight analysis of framework on the latest local feature descriptors and promote further research on the same aspects in computer vision applications to provide most recent and advanced innovations for researchers.

The local feature descriptor is a key technique that plays an essential role in computer vision and pattern recognition. However, there are still some issues that remained un-solved for the local feature descriptions as follows:

- The detection of stable features is the first step for image registration, that directly affect the accuracy for feature correspondence. Therefore, it is suggested that in future researchers are suggested to detect reference image and sensed image to solve this problem.
- Describing the features detected is a critical and very difficult step, and the features by local description should be invariant and robust to affine, scale, rotation, occlusion and illumination.
- To find the accurate feature correspondence is very important step, that measure criterions and optimization algorithms which are used to accelerate and improve image registration.
- How to choose the suitable transformation function is an essential step, because some images are global, local or global and local deformation.

In summary, local feature descriptors play an essential role in many computer vision applications for future in image registration, image fusion, image retrieval, object recognition and change detection.

CONFLICT OF INTERESTS

The authors declare that there is no conflict of interests regarding the publication of this paper.

REFERENCES

[1] B. Zitová and J. Flusser, "Image registration methods: A survey," *Image Vis. Comput.*, vol. 21, no. 11, pp. 977–1000, 2003.

[2] W. K. Pratt, "Correlation techniques of image registration," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-10, no. 3, pp. 353–358, May 1974.

[3] P. Viola and W. M. Wells, III, "Alignment by maximization of mutual information," *Int. J. Comput. Vis.*, vol. 24, no. 2, pp. 137–154, Sep. 1997.

[4] A. Myronenko and X. Song, "Intensity-based image registration by minimizing residual complexity," *IEEE Trans. Med. Imag.*, vol. 29, no. 11, pp. 1882–1891, Nov. 2010.

[5] X. Liu, Z. Tian, and M. T. Ding, "A novel adaptive weights proximity matrix for image registration based on R-SIFT," *AEU-Int. J. Electron. Commun.*, vol. 65, no. 12, pp. 1040–1049, 2011.

[6] C. Leng, J. Xiao, M. Li, and H. P. Zhang, "Robust adaptive principal component analysis based on intergraph matrix for medical image registration," *Comput. Intell. Neurosci.*, vol. 2015, Mar. 2015, Art. no. 829528.

[7] L. G. Brown, "A survey of image registration techniques," *ACM Comput. Surv.*, vol. 24, no. 4, pp. 325–376, Dec. 1992.

[8] X. Duan, Z. Tian, M. Ding, and W. Zhao, "Registration of remote-sensing images using robust weighted kernel principal component analysis," *AEU-Int. J. Electron. Commun.*, vol. 67, no. 1, pp. 20–28, Jan. 2013.

[9] H. P. Moravec, "Rover visual obstacle avoidance," in *Proc. 7th Int. Joint Conf. Artif. Intell.*, Vancouver, BC, Canada, 1981, pp. 785–790.

[10] C. Harris and M. Stephens, "A combined corner and edge detector," in *Proc. Alvey Vis. Conf.*, 1988, vol. 15, no. 50, pp. 147–151.

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

[12] S. M. Smith and J. M. Brady, "SUSAN—A new approach to low level image processing," *Int. J. Comput. Vis.*, vol. 23, no. 1, pp. 45–78, 1997.

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

[14] Y. Ke and R. Sukthankar, "PCA-SIFT: A more distinctive representation for local image descriptors," in *Proc. IEEE Comput. Vis. Pattern Recognit.*, vol. 2, Jun./Jul. 2004, pp. II-506–II-513.

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

[16] S. Lazebnik, C. Schmid, and J. Ponce, "A sparse texture representation using local affine regions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 8, pp. 1265–1278, Aug. 2005.

[17] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (SURF)," *Comput. Vis. Image Understand.*, vol. 110, no. 3, pp. 346–359, 2008.

[18] J. Chen *et al.*, "WLD: A robust local image descriptor," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 9, pp. 1705–1720, Sep. 2010.

[19] J. Wu, Y. Zhang, and W. Lin, "Towards good practices for action video encoding," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2014, pp. 2577–2584.

[20] J. Wu, Y. Zhang, and W. Lin, "Good practices for learning to recognize actions using FV and VLAD," *IEEE Trans. Cybern.*, vol. 46, no. 12, pp. 2978–2990, Dec. 2016.

[21] W. Lin *et al.*, "A tube-and-droplet-based approach for representing and analyzing motion trajectories," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 8, pp. 1489–1503, Aug. 2016.

[22] S. Leutenegger, M. Chli, and R. Y. Siegwart, "BRISK: Binary robust invariant scalable keypoints," in *Proc. IEEE Int. Conf. Comput. Vis.*, Nov. 2011, pp. 2548–2555.

[23] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "ORB: An efficient alternative to SIFT or SURF," in *Proc. IEEE Int. Conf. Comput.*, Nov. 2011, pp. 2564–2571.

[24] M. Calonder, V. Lepetit, M. Ozuysal, T. Trzcinski, C. Strecha, and P. Fua, "BRIEF: Computing a local binary descriptor very fast," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 7, pp. 1281–1298, Jul. 2012.

[25] C. Strecha, A. M. Bronstein, M. M. Bronstein, and P. Fua, "LDAHash: Improved matching with smaller descriptors," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 1, pp. 66–78, Jan. 2012.

[26] B. Fan, F. Wu, and Z. Hu, "Rotationally invariant descriptors using intensity order pooling," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 10, pp. 2031–2045, Oct. 2012.

[27] T. Trzcinski, M. Christoudias, P. Fua, and V. Lepetit, "Boosting binary keypoint descriptors," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 2874–2881.

[28] T. Trzcinski, M. Christoudias, and V. Lepetit, "Learning image descriptors with boosting," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 3, pp. 597–610, Mar. 2015.

[29] S. Zhang, Q. Tian, Q. Huang, W. Gao, and Y. Rui, "USB: Ultrashort binary descriptor for fast visual matching and retrieval," *IEEE Trans. Image Process.*, vol. 23, no. 8, pp. 3671–3683, Aug. 2014.

[30] X. Xu, L. Tian, J. Feng, and J. Zhou, "OSRI: A rotationally invariant binary descriptor," *IEEE Trans. Image Process.*, vol. 23, no. 7, pp. 2983–2995, Jul. 2014.

[31] F. Liu, Z. Tang, and J. Tang, "WLBP: Weber local binary pattern for local image description," *Neurocomputing*, vol. 120, pp. 325–335, Nov. 2013.

[32] Y. Liu, P. Lasang, M. Siegel, and Q. Sun, "Geodesic invariant feature: A local descriptor in depth," *IEEE Trans. Image Process.*, vol. 24, no. 1, pp. 236–248, Jan. 2015.

[33] B. Luo and E. R. Hancock, "Structural graph matching using the EM algorithm and singular value decomposition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 10, pp. 1120–1136, Oct. 2001.

[34] C. Leng, Z. Tian, J. Li, and M. Ding, "Image registration based on matrix perturbation analysis using spectral graph," *Chin. Opt. Lett.*, vol. 7, no. 11, pp. 996–1000, 2009.

[35] C. Leng, W. Xu, I. Cheng, and A. Basu, "Graph matching based on stochastic perturbation," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 4862–4875, Dec. 2015.

[36] 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.

[37] L. He and H. Zhang, "Kernel K-means sampling for Nyström approximation," *IEEE Trans. Image Process.*, vol. 27, no. 5, pp. 2108–2120, May 2018.

[38] L. He, N. Ray, Y. Guan, and H. Zhang, "Fast large-scale spectral clustering via explicit feature mapping," *IEEE Trans. Cybern.*, to be published, doi: 10.1109/TCYB.2018.2794998.

[39] J. Yang, J. Liang, H. Shen, K. Wang, P. L. Rosin, and M.-H. Yang, "Dynamic match kernel with deep convolutional features for image retrieval," *IEEE Trans. Image Process.*, vol. 27, no. 11, pp. 5288–5302, Nov. 2018.

[40] Y. Li, Y. Zhang, X. Huang, H. Zhu, and J. Ma, "Large-scale remote sensing image retrieval by deep hashing neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 2, pp. 950–965, Feb. 2018.

[41] J. Chen, J. Ma, C. Yang, L. Ma, and S. Zheng, "Non-rigid point set registration via coherent spatial mapping," *Signal Process.*, vol. 106, pp. 62–72, Jan. 2015.

[42] K. Nai, Z. Li, G. Li, and S. Wang, "Robust object tracking via local sparse appearance model," *IEEE Trans. Image Process.*, vol. 27, no. 10, pp. 4958–4970, Oct. 2018.

[43] F. Pernici and A. D. Bimbo, "Object tracking by oversampling local features," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 12, pp. 2538–2551, Dec. 2014.

[44] C. Ding and D. Tao, "Trunk-branch ensemble convolutional neural networks for video-based face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 1002–1014, Apr. 2018.

[45] S. A. Khan, A. Hussain, M. Usman, M. Nazir, N. Riaz, and A. M. Mirza, "Robust face recognition using computationally efficient features," *J. Intell. Fuzzy Syst.*, vol. 27, no. 6, pp. 3131–3143, 2014.

[46] J. Ma, J. Zhao, Y. Ma, and J. Tian, "Non-rigid visible and infrared face registration via regularized Gaussian fields criterion," *Pattern Recognit.*, vol. 48, no. 3, pp. 772–784, 2015.

[47] Z. Zhang, Z. Tian, M. Ding, and A. Basu, "Improved robust kernel subspace for object-based registration and change detection," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 4, pp. 791–795, Jul. 2013.

[48] M. Amiri and H. R. Rabiee, "RASIM: A novel rotation and scale invariant matching of local image interest points," *IEEE Trans. Image Process.*, vol. 20, no. 12, pp. 3580–3591, Dec. 2011.

[49] D. Weng, Y. Wang, M. Gong, D. Tao, H. Wei, and D. Huang, "DERF: Distinctive efficient robust features from the biological modeling of the P ganglion cells," *IEEE Trans. Image Process.*, vol. 24, no. 8, pp. 2287–2302, Aug. 2015.

[50] Y. Pang, W. Li, Y. Yuan, and J. Pan, "Fully affine invariant SURF for image matching," *Neurocomputing*, vol. 85, pp. 6–10, May 2012.

[51] M. Lourenco, J. P. Barreto, and F. Vasconcelos, "sRD-SIFT: Keypoint detection and matching in images with radial distortion," *IEEE Trans. Robot.*, vol. 28, no. 3, pp. 752–760, Jun. 2012.

[52] S. Saleem and R. Sablatnig, "A robust SIFT descriptor for multispectral images," *IEEE Signal Process. Lett.*, vol. 21, no. 4, pp. 400–403, Apr. 2014.
[53] C.-C. Chen and S.-L. Hsieh, "Using binarization and hashing for efficient SIFT matching," *J. Vis. Commun. Image Represent.*, vol. 30, pp. 86–93, Jul. 2015.
[54] Q. Li, G. Wang, J. Liu, and S. Chen, "Robust scale-invariant feature matching for remote sensing image registration," *IEEE Geosci. Remote Sens. Lett.*, vol. 6, no. 2, pp. 287–291, Apr. 2009.
[55] E. Tola, V. Lepetit, and P. Fua, "DAISY: An efficient dense descriptor applied to wide-baseline stereo," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 5, pp. 815–830, May 2010.
[56] Q. Shi, G. Ma, F. Zhang, W. Chen, Q. Qin, and H. Duo, "Robust image registration using structure features," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 12, pp. 2045–2049, Dec. 2014.
[57] C. Cui and K. N. Ngan, "Scale- and affine-invariant fan feature," *IEEE Trans. Image Process.*, vol. 20, no. 6, pp. 1627–1640, Jun. 2011.
[58] X. Su, W. Lin, X. Zheng, X. Han, H. Chu, and X. Zhang, "A new local-main-gradient-orientation HOG and contour differences based algorithm for object classification," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, May 2013, pp. 2892–2895.
[59] K. Huang *et al.*, "Improved human head and shoulder detection with local main gradient and tracklets-based feature," in *Proc. Asia-Pacific Signal Inf. Process. Assoc. Annu. Summit Conf. (APSIPA)*, Dec. 2014, pp. 1–4.
[60] J. Baber, M. N. Dailey, S. Satoh, N. Afzulpurkar, and M. Bakhtyar, "BIG-OH: BInarization of gradient orientation histograms," *Image Vis. Comput.*, vol. 32, no. 11, pp. 940–953, Nov. 2014.
[61] D. Huang, C. Zhu, Y. Wang, and L. Chen, "HSOG: A novel local image descriptor based on histograms of the second-order gradients," *IEEE Trans. Image Process.*, vol. 23, no. 11, pp. 4680–4695, Nov. 2014.
[62] L. Xie, J. Wang, W. Lin, B. Zhang, and Q. Tian, "RIDE: Reversal invariant descriptor enhancement," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 100–108.
[63] L. Xie, J. Wang, W. Lin, B. Zhang, and Q. Tian, "Towards reversal-invariant image representation," *Int. J. Comput. Vis.*, vol. 123, no. 2, pp. 226–250, 2017.
[64] A. Sedaghat and H. Ebadi, "Remote sensing image matching based on adaptive binning SIFT descriptor," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 10, pp. 5283–5293, Oct. 2015.
[65] F. Dellinger, J. Delon, Y. Gousseau, J. Michel, and F. Tupin, "SAR-SIFT: A SIFT-like algorithm for SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 1, pp. 453–466, Jan. 2015.
[66] Y. Xiang, F. Wang, and H. You, "OS-SIFT: A robust SIFT-like algorithm for high-resolution optical-to-SAR image registration in suburban areas," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 6, pp. 3078–3090, Jun. 2018.
[67] J. Chen, J. Tian, N. Lee, J. Zheng, R. T. Smith, and A. F. Laine, "A partial intensity invariant feature descriptor for multimodal retinal image registration," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 7, pp. 1707–1718, Jul. 2010.
[68] B. Kim, H. Yoo, and K. Sohn, "Exact order based feature descriptor for illumination robust image matching," *Pattern Recognit.*, vol. 46, no. 12, pp. 3268–3278, 2013.
[69] T. Ojala, M. Pietikäinen, and D. Harwood, "A comparative study of texture measures with classification based on featured distributions," *Pattern Recognit.*, vol. 29, no. 1, pp. 51–59, 1996.
[70] T. Ojala, M. Pietikäinen, and T. Mäenpää, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, Jul. 2002.
[71] J. Lu, V. E. Liong, and J. Zhou, "Simultaneous local binary feature learning and encoding for homogeneous and heterogeneous face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 8, pp. 1979–1993, Aug. 2018.
[72] Y. Duan, J. Lu, J. Feng, and J. Zhou, "Context-aware local binary feature learning for face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 5, pp. 1139–1153, May 2018.
[73] L. Ji, Y. Ren, G. Liu, and X. Pu, "Training-based gradient LBP feature models for multiresolution texture classification," *IEEE Trans. Cybern.*, vol. 48, no. 9, pp. 2683–2696, Sep. 2018.
[74] Z. Guo and D. Zhang, "A completed modeling of local binary pattern operator for texture classification," *IEEE Trans. Image Process.*, vol. 19, no. 6, pp. 1657–1663, Jan. 2010.
[75] X. Qian, X.-S. Hua, P. Chen, and L. Ke, "PLBP: An effective local binary patterns texture descriptor with pyramid representation," *Pattern Recognit.*, vol. 44, nos. 10–11, pp. 2502–2515, Oct./Nov. 2011.
[76] X. Hong, G. Zhao, M. Pietikäinen, and X. Chen, "Combining LBP difference and feature correlation for texture description," *IEEE Trans. Image Process.*, vol. 23, no. 6, pp. 2557–2568, Jun. 2014.
[77] Y.-F. Yu, C.-X. Ren, D.-Q. Dai, and K.-K. Huang, "Kernel embedding multiorientation local pattern for image representation," *IEEE Trans. Cybern.*, vol. 48, no. 4, pp. 1124–1135, Apr. 2018.
[78] G. Zhao, T. Ahonen, J. Matas, and M. Pietikäinen, "Rotation-invariant image and video description with local binary pattern features," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 1465–1477, Apr. 2012.
[79] B. Fan, Q. Kong, T. Trzcinski, Z. Wang, C. Pan, and P. Fua, "Receptive fields selection for binary feature description," *IEEE Trans. Image Process.*, vol. 23, no. 6, pp. 2583–2595, Jun. 2014.
[80] M. Heikkilä, M. Pietikäinen, and C. Schmid, "Description of interest regions with local binary patterns," *Pattern Recognit.*, vol. 42, no. 3, pp. 425–436, 2009.
[81] C. Zhu, C.-E. Bichot, and L. Chen, "Image region description using orthogonal combination of local binary patterns enhanced with color information," *Pattern Recognit.*, vol. 46, no. 7, pp. 1949–1963, Jul. 2013.
[82] T.-N. Nguyen and K. Miyata, "Multi-scale region perpendicular local binary pattern: An effective feature for interest region description," *Vis. Comput.*, vol. 31, no. 4, pp. 391–406, 2015.
[83] Z. Li, G. Liu, Y. Yang, and J. You, "Scale- and rotation-invariant local binary pattern using scale-adaptive texton and subuniform-based circular shift," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 2130–2140, Apr. 2012.
[84] K. Simonyan, A. Vedaldi, and A. Zisserman, "Learning local feature descriptors using convex optimisation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 8, pp. 1573–1585, Aug. 2014.
[85] X. Yang and K. T. Cheng, "Local difference binary for ultrafast and distinctive feature description," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 1, pp. 188–194, Jan. 2014.
[86] S. R. Dubey, S. K. Singh, and R. K. Singh, "Rotation and illumination invariant interleaved intensity order-based local descriptor," *IEEE Trans. Image Process.*, vol. 23, no. 12, pp. 5323–5333, Dec. 2014.
[87] D. Mahapatra and Y. Sun, "MRF-based intensity invariant elastic registration of cardiac perfusion images using saliency information," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 4, pp. 991–1000, Apr. 2011.
[88] P. E. Forssén and D. G. Lowe, "Shape descriptors for maximally stable extremal regions," in *Proc. IEEE 11th Int. Conf. Comput. Vis.*, Oct. 2007, pp. 1–8.
[89] A. M. Peter and A. Rangarajan, "Information geometry for landmark shape analysis: Unifying shape representation and deformation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 337–350, Feb. 2009.
[90] O. van Kaick, H. Zhang, G. Hamarneh, and D. Cohen-Or, "A survey on shape correspondence," *Comput. Graph. Forum*, vol. 30, no. 6, pp. 1681–1707, 2011.
[91] E. Hasanbelliu, L. S. Giraldo, and J. C. Príncipe, "Information theoretic shape matching," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 12, pp. 2436–2451, Dec. 2014.
[92] R. Litman and A. M. Bronstein, "Learning spectral descriptors for deformable shape correspondence," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 1, pp. 171–180, Jan. 2014.
[93] B. Wang, D. Brown, Y. Gao, and J. L. Salle, "MARCH: Multiscale-arch-height description for mobile retrieval of leaf images," *Inf. Sci.*, vol. 302, pp. 132–148, May 2015.
[94] B.-W. Hong and S. Soatto, "Shape matching using multiscale integral invariants," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 1, pp. 151–160, Jan. 2015.
[95] D. Jayachandra and A. Makur, "Directionlets using in-phase lifting for image representation," *IEEE Trans. Image Process.*, vol. 23, no. 1, pp. 240–249, Jan. 2014.
[96] W. T. Freeman and E. H. Adelson, "The design and use of steerable filters," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 13, no. 9, pp. 891–906, Sep. 1991.
[97] Z. Liu, Y. K. Ho, K. Tsukada, K. Hanasaki, Y. Dai, and L. Li, "Using multiple orientational filters of steerable pyramid for image registration," *Inf. Fusion*, vol. 3, no. 3, pp. 203–214, Sep. 2002.
[98] T. Tuytelaars and L. Van Gool, "Matching widely separated views based on affine invariant regions," *Int. J. Comput. Vis.*, vol. 59, no. 1, pp. 61–85, 2004.

[99] P. Matsakis, J. M. Keller, O. Sjahputera, and J. Marjamaa, "The use of force histograms for affine-invariant relative position description," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 1, pp. 1–18, Jan. 2004.

[100] I. Zavorin and J. Le Moigne, "Use of multiresolution wavelet feature pyramids for automatic registration of multisensor imagery," *IEEE Trans. Image Process.*, vol. 14, no. 6, pp. 770–782, Jun. 2005.

[101] W. Pan, K. Qin, and Y. Chen, "An adaptable-multilayer fractional Fourier transform approach for image registration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 3, pp. 400–414, Mar. 2009.

[102] T. Song and H. Li, "Local polar DCT features for image description," *IEEE Signal Process. Lett.*, vol. 20, no. 1, pp. 59–62, Jan. 2013.

[103] B. Fan, F. Wu, and Z. Hu, "Robust line matching through line–point invariants," *Pattern Recognit.*, vol. 45, no. 2, pp. 794–805, 2012.

[104] J. Meltzer and S. Soatto, "Edge descriptors for robust wide-baseline correspondence," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2008, pp. 1–8.

[105] X. Dong and J. Dong, "The visual word booster: A spatial layout of words descriptor exploiting contour cues," *IEEE Trans. Image Process.*, vol. 27, no. 8, pp. 3904–3917, Aug. 2018.

[106] L. Zhang and R. Koch, "An efficient and robust line segment matching approach based on LBD descriptor and pairwise geometric consistency," *J. Vis. Commun. Image Represent.*, vol. 24, no. 7, pp. 794–805, 2013.

[107] Z. Wang, Z. Tang, and X. Zhang, "Reflection symmetry detection using locally affine invariant edge correspondence," *IEEE Trans. Image Process.*, vol. 24, no. 4, pp. 1297–1301, Apr. 2015.

[108] J. M. Murphy, J. Le. Moigne, and D. J. Harding, "Automatic image registration of multimodal remotely sensed data with global shearlet features," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 3, pp. 1685–1704, Mar. 2016.

[109] X. Dai and S. Khorram, "A feature-based image registration algorithm using improved chain-code representation combined with invariant moments," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 5, pp. 2351–2362, Sep. 1999.

[110] Z. Chen and S.-K. Sun, "A Zernike moment phase-based descriptor for local image representation and matching," *IEEE Trans. Image Process.*, vol. 19, no. 1, pp. 205–219, Jan. 2010.

[111] P. H. S. Torr and C. Davidson, "IMPSAC: Synthesis of importance sampling and random sample consensus," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 3, pp. 354–364, Mar. 2003.

[112] H. Li and G. Hua, "Probabilistic elastic part model: A pose-invariant representation for real-world face verification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 918–930, Apr. 2018.

[113] S.-K. Hwang, M. Billinghurst, and W.-Y. Kim, "Local descriptor by Zernike moments for real-time keypoint matching," in *Proc. IEEE Congr. Image Signal Process.*, vol. 2, 2008, pp. 781–785.

[114] J. Heikkilä, "Pattern matching with affine moment descriptors," *Pattern Recognit.*, vol. 37, no. 9, pp. 1825–1834, 2004.

[115] M. Cho and K. M. Lee, "Progressive graph matching: Making a move of graphs via probabilistic voting," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 398–405.

[116] X.-H. Han, Y.-W. Chen, and G. Xu, "High-order statistics of weber local descriptors for image representation," *IEEE Trans. Cybern.*, vol. 45, no. 6, pp. 1180–1193, Jun. 2015.

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

[118] 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.

[119] J. Ma, W. Qiu, J. Zhao, Y. Ma, A. L. Yuille, and Z. Tu, "Robust L2E estimation of transformation for non-rigid registration," *IEEE Trans. Signal Process.*, vol. 63, no. 5, pp. 1115–1129, Mar. 2015.

[120] J. Ma, J. Zhao, and A. L. Yuille, "Non-rigid point set registration by preserving global and local structures," *IEEE Trans. Image Process.*, vol. 25, no. 1, pp. 53–64, Jan. 2016.

[121] Y. Liu, L. De Dominicis, B. Wei, L. Chen, and R. R. Martin, "Regularization based iterative point match weighting for accurate rigid transformation estimation," *IEEE Trans. Vis. Comput. Graph.*, vol. 21, no. 9, pp. 1058–1071, Sep. 2015.

[122] H. Zhou, J. Ma, C. Yang, S. Sun, R. Liu, and J. Zhao, "Nonrigid feature matching for remote sensing images via probabilistic inference with global and local regularizations," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 3, pp. 374–378, Mar. 2016.

[123] M. Brown, G. Hua, and S. Winder, "Discriminative learning of local image descriptors," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 1, pp. 43–57, Jan. 2011.

[124] L. Shao, L. Liu, and X. Li, "Feature learning for image classification via multiobjective genetic programming," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 7, pp. 1359–1371, Jul. 2014.

[125] Z. Feng, J. Lai, and X. Xie, "Learning view-specific deep networks for person re-identification," *IEEE Trans. Image Process.*, vol. 27, no. 7, pp. 3472–3483, Jul. 2018.

[126] L. Shao, D. Wu, and X. Li, "Learning deep and wide: A spectral method for learning deep networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 12, pp. 2303–2308, Dec. 2014.

[127] Y. Yuan, L. Mou, and X. Lu, "Scene recognition by manifold regularized deep learning architecture," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 10, pp. 2222–2233, Oct. 2015.

[128] L. Liu, L. Shao, X. Li, and K. Lu, "Learning spatio-temporal representations for action recognition: A genetic programming approach," *IEEE Trans. Cybern.*, vol. 46, no. 1, pp. 158–170, Jan. 2016.

[129] G. Wu, M. Kim, Q. Wang, B. C. Munsell, and D. Shen, "Scalable high-performance image registration framework by unsupervised deep feature representations learning," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 7, pp. 1505–1516, Jul. 2016.

[130] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.

[131] P. Fischer, A. Dosovitskiy, and T. Brox. (2014). "Descriptor matching with convolutional neural networks: A comparison to SIFT." [Online]. Available: https://arxiv.org/abs/1405.5769

[132] Y. Gong, L. Wang, R. Guo, and S. Lazebnik, "Multi-scale orderless pooling of deep convolutional activation features," in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 392–407.

[133] S. Zagoruyko and N. Komodakis, "Learning to compare image patches via convolutional neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 4353–4361.

[134] E. Simo-Serra, E. Trulls, L. Ferraz, I. Kokkinos, P. Fua, and F. Moreno-Noguer, "Discriminative learning of deep convolutional feature point descriptors," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2016, pp. 118–126.

[135] Y. Tian, B. Fan, and F. Wu, "L2-Net: Deep learning of discriminative patch descriptor in Euclidean space," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 6128–6136.

[136] K. Nguyen, C. Fookes, A. Ross, and S. Sridharan, "Iris recognition with off-the-shelf CNN features: A deep learning perspective," *IEEE Access*, vol. 6, pp. 18848–18855, 2018.

[137] G. Cheng, P. Zhou, and J. Han, "Learning rotation-invariant convolutional neural networks for object detection in VHR optical remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 12, pp. 7405–7415, Dec. 2016.

[138] J. Y. Ma and J. Zhao, "Robust topological navigation via convolutional neural network feature and sharpness measure," *IEEE Access*, vol. 5, no. 99, pp. 20707–20715, 2017.

[139] W. Luo, J. Li, J. Yang, W. Xu, and J. Zhang, "Convolutional sparse autoencoders for image classification," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 7, pp. 3289–3294, Jul. 2018.

[140] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. Int. Conf. Learn. Represent.*, 2015.

[141] C. Szegedy *et al.*, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 1–9.

[142] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 770–778.

[143] L. Zheng, Y. Yang, and Q. Tian, "SIFT meets CNN: A decade survey of instance retrieval," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 5, pp. 1224–1244, May 2018.

[144] X. Zhang, H. Xiong, W. Zhou, W. Lin, and Q. Tian, "Picking neural activations for fine-grained recognition," *IEEE Trans. Multimedia*, vol. 19, no. 12, pp. 2736–2750, Dec. 2017.

[145] L. Lin, G. Wang, W. Zuo, X. Feng, and L. Zhang, "Cross-domain visual matching via generalized similarity measure and feature learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1089–1102, Jun. 2017.

[146] Y. Wen, B. Sheng, P. Li, W. Lin, and D. D. Feng, "Deep color guided coarse-to-fine convolutional network cascade for depth image super-resolution," *IEEE Trans. Image Process.*, vol. 28, no. 2, pp. 994–1006, Feb. 2019.
[147] J. Sivic and A. Zisserman, "Video Google: A text retrieval approach to object matching in videos," in *Proc. 9th IEEE Int. Conf. Comput. Vis.*, vol. 2, Oct. 2003, pp. 1470–1477.
[148] F. Perronnin, J. Sánchez, and T. Mensink, "Improving the fisher kernel for large-scale image classification," in *Proc. Eur. Conf. Comput. Vis.*, 2010, pp. 143–156.
[149] H. Jégou, M. Douze, C. Schmid, and P. Pérez, "Aggregating local descriptors into a compact image representation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 3304–3311.
[150] Y. Yuan and X. Hu, "Bag-of-words and object-based classification for cloud extraction from satellite imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 8, pp. 4197–4205, Aug. 2015.
[151] X. Li, L. Zhang, L. Wang, and X. Wan, "Effects of BOW model with affinity propagation and spatial pyramid matching on polarimetric SAR image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 7, pp. 3314–3322, Jul. 2017.
[152] J. Redolfi, J. Sánchez, and A. G. Flesia, "Fisher vectors for PolSAR image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 11, pp. 2057–2061, Nov. 2017.
[153] L. Liu *et al.*, "Compositional model based Fisher vector coding for image classification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2335–2348, Dec. 2017.
[154] Q. Li, Q. Peng, and C. Yan, "Multiple VLAD encoding of CNNs for image classification," *Comput. Sci. Eng.*, vol. 20, no. 2, pp. 52–63, Mar./Apr. 2018.
[155] L. Wang, D. Meng, X. Hu, J. Lu, and J. Zhao, "Instance annotation via optimal BoW for weakly supervised object localization," *IEEE Trans. Cybern.*, vol. 47, no. 5, pp. 1313–1324, May 2017.
[156] H. Wang, J. Hu, and W. Deng, "Compressing Fisher vector for robust face recognition," *IEEE Access*, vol. 5, pp. 23157–23165, 2017.
[157] B. Sheng, C. Shen, G. Lin, J. Li, W. Yang, and C. Sun, "Crowd counting via weighted VLAD on a dense attribute feature map," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 28, no. 8, pp. 1788–1797, Aug. 2018.
[158] Y. Xu, Y. Han, R. Hong, and Q. Tian, "Sequential video VLAD: Training the aggregation locally and temporally," *IEEE Trans. Image Process.*, vol. 27, no. 10, pp. 4933–4944, Oct. 2018.

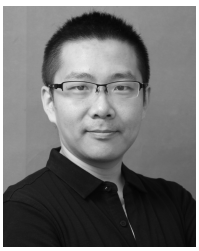
CHENGCAI LENG received the Ph.D. degree in applied mathematics from Northwestern Polytechnical University, Xi'an, China, in 2012. From 2010 to 2011, and, from 2017 to 2018, he was a Visiting Student and a Visiting Scholar with the Department of Computing Science, University of Alberta, Edmonton, AB, Canada, respectively. He is currently an Associate Professor with the School of Mathematics, Northwest University, Xi'an. His current research interests include image processing, computer vision, and optical molecular imaging.

HAI ZHANG received the Ph.D. degree in applied mathematics from Xi'an Jiaotong University, Xi'an, in 2012. He is currently a Professor with the Department of Financial Mathematics and Statistics, Northwest University. His research interests include statistical machine learning, high-dimensional statistics, and social network analysis.

BO LI received the Ph.D. degree in computational mathematics from the Dalian University of Technology, Dalian, China, in 2008. He is currently an Associate Professor with the School of Mathematics and Informatics Science, Nanchang Hangkong University. His current research interests include the areas of image processing and computer graphics.

GUORONG CAI received the Ph.D. degree in computer science from Xiamen University, Fujian, China, in 2013. He is currently an Associate Professor with the Computer Engineering College, Jimei University, Xiamen, China. His research interests include 3D reconstruction, machine learning, object detection/recognition, and image/video retrieval.

ZHAO PEI received the joint Ph.D. degree from the Department of Computing Science, University of Alberta, Edmonton, AB, Canada, in 2011, and the Ph.D. degree from Northwestern Polytechnic University, Xi'an, China, in 2013. He is currently an Associate Professor with the School of Computer Science, Shaanxi Normal University, Xi'an, China. His research interests include camera array synthetic aperture imaging, object detection and tracking, computer vision, and pattern recognition.



LI HE received the B.Sc., M.Sc., and Ph.D. degrees from the Department of Automation, Northwestern Polytechnical University, Xi'an, China, in 2006, 2009, and 2014, respectively. He was a Visiting Ph.D. Student with the Department of Computing Science, University of Alberta, from 2010 to 2011. He served as a Post-Doctoral Fellow with the Department of Computing Science, University of Alberta, from 2014 to 2017. He is currently an Associate Professor with the School of Electromechanical Engineering, Guangdong University of Technology, Guangzhou, China. He has more than 20 publications on high-rank venues, such as TCYB, TIP, PR, and IROS. His current research interests include machine learning, visual SLAM, and computer vision. He is an Associate Editor of the IEEE Access and a Managing Guest Editor of *Computers & Electrical Engineering*.