Hand-Dorsa Vein Recognition by Matching Local Features of Multisource Keypoints

Di Huang, Yinhang Tang, Yiding Wang, Liming Chen, Yunhong Wang

Hand-Dorsa Vein Recognition by Matching Local Features of Multisource Keypoints

Di Huang, *Member, IEEE*, Yinhang Tang, Yiding Wang, Liming Chen, *Senior Member, IEEE*, and Yunhong Wang, *Member, IEEE*

***Abstract*—As an emerging biometric for people identification, the dorsal hand vein has received increasing attention in recent years due to the properties of being universal, unique, permanent, and contactless, and especially its simplicity of liveness detection and difficulty of forging. However, the dorsal hand vein is usually captured by near-infrared (NIR) sensors and the resulting image is of low contrast and shows a very sparse subcutaneous vascular network. Therefore, it does not offer sufficient distinctiveness in recognition particularly in the presence of large population. This paper proposes a novel approach to hand-dorsa vein recognition through matching local features of multiple sources. In contrast to current studies only concentrating on the hand vein network, we also make use of person dependent optical characteristics of the skin and subcutaneous tissue revealed by NIR hand-dorsa images and encode geometrical attributes of their landscapes, e.g., ridges, valleys, etc., through different quantities, such as cornerness and blobness, closely related to differential geometry. Specifically, the proposed method adopts an effective keypoint detection strategy to localize features on dorsal hand images, where the speciality of absorption and scattering of the entire dorsal hand is modeled as a combination of multiple (first-, second-, and third-) order gradients. These features comprehensively describe the discriminative clues of each dorsal hand. This method further robustly associates the corresponding keypoints between gallery and probe samples, and finally predicts the identity. Evaluated by extensive experiments, the proposed method achieves the best performance so far known on the North China University of Technology (NCUT) Part A dataset, showing its effectiveness. Additional results on NCUT Part B illustrate its generalization ability and robustness to low quality data.**

***Index Terms*—Hand-dorsa vein recognition, multilevel keypoint detection, optical properties of dorsa hand subcutaneous tissue, oriented gradient maps (OGMs).**

This work was supported in part by the National Basic Research Program of China under Grant 2010CB327902, in part by the National Natural Science Foundation of China under Grant 61202237 and Grant 61271368, in part by the Beijing Municipal Natural Science Foundation under Grant 4142032 and Grant KZ201410009012, in part by the French Research Agency (Agence Nationale de Recherche, ANR) under Grant ANR-2010-INTB-0301-01 and Grant ANR-13-INSE-0004-02, in part by the Specialized Research Fund for the Doctoral Program of Higher Education under Grant 20121102120016, in part by the Research Program of State Key Laboratory of Software Development Environment under Grant SKLSDE-2013ZX-31, in part by the Joint Project by the LIA2MCSI Laboratory between the Group of Ecoles Centrales and Beihang University, and in part by the Fundamental Research Funds for the Central Universities. This paper was recommended by Associate Editor P. Bhattacharya.

D. Huang and Y. Wang are with the State Key Laboratory of Software Development Environment, School of Computer Science and Engineering, Beihang University, Beijing 100191, China (e-mail: dhuang@buaa.edu.cn; yhwang@buaa.edu.cn).

Y. Tang and L. Chen are with the Department of Mathematics and Computer Science, Ecole Centrale de Lyon, Lyon 69134, France (e-mail: germain.tang-yinhang@doctorant.ec-lyon.fr; liming.chen@ec-lyon.fr).

Y. Wang is with the College of Information Engineering, North China University of Technology, Beijing 100041, China (e-mail: wangyd@ncut.edu.cn).

I. Introduction

DRIVEN mainly by increasing requirements in public security against terrorist activities, sophisticated crimes, and electronic frauds, biometric solutions have witnessed an accelerated pace of growth in the global market of security over the past several decades. Recently, the vein has emerged as a new biometric trait for the purpose of people identification, and has received growing attention within the community.

Anatomically, veins are blood carrying vessels interweaved with muscles and bones, and the key function of the vascular system is to supply oxygen to each part of the body. The spatial arrangement of vascular network in the human body is stable and unique, and vein patterns of individuals are different, even between identical twins [1]. In this paper, we focus on the vein pattern of the back of the hand (i.e., dorsal hand) because it is distinctly visible, easy to acquire, and efficient to process. As compared with other popular biometric traits, such as face or fingerprint, the hand vein has several distinguished merits, in particular the following ones.

1) *Direct Liveness Detection:* Hand veins are sensed using far or near-infrared (NIR) lighting to capture the temperature difference between hot blood flow inside vein vessels and the surrounding skin, therefore, they can only be imaged on the live body and the images taken on nonlive bodies do not contain their spatial vein arrangement.
2) *Safety:* Blood vessel patterns are hardwired underneath the skin at birth; they are hence much harder for intruders to forge.

The pattern of vein as a biometric trait is relatively recent. It was not presented until 1990 when MacGregor and Welford [2] came up with the system named "vein check" for identification. Despite the vast vascular network in the human body, hand veins are favored for their simplicity in terms of acquisition and processing. In last decades, there exist increasing amount of research works focusing on hand vein recognition using the vein pattern in the palm part [3]–[5], the back of the hand [6]–[8], or fingers [9].

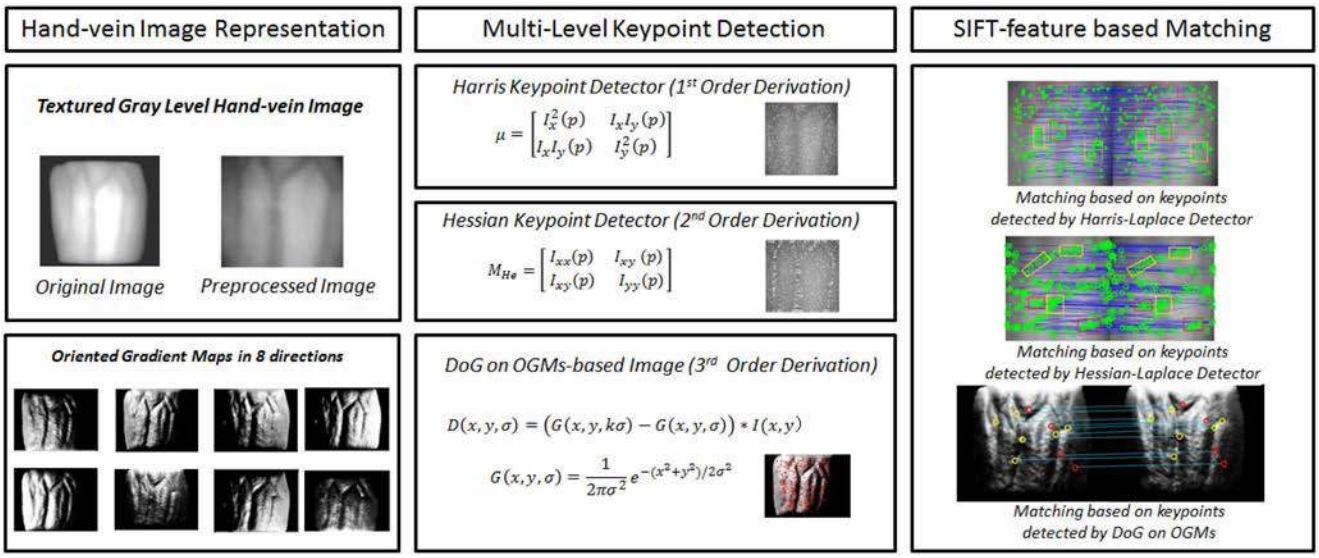

Fig. 1. Framework of the proposed approach, including comprehensive representation of optical properties through multiorder gradient quantities and robust matching with scale-invariant feature transform (SIFT) based features.

Although there have been already several attempts on hand vein recognition by adopting holistic techniques, e.g., principal component analysis (PCA) [10], linear discriminant analysis (LDA) [11], etc., the changes of viewpoint, lighting intensity, distortion, and occlusion largely imped their development. In contrast, local feature based approaches become dominant due to its robustness to the aforementioned disturbing factors. Most of the methods in the literature follow the framework that first segments the region of interest and the hand subcutaneous vascular network from the hand vein image, and then extracts local geometric features for matching such as the positions and angles of short straight vectors [12], vein minutiae and knuckle shapes [1], endpoints and crossing points [13], dominant points [3], etc. All these methods demonstrate reasonable recognition rates on small databases ranging from 32 [3] to 100 subjects [1], [7]. However, when regarding the problem of dorsal hand vein recognition, the above techniques suffer from very limited local features because compared with the palm and finger part, the number of vein minutiae on the dorsal hand is really few, directly leading to the deficiency in capturing the difference of hand vein networks between subjects. Hand dorsa vein images are mostly sensed by the NIR imaging system, irradiating the hand dorsa with the NIR light. In delivering the vein pattern of the hand dorsa, these images hence also convey the optical properties, i.e., the absorption and scattering speciality, of the skin and subcutaneous tissue which mainly consists of three different layers, namely epidermis, dermis, and hypodermis. The randomly inhomogeneous distribution of blood and various chromophores and pigments produces variations of optical properties of these skin layers that are subject dependent [14]. These optical properties are investigated as such in medicine for various purposes, e.g., diagnostics, surgery, therapy. In this paper, we propose to make full use of these optical properties of the hand dorsa for people identification.

Specifically, in this paper, we propose a novel and effective approach to hand-dorsa vein recognition based on local feature matching. Unlike the overwhelming majority of state of the art techniques which only focus on the venous network, the proposed method makes full use of discriminative clues as offered by the optical properties of NIR dorsal hand images that cover not only the vein areas but also their surrounding skin and subcutaneous regions. In the same way as the retinal image [15], the optical properties conveyed by NIR dorsal hand images are interpreted as landscapes or surfaces, consisting of geometric features like ridges, valleys, summits, etc. Their properties are comprehensively analyzed using differential geometry quantities, resulting in a set of keypoints of multiple-order gradient cues (from the first to third order). More precisely, we introduce the Harris–Laplace detector to characterize the elasticity, i.e., length and angle variations, of the underlying surface, through the cornerness measurement of the first order gradients [16]. We then describe the hand-dorsa vein areas which coincide with the valley regions of the underlying landscape because of their absorption and scattering properties. They are identified using the Hessian–Laplace detector [17] which relies on the blobness measurement of the Hessian matrix of the second order gradients. In order to further thoroughly highlight shape changes, i.e., the changes in optical properties, of the whole hand dorsa skin and subcutaneous tissue, we also compute a human vision inspired representation, namely oriented gradient maps (OGMs) [18], of the original image and then identify feature points through the difference of Gaussian (DoG) [19]. Because OGMs are first order gradient based and DoG (an approximation of Laplace of Gaussian, LoG) is second order gradient based, these features are essentially third order gradient based and correspond to the points whose curvatures change most on the surfaces. Finally, the keypoints as detected by the previous process between the hand-dorsa images of the same subject are robustly associated using local feature matching for decision making, accounting for moderate geometrical transformations and possible lighting variations that often occur in image acquisition. See Fig. 1 for the approach

framework. The proposed approach is extensively evaluated on North China University of Technology (NCUT) Part A and NCUT Part B, both of which are among the largest dorsal hand vein datasets so far known in the literature. Experimental results clearly demonstrate the effectiveness of the proposed method.

The contributions of this paper can be summarized as follows.

1) We prove that dorsal hand vein based people identification can not only rely on the vascular network, but also depend on the optical characteristics, i.e., the absorption and scattering properties, of surrounding skin as well as subcutaneous tissue, since the randomly inhomogeneous distribution of blood and various chromophores and pigments is subject dependent.
2) We interpret NIR dorsal hand images as landscapes and surfaces and identify these keypoints of their optical properties using geometric features through the quantities of multiorder (i.e., the first-, second-, and third-order) gradient cues, namely Harris cornerness measurement, Hessian blobness measurement, and curvature extrema by operating the DoG detector on a human vision inspired image representation, OGMs, which are closely related to the quantities in differential geometry.
3) We demonstrate that these keypoints as localized by the aforementioned multiorder gradient based quantities capture different geometric attributes corresponding to complementary facets of the optical properties of the vein network as well as its surrounding skin and subcutaneous tissue. As a result, we further propose to combine these local features for identification and achieve the best recognition accuracy so far known on the NCUT Part A dataset.

Preliminary results appear in [20] and [21]. This paper includes previous results but significantly extends them in the following ways. Firstly, according to recent studies of optical techniques for medicine, we state the motivation and the rational of using both the hand vein area and the optical properties of the surrounding skin and subcutaneous tissue in people identification. Secondly, in the same way as retinal images, we interpret hand dorsa vein images as landscapes or surfaces and explain why the different facets of the optical properties of the skin and subcutaneous tissue can be captured by geometric features through quantities related with differential geometry. Thirdly, because the keypoints are localized by multiorder gradient quantities, i.e., Harris cornerness and Hessian blobness measurements and OGMs with DoG, to represent different optical characteristics of dorsal hand images, we further propose an effective fusion approach that integrates and associates these local features for matching. Fourthly, we comprehensively evaluate the approach considering not only the identification scenario as did in [20] and [21], but also the verification scenario, thereby illustrating the general nature of the proposed method for the most common applicative conditions. Finally, we also assess and discuss the time complexity of the system.

The remainder of this paper is organized as follows. Section II introduces the acquisition process of NIR dorsal hand vein images. Section III presents the multilevel keypoint detection method and Section IV describes the OGMs based dorsal hand representation. The local matching step is shown in Section V. The experimental results of both scenarios in dorsal hand vein recognition and verification are displayed and analyzed in Section VI. Section VII concludes this paper.

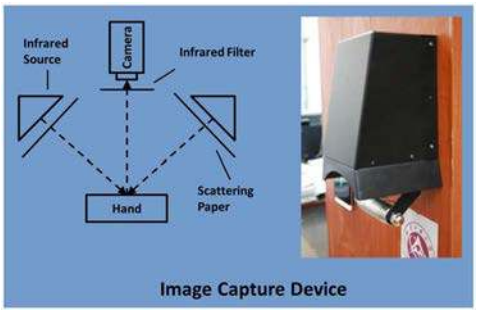

Fig. 2. Illustration of the NIR imaging system.

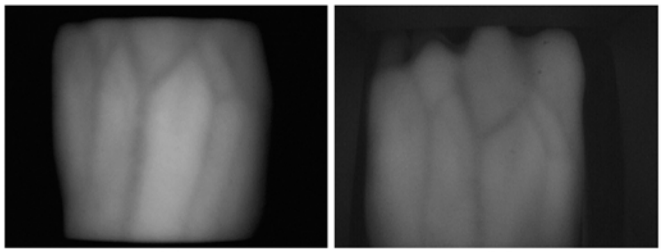

Fig. 3. Hand-dorsa vein images. From the NCUT (a) Part A dataset and (b) Part B dataset.

II. Vein Image Acquisition

Fig. 2 illustrates the system setup where an LED array lamp is exploited to shine infrared light onto the back of the hand. The incident infrared light can penetrate into the biological tissue with an approximate depth of 3 mm, and the randomly inhomogeneous distribution of blood and various chromophores and pigments produces optical characteristics, i.e., absorption and scattering properties that are subject dependent [14]. Since the flow of hot blood inside the vein network generally absorbs and scatters more infrared radiation than the surrounding skin and subcutaneous tissue, its curvilinear structures are imaged through a CCD camera associated with an IR filter where the veins appear darker valleys whereas the surrounding skin and subcutaneous tissue displays a landscape or surface, containing various features, e.g., cliffs, ridges, plateaux, basins, etc., (see Fig. 3). The spectral responses or the variations of these optical attributes of the hand-dorsa skin and the subcutaneous tissue, including in particular the vascular network, are thus perfectly modeled by using differential geometric quantities.

Using such a hardware setup depicted in Fig. 2, a database of 2040 dorsal hand vein images of both hands of 102 subjects was built by North China University of Technology in 2010, and it was marked as the NCUT Part A database. In order to make the device more practical, another sensor was proposed by NCUT, in which a trade-off was considered between the expenditure of hardware and the quality of hand vein images.

The CCD camera and IR optical filter were substituted, leading to around half reduction in total cost. Another dataset, namely NCUT Part B, was collected through this novel device in 2011. It consists of 2020 dorsal hand vein images of 101 subjects, each of which owns 20 images; half for the left hand and half for the right hand. In contrast to the NCUT Part A database, NCUT Part B dataset is composed of dorsal hand vein images under different acquisition conditions, and the images are more noisy. Since the vein patterns are best described when the skin on the dorsal hand is taut, a handle was mounted at the bottom of the device to position the hand, and the images were thereby roughly aligned. Fig. 3 shows samples of NCUT Part A and Part B captured with a resolution of 640 by 480 pixels. There were no major illumination variations, but moderate changes in viewpoint (i.e., differing by rotations as well as translations) still can occur since these images were collected in different periods and environmental situations.

As we can see from Fig. 3, the pattern of the dorsal hand vein is captured and it appears darker within the NIR image. The widths of these vein profiles change in the range of 30 to 50 pixels. Even though the vein spatial arrangement is visible, it is not very distinguishable from the surrounding bio-tissue. Furthermore, the number of local features, e.g., endpoints and crossing points, is quite limited and usually varies from 5 to 10, thereby making local feature-based approach questionable for the discriminative power as directly applied to these dorsal hand images. On the other hand, the spectral response of the surrounding skin as well as subcutaneous tissue translates the subject-dependent inhomogeneous compositions of blood and various chromophores and pigments, and their variations are also imaged by the NIR sensor. In interpreting the hand vein image as a landscape or a surface in the same way as retinal images [15], the key features of these variations in absorption and scattering characteristics can be perfectly captured through differential geometric properties, e.g., cliffs, ridges, plateaux, basins, including in particular the valley which corresponds to the vascular network. In order to localize these geometric features and hence increase the number of local features for more distinctiveness, we propose to make use of quantities closely related to differential geometry, namely Harris cornerness and Hessian blobness measurements grouped under the multilevel keypoint detection on the dorsal hand image, and DoG based curvature extrema on a human vision inspired representation, i.e., OGMs. They describe complementary geometric attributes, and we introduce them in the subsequent two sections, respectively.

III. Multilevel Keypoint Detection

For local feature-based matching approaches, keypoint detection is a critical step which is expected to locate a sufficient number of local feature points for a comprehensive description of the target image while providing some properties of invariance, e.g., scale, translation, rotation, etc. There exist several state of the art keypoint detection methods, such as DoG, Harris, Hessian, Harris–Laplace, Hessian–Laplace, whose properties on textured gray level images have been explicitly investigated by Roth and Winter [22] in object retrieval (see Table I). However, due to the optical properties, the dorsal hand vein image contains very few texture details. In this section, we are interested in the geometric attributes of these keypoint detectors when the underlying hand vein images are interpreted as surfaces. This geometric analysis results in the design of our multilevel keypoint detection for hand vein images.

TABLE I
Comparison of Different Local Feature Detectors

Detector	*Invarance*	*Repeat*	*KeyPt.Num.*
Harris	Rotation	High	High
Hessian	Rotation	High	High
Harris-Laplace	Rotation & Scale	High	High
Hessian-Laplace	Rotation & Scale	High	High
DoG	Rotation & Scale	High	Medium

A. DoG Detector

DoG, proposed by Lowe [19], is one of the most widely used detectors, and it serves the scale-invariant feature transform (SIFT) feature extraction and matching.

The image is first repeatedly convolved with Gaussian filters of different scales separated by a constant factor, k, to generate an octave in the scale space. As for an input image, $I(x, y)$, its scale space is defined as a function, $L(x, y, \alpha)$, produced by a convolution of a variable scale Gaussian $G(x, y, \alpha)$ with the input image I, and the DoG function $D(x, y, \alpha)$ can be computed from the difference of two nearby scales

$$\begin{aligned} D(x, y, \alpha) &= (G(x, y, k\alpha) - G(x, y, \alpha)) \times I(x, y) \\ &= L(x, y, k\alpha) - L(x, y, \alpha). \end{aligned} \tag{1}$$

The extrema of $D(x, y, \alpha)$ can be detected by comparing each pixel value with those of its 26 neighbors within a 3×3 area at the current and adjacent scales. At each scale, gradient magnitude and orientation, $m(x, y)$ and $\theta(x, y)$ [as shown in (2) and (3)], are computed by exploiting pixel differences. The confirmed stable extremes are regarded as the scale-invariant keypoints located by DoG

$$\begin{aligned} m^2(x, y) = &(L(x+1, y) - L(x-1, y))^2 \\ &+ (L(x, y+1) - L(x, y-1))^2 \end{aligned} \tag{2}$$

$$\theta(x, y) = \tan^{-1} \frac{L(x+1, y) - L(x, y-1)}{L(x+1, y) - L(x-1, y)}. \tag{3}$$

DoG has proved competent at blob detection on gray level images. In hand vein analysis, the optical properties of the vein and its nearby tissue result in images with very limited texture details. Therefore, DoG locates very few feature points which are not located on hand vein regions [see Fig. 4(a)]. However, when the dorsal hand image is considered as a surface, we can give a geometric interpretation of these local features detected by using DoG. In our implementation, DoG can be regarded as an approximation of the Laplacian of Gaussian (LoG) with the ratio of the scales equal to 1.6. In this case, the Laplacian calculates the addition of these second partial derivatives and delivers the sum of both the curvatures in the x and y direction. Keeping this property in mind, we can see from Fig. 4(a) that DoG has actually located a few points on the surface displayed

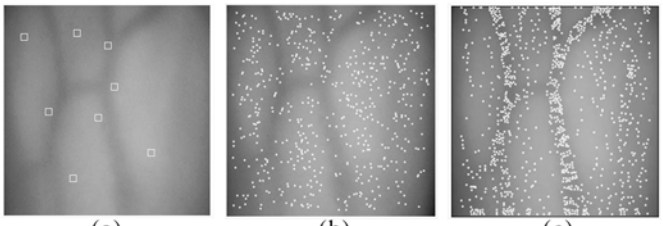

Fig. 4. Distribution of keypoints detected by (a) DoG, (b) Harris–Laplace, and (c) Hessian–Laplace (1000 clusters), on a hand-dorsa surface. DoG locates very few feature points whose sums of x and y curvatures are extrema; Harris–Laplace identifies the keypoints whose elasticities are greater than a threshold; Hessian–Laplace detects the keypoints which carry shape information in terms of curvatures, and localizes in particular the ones which densely populate the valley regions corresponding to veins.

through the dorsal hand image whose sums of the curvatures in the x and y directions are extrema either on ridges or on basins. Of course, the number of these feature points is not sufficient to comprehensively capture the whole geometric attributes of the underlying surface. We therefore study other state of the art local feature detection techniques and analyze their geometric properties.

B. Harris and Hessian Keypoint Detection

From Table I, we can see that compared with DoG, the other detectors, i.e., Harris, Hessian, Harris–Laplace, and Hessian–Laplace, not only have high repeatability, but also locate more keypoints on the gray level image. Both the Harris–Laplace and Hessian–Laplace are similar with DoG in the performance of scale invariance. Instead of Laplacian approximation, Harris- and Hessian–Laplace apply the scale normalized Laplacian to create the scale space which gives the benefit to local feature extraction and matching.

Specifically, the Harris detector was proposed by Harris and Stephens who defined the product of two first derivation matrices [22]

$$\mu = \begin{bmatrix} I_x^2(p) & I_xI_y(p) \\ I_xI_y(p) & I_y^2(p) \end{bmatrix} = \begin{bmatrix} A & B \\ B & C \end{bmatrix} \tag{4}$$

and it responds to corner features on gray level images. I_x and I_y denote the first derivation of the image I at position p in the x and y direction, respectively. The corner response threshold c calculated by avoiding the eigenvalue decomposition of the second moment matrix above by

$$c = Det(\mu) - k \times Tr(\mu)^2 = (AC - B^2) - k \times (A + C)^2. \tag{5}$$

The Hessian matrix-based detector is similar with the Harris detector but presents strong responses on blob features, instead of the corner ones, because the Hessian matrix-based detector replaces the elements of (4) with the second derivation

$$M_{He} = \begin{bmatrix} I_{xx}(p) & I_{xy}(p) \\ I_{xy}(p) & I_{yy}(p) \end{bmatrix} \tag{6}$$

where I_{xx} and I_{yy} are the second derivatives of the image I at the position p in the x and y direction, respectively; and I_{xy} is the mixed derivative in both directions.

The two detectors above, i.e., Harris and Hessian, only own the property of rotation invariance. To achieve scale invariance, the scale normalized Laplacian S defined in (7) is introduced as a scale selection criterion by Harris–Laplace and Hessian–Laplace, and both detectors thus possess the property of scale invariant as DoG does

$$S = s^2 \times \left| I_{xx}(p) + I_{yy}(p) \right|. \tag{7}$$

While the Harris and Hessian keypoint detectors are mostly analyzed in terms of properties of cornerness and blobness on gray level images, they can also be described as geometrical attributes on surfaces because of their close relationships with the first and second fundamental forms in differential geometry [23]. Indeed, the matrix of the Harris detector as defined in (4) is related to the symmetric matrix of the fundamental form which characterizes the metric properties of a surface, i.e., how the length and area are changed on the surface with regard to the ambient space. In other words, the matrix of Harris in (4) and the cornerness response in (5) characterize somehow the elasticity of a surface as we can see in Fig. 4(b). In this figure, the Harris–Laplacian detector locates much more keypoints in comparison with DoG. Furthermore, these keypoints cover the whole hand vein image and identify those points on the hand-dorsa surface whose elasticity is greater than a given threshold. They can thereby contribute to people identification using the optical properties of the hand dorsa subcutaneous tissue.

In regard to the Hessian–Laplace detector, the matrix in (6) with the second derivatives is related to the matrix of the second fundamental form which characterizes how an embedded surface is curved in the ambient space using curvature metrics, e.g., principal curvatures, mean, and Gaussian curvatures. The Hessian–Laplace detector hence delivers keypoints on the hand dorsa surface with shape clues in terms of curvatures. As we can see in Fig. 4(c), the keypoints densely populate the valley regions, i.e., the hand vein regions.

C. Design of Multilevel Keypoint Detection

Given the fact that both the vein and the optical attributes of the surrounding subcutaneous tissue are subject dependent, an effective way to characterize a person is to adopt the keypoints localized by the Harris–Laplace and Hessian–Laplace detectors. The former captures the elasticity of the underlying surface of the dorsal hand whereas the latter delivers the points of shape information, in particular those populating the valley regions of hand veins. The resulting method for locating these keypoints is called in the subsequent multilevel keypoint detector.

Some statistical analysis has been conducted along with the experiments in this paper using the images on the NCUT Part A database. DoG only detects less than ten keypoints on each hand-dorsa image, and such a sparsity in local features cannot provide sufficient distinctiveness and thus fails to result in a reasonable recognition accuracy. When exploiting the Harris–Laplace detector, we can averagely locate 640 keypoints, and this number is indeed much larger than that of DoG.

Regarding Hessian–Laplace, around 3000 local features can be found on each dorsal hand image, and this amount causes

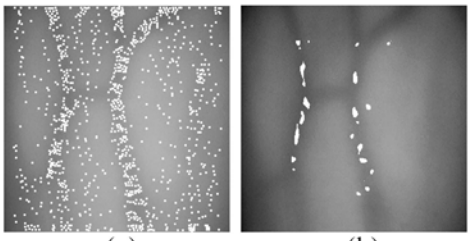

Fig. 5. 1000 selected keypoints located by the Hessian–Laplace detector, based on (a) clustering and (b) strongest responses.

a sharp increase of computational cost in matching. We hence consider selecting a subset of the most representative features. Generally, it is straightforward to choose the strongest points in terms of their responses in detection. However, those points, achieved in this operation whose responses are higher than the others, only distribute on the partial dorsal hand vein network [as in Fig. 5(b)], leading to a loss in discriminative power. As a result, to reduce the number of Hessian–Laplace based points while keeping the distinctiveness, the number of keypoints is reduced through clustering [as shown in Fig. 5(a)], where only their locations (i.e., x and y coordinates) are considered. In our case, the k-means algorithm is employed to randomly cluster the points into 500, 700, and 1000, respectively, to balance the performance and efficiency.

IV. OGMs Based Representation

In the previous section, we analyze the geometric properties of several state of the art keypoint detectors when dorsal hand vein images are interpreted as surfaces, and propose a multilevel keypoint detector to localize features not only on the vein but also on its surrounding subcutaneous tissue. In this section, we further increase the descriptive completeness of these local features through an approach inspired by human vision, using OGMs which are originally applied to represent the texture as well as shape information for 3-D face recognition [24].

The objective of the OGMs is to provide a visual description simulating the operation of human complex cells in the visual cortex [25]. These complex neurons respond to a gradient at a particular orientation and spatial frequency, but the location of the gradient is allowed to shift over a small receptive field rather than being precisely localized.

A. Representation of Complex Neuron Response

The proposed OGM based representation simulates the response of complex neurons through a convolution of gradients in specific directions within a predefined neighborhood. Since the scale of the dorsal hand vein image changes slightly thanks to the hardware setup, we only employ a circular neighborhood R, as demonstrated in Fig. 6. The precise radius value of the circular area needs to be fixed experimentally. The response of a complex neuron at a given pixel location is a set of gradient maps in different orientations convolved by a Gaussian kernel.

Specifically, given an input image (a dorsal hand vein image in our case) I, a certain number of gradient maps

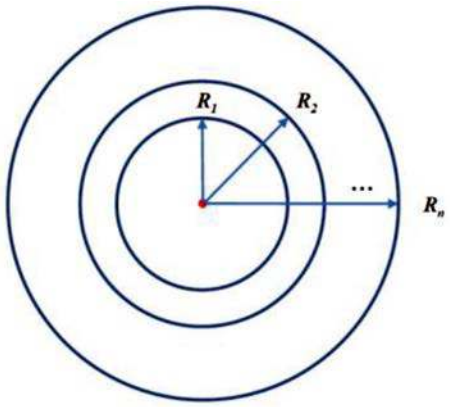

Fig. 6. Neighborhood of the complex neutrons is a circular area and its radius can be changed according to the scale.

G_1, G_2,...,G_o, one for each quantized direction o, are firstly computed. They are defined as

$$G_o = \left(\frac{\partial I}{\partial o}\right)^+. \tag{8}$$

The "+" sign indicates that only the positive values are kept to preserve the polarity of the intensity changes, and the negative ones are set to zero.

Each of gradient maps describes gradient norms of the input image in an orientation o at every pixel. We further simulate the response of the complex neurons by convolving its gradient maps with a Gaussian kernel G, and its standard deviation is proportional to the radius value of the given neighborhood, R, as in

$$\rho_o^R = G_R * G_o. \tag{9}$$

The purpose of the convolution with Gaussian kernels is to allow the gradients to shift in a neighborhood without abrupt changes.

At a given pixel location (x, y), we collect all values of the convolved gradient maps at that location and form the vector $\rho^R(x, y)$, and it hence possesses a response value of complex neurons for each orientation o

$$\rho^R(x, y) = \left[\rho_1^R(x, y),\ \ldots,\ \rho_O^R(x, y)\right]^t. \tag{10}$$

$\rho^R(x, y)$, is then normalized to an unit norm vector, which is called response vector and denoted by $\underline{\rho}^R$.

B. OGMs by Response Vectors

According to the definition of the response vector, the dorsal hand vein image can be represented by its perceived values of complex neurons. Specifically, given a hand vein image I, we generate an OGM J_o using complex neurons for each orientation o defined as in

$$J_o(x, y) = \underline{\rho}_o^R(x, y). \tag{11}$$

Fig. 7 depicts such a process applied to a dorsal hand vein image. In this paper, we generate eight OGMs for eight predefined quantized directions. Instead of the original NIR hand dorsa images, these OGMs are thus exploited in the subsequent local feature extraction and matching for identification.

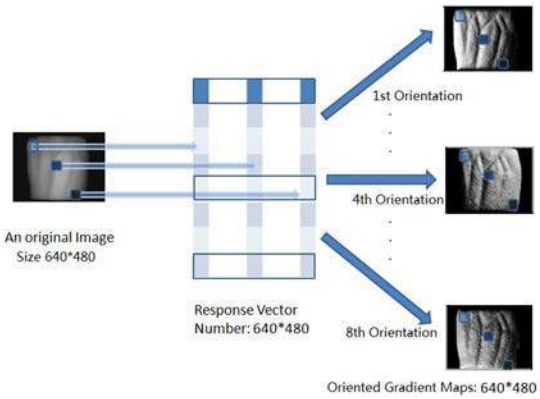

Fig. 7. OGMs describe a perceived near-infrared hand-dorsa vein image in eight orientations.

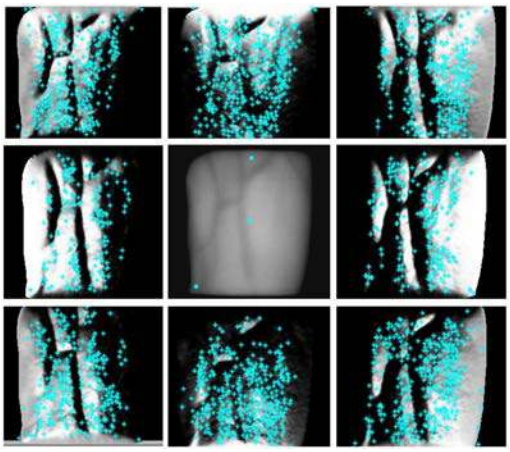

Fig. 8. Comparison in keypoint detection by DoG in the raw hand-dorsa vein image (center) and its corresponding OGMs in the eight predefined quantized orientations (around).

C. Properties of Distinctiveness and Invariance

The OGMs potentially offer high distinctiveness since they highlight the details of local texture changes. Meanwhile, they also possess some interesting properties of robustness to affine lighting variations.

When applied OGMs to dorsal hand vein images, they offer the property of being robust to affine illumination transformations. Indeed, each OGM, J_o, is simply normalized convolved gradient maps at the orientation o, while monotonic illumination change often adds a constant intensity value, as a result, it does not affect the computation of gradients. Furthermore, a change in image contrast in which the intensities of all pixels are multiplied by a constant will lead to the multiplication of gradient computation; however, such a contrast change will be canceled by the normalization of response vectors.

OGMs can be made even rotation invariant if we choose to quantize directions starting from that of the principal gradient of all the gradients within the neighborhood, and the tolerance to scale variations can also be largely improved by embedding the multiscale strategy. Nevertheless, we do not perform such steps to save computational cost as the dorsal hand vein images in our study were already roughly aligned.

D. Design of the OGM Based Keypoint Detector

After the OGMs of a dorsal hand vein image are computed to highlight the details of optical properties of the underlying vein network and its nearby subcutaneous tissue, they are then interpreted as retinal images [15], i.e., surfaces or landscapes, and their geometric attributes can be further analyzed. In this paper, we concentrate on the variations of shape of these OGM-based surfaces and employ DoG which identifies the keypoints whose sums of curvatures in the x and y directions change the most.

Fig. 8 demonstrates the distribution of the keypoints detected by DoG, from the hand-dorsa vein image and its corresponding OGMs, respectively. Because OGMs simulate the operation of complex cells of the visual cortex and therefore highlight the details of the vein patterns and their surrounding subcutaneous tissue, DoG locates much more keypoints, including in particular the dorsal hand vein minutia, on these OGM-based surfaces in comparison with the smooth raw hand-dorsa vein surface, which comprehensively describe these optical characteristics. The statistics that we computed show that the average number of keypoints extracted from each of OGM can rise up to 627, while that from the original dorsal hand vein image is less than ten as stated in Section III. Fig. 8 illustrates this phenomenon.

Recall that the Harris–Laplace detector locates the keypoints whose elasticities are greater than a threshold from the hand-dorsa surface and the Hessian–Laplace detector mostly localizes the keypoints in the valley regions, i.e., the vein areas. As compared to these features located by the multilevel detector, i.e., Harris–Laplace as well as Hessian–Laplace, DoG identifies the keypoints whose shape changes the most at a given OGM-based surface. In the viewpoint of differential calculus, Harris–Laplace provides the first order gradient information; Hessian–Laplace offers the second order gradient information; whereas DoG associated with OGMs generates the third order gradient information from an input hand-dorsa surface. These keypoints are thus complementary for a comprehensive description of the hand-dorsa vein image and can be used for people identification through a local feature matching process.

V. Local Feature Matching

Once identified the keypoints using the multilevel keypoint detection and DoG with OGM as described in the previous two sections, we further extract the widely-used SIFT features [19] at these positions to enable the matching between two dorsal hand vein images for similarity measure computation and final decision making.

A. SIFT-Feature Based Matching

For each detected keypoint, a feature vector is extracted as a descriptor from these gradients of sampling points within its neighborhood. In order to obtain the orientation invariance, the coordinates and gradient orientations of sampling points in the neighborhood are rotated relative with the keypoint orientation. Then a Gaussian function is employed to assign a weight to the

gradient magnitude of each point. Points close to the keypoint are given more emphasis than the ones far from it (see [19] for more details about the SIFT parameter setting). The orientation histograms of 4×4 sampling regions are calculated, each with eight orientation bins. Thus, a feature vector with a dimension of 128 ($4 \times 4 \times 8$) is produced.

Given these local features extracted from the original image pair or each of their corresponding OGM pairs in the gallery and probe sets, respectively, the two sets of keypoints on dorsal hands can be associated. Matching one keypoint to another is accepted only if the similarity distance is below a predefined threshold t times the distance to the second closest match. In this paper, t is empirically set at 0.6 as in [19]. The number of matched keypoints is accounted as the similarity measurement between the gallery and probe samples, and a larger matching score indicates a bigger probability that the hand-dorsa images are from the same hand.

B. Score Level Fusion

For a hand-dorsa vein image, we extract a set of keypoints of multiple sources, i.e., the multilevel detection based ones directly localized on the original image by the Harris–Laplace and Hessian–Laplace detectors as well as the ones detected by DoG on its corresponding OGMs at different orientations. As a result, multisource keypoint matches can be associated for identification. We then combine their similarity measurements at the matching score level to take all these contributions into account for final decision making.

Specifically, we denote the number of the matched keypoints by N_{Harr} for the ones localized using Harris–Laplace and by N_{Hess} for the ones detected employing Hessian–Laplace from the original hand-dorsa vein image pair; and by N_{OGM_o} for the ones found exploiting DoG from each of their corresponding OGM pairs at the oth direction. The bigger the value of N is, the more likely that the two dorsal hand images belong to the same subject, indicating that the similarity measurements, i.e., N_{Harr}, N_{Hess}, and N_{OGM_o}, are all with the positive polarity (a bigger value means a better matching relationship). A dorsal hand vein image in the probe set is compared with the ones in the gallery set, respectively, leading to a matching score vector. The nth element in a matching score vector corresponds to the similarity between the probe and the nth gallery sample. The score vectors from multiple sources are further normalized to the interval of [0, 1] using the max-min rule. These matching scores are finally fused by a basic weighted sum rule

$$S = \sum_{i=1}^{o+2} w_i \cdot S_i. \tag{12}$$

There are totally $o + 2$ similarity scores including the one of S_{Harr}, the one of S_{Hess}, and the ones (o) of S_{OGM_o}. The corresponding weight w_i is calculated dynamically during the online step using the scheme as in [26]

$$w_i = \frac{\max_1(S_i) - \text{mean}(S_i)}{\max_2(S_i) - \text{mean}(S_i)} \tag{13}$$

where the operators $\max_1(S)$ and $\max_2(S)$ produce the first and second maximum values of the score S, respectively. The gallery dorsal hand vein image that holds the maximum value is declared as the identity of the probe image.

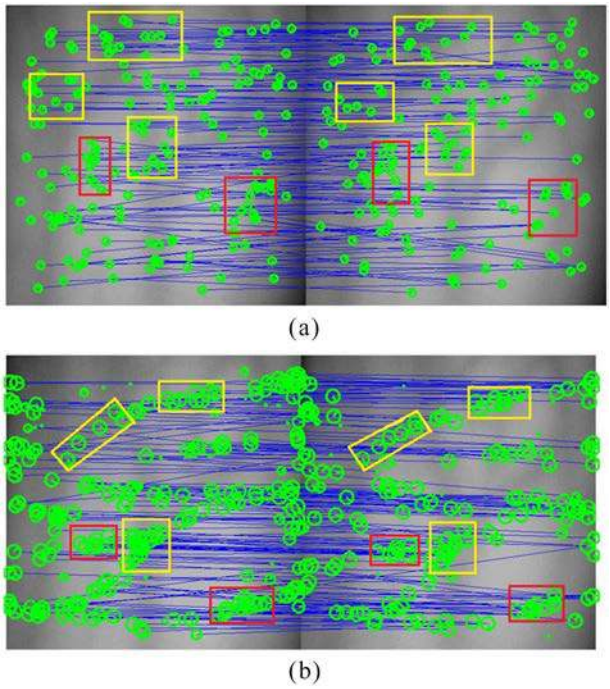

Fig. 9. Matching example between the dorsal hand vein images belonging to the same person based on these keypoints detected using (a) Harris–Laplace and (b) Hessian–Laplace. The matched keypoints marked in yellow boxes are located in the vein region and the ones in red boxes are located in the nearby subcutaneous tissue.

C. Illustration of Matching Samples

Fig. 9 displays a matching example adopting multilevel keypoint detection applied directly to the dorsal hand vein image: i.e., Harris–Laplace [Fig. 9(a)] and Hessian–Laplace [Fig. 9(b)]. According to the locations of the matched keypoints, they are highlighted by using two different colors. The ones in the vein area are marked in yellow whilst the ones in the surrounding subcutaneous tissue are marked in red. This figure highlights the following facts: 1) the positions of the keypoints provided by both the detectors, i.e., Harris–Laplace and Hessian–Laplace, are different and the two sets of points are complementary to each other and 2) these details outside the vein regions, i.e., these matched keypoints in the red boxes, are as important as those within the vein regions in final decision making.

Fig. 10 depicts a matching example between two hand-dorsa images of the same subject utilizing OGMs with DoG. We can observe a similar phenomenon and draw the same conclusion, i.e., the clues conveyed within the vein region (marked in yellow) and its surrounding subcutaneous tissue (marked in red) are both discriminative; meanwhile, these OGMs at different orientations contain complementary information.

VI. Experimental Results

In order to comprehensively evaluate the proposed method, we designed several experiments that are explicitly introduced in the subsequent. The experiments (in Sections VI-A–VI-F)

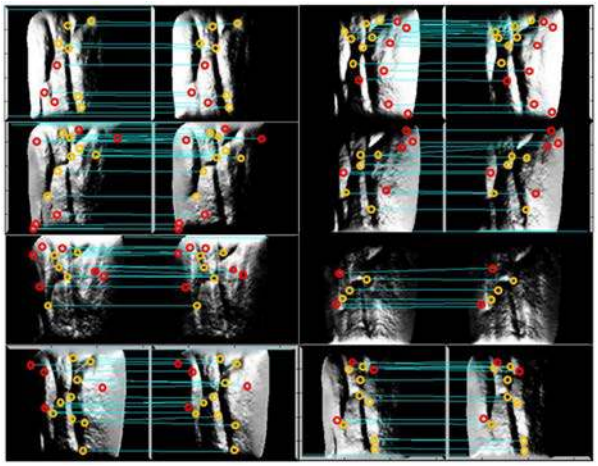

Fig. 10. Matching example using these OGM pairs of two left hands of the same person. The left column from top to bottom: OGM1 to OGM4; while the right column with the same order: OGM5 to OGM8. The matched keypoints marked in yellow are located in the vein area and the ones marked in red are detected in the nearby subcutaneous tissue.

were mainly conducted both in the scenarios of identification and verification as in the state-of-the-art work using the NCUT Part A database. In the meantime, to check the generalization ability of the proposed approach, we also carried out additional experiments (in Section VI-G) on NCUT Part B collected by a device whose cost is only a half as that for Part A, and its images are thus with more noise. Recall that both databases are among the largest ones of NIR hand-dorsa vein images. Part A contains ten right and ten left dorsal hand images, respectively, for each of the 102 subjects (totally 2040 samples), while Part B consists of the same number of images from both hands of 101 subjects (totally 2020 samples). All the hand images were roughly aligned thanks to the hardware configuration, but they still have moderate viewpoint (i.e., rotation and translation) and slight lighting intensity variations.

A. Effectiveness of Multilevel Keypoint Detection

We evaluated the effectiveness of the proposed multilevel keypoint detection approach in terms of the rank-one recognition rate in the scenario of identification. For experimental setup, the first five images of a subject were used in the gallery set and the remaining five images were exploited as probes. Because it was found out that the hand vein pattern is unique to some level for each person and each hand [27], we considered the left and right hand-dorsa vein images separately as if we had 204 different subjects each of which possesses ten samples in the dataset.

From the results in Table II, we can conclude in these points.

1) When we increase the number of these clustered centers (i.e., from 500 to 1000) for the keypoints detected by Hessian–Laplace, the rank-one recognition rate is improved, indicating that more keypoints lead to better accuracy. As we continue to increase it, the improvement is more and more limited. We thus set this number at 1000 to balance the accuracy and time cost in the following experiments to compute the performance of Hessian–Laplace.

TABLE II
RESULTS OF DIFFERENT DETECTORS, i.e., HARRIS–LAPLACE AND HESSIAN–LAPLACE AND THEIR DIFFERENT FUSION SCHEMES FOR THE MULTILEVEL KEYPOINT DETECTION BASED METHOD ON NCUT PART A

Harris	Hessian	*Sum*	*Product*	*Max*	*Min*
640kpt.: 96.27%	500kpt.: 83.53%	97.43%	87.65%	95.49%	86.08%
	700kpt.: 93.63%	97.55%	97.55%	95.49%	94.61%
	1000kpt.: 95.59%	97.55%	**98.04%**	95.88%	96.18%

TABLE III
PERFORMANCE OF EACH OGM AND THEIR COMBINATION IN THE SETUP OF LEFT-HAND ONLY, RIGHT-HAND ONLY, AND BOTH-HANDS ON THE NCUT PART A DATABASE

Directions	Left Hand	Right Hand	Both Hands
OGM-1	92.94%	93.53%	93.04%
OGM-2	81.18%	79.41%	78.53%
OGM-3	75.88%	77.45%	75.10%
OGM-4	73.14%	60.78%	66.18%
OGM-5	97.57%	92.55%	91.57%
OGM-6	78.82%	74.90%	75.49%
OGM-7	77.65%	84.51%	80.69%
OGM-8	78.63%	82.16%	78.82%
Fusion	99.02%	99.02%	99.02%

2) Making use of a comparable number of detected keypoints, the performance achieved by Harris–Laplace is superior to that of Hessian–Laplace, demonstrating that the keypoints localized by Harris–Laplace, including in particular the ones outside the vein areas, provide more discriminative information than those detected by Hessian–Laplace which mainly focuses on the vein regions. This phenomenon further illustrates the fact that these optical properties of subcutaneous tissue surrounding the vein network convey subject dependent cues.
3) No matter which score level fusion scheme (sum, product, max, and min rule) we take, the recognition rate is better than either of the Harris–Laplace or Hessian–Laplace, proving that the two detectors provide complementary clues to each other and highlighting the effectiveness of the multilevel keypoint detection approach. To keep the consistency in our approach, the sum rule was used in the following experiments to combine the results of Harris- and Hessian–Laplace.

B. Discriminative Power of OGMs

We then tested the discriminative power of the OGM based image representation in terms of the rank-one recognition rate in the identification scenario as well, following the same protocol as in the previous experiment. We calculated recognition rates of each OGM (for different quantized orientations) and their combination as displayed in Table III.

As we discussed in Section IV, each dorsal hand image has quite limited number of keypoints if DoG is directly applied to the original data, thus leading to a very partial description for the following matching step. This observation was our major

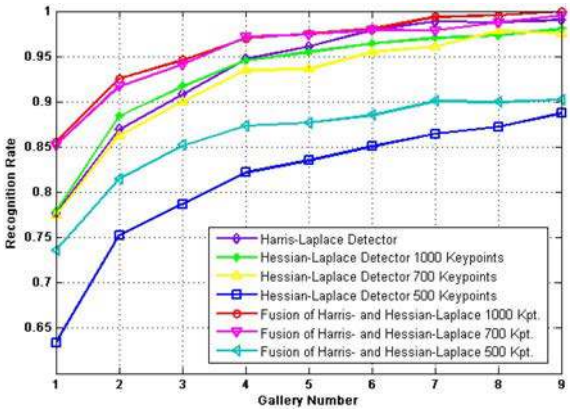

Fig. 11. Accuracy curves based on multilevel keypoint detection with respect to the gallery size of each subject on NCUT Part A.

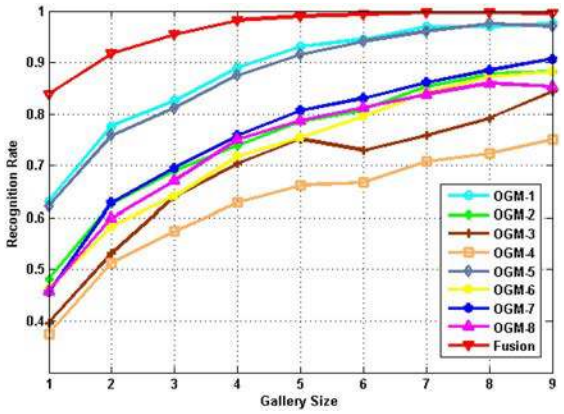

Fig. 12. Accuracy curves based on individual OGMs and their combination with respect to the gallery size of each subject on NCUT Part A.

motivation to develop OGMs which simulate the response of complex cells in the visual cortex in highlighting the gradients at different orientations. In Table III, we can see that the fusion of all these OGMs reaches a much better result than any of the single one. Such a fact accords with our preliminary study for this issue in adopting subspace techniques [28]. Unfortunately, in that work, due to the sensitivity of holistic methods to NIR intensity variations and hand geometric transformations, only about 70%–80% rank-one recognition rates were reported even with an easier experimental setup. Obviously, that performance is not accurate enough for a biometric system.

Meanwhile, we can see that the results of these OGMs are different, and the ones of OGM-1 and 5 are largely better than the others. The reason lies in that most of the dorsal hand veins are vertically distributed as shown in Fig. 8, which can be best highlighted by the horizontal gradient responses, i.e., OGM-1 and 5. Moreover, there exist a few horizontal and oblique vein furcations, and their corresponding best gradient responses are also necessary to comprehensively represent the entire venous network. As a result, the joint use of all these OGMs leads to the final highest score, indicated by the fusion performance.

On the other hand, we compared these results in the three columns of Table III, and found out that the performance only using left hand images was comparable to that only using right hand images. When left and right hand vein images were both used and considered as captured from different subjects, the result generally remains stable, showing that our method works well as the class size is doubled.

C. Impact of Gallery Size

An important property of a biometric system is its stability when the gallery size changes. For this purpose, we varied the number of gallery samples of each person from 1 to 9 (since at least one sample per person should be used in the probe set) to analyze the impact of the gallery size on the proposed method, employing the predefined experimental setup in identification. We can find that the rank-one recognition rate based on multilevel keypoint detection decreases from 97.55% to 85.57% (as in Fig. 11) and the accuracy by combining these OGMs at all orientations falls from 99.02% to 83.88% (as in Fig. 12) when the gallery size drops from 5 to 1. It indicates that the problem of limited enrolled samples seriously challenges the biometric system. In the meantime, we can also see that they both display certain robustness to such a challenge, and achieve acceptable rank-one recognition rates at 92.52% and 91.42%, respectively, when only two samples were enrolled as gallery for each subject.

TABLE IV
RANK-ONE RECOGNITION RATES OF KEYPOINT MATCHING IN EACH SOURCE AS WELL AS THEIR COMBINATION, i.e., MULTIPLE SOURCES, WITH RESPECT TO THE GALLERY SIZE OF EACH SUBJECT ON NCUT PART A

GalleryNum.	*Harris*	*Hessian*	*OGMs*	*Fusion*
1	77.67%	77.89%	83.88%	91.29%
2	86.95%	88.42%	91.42%	96.38%
3	90.83%	91.74%	94.82%	98.18%
4	94.69%	94.53%	97.88%	99.35%
5	96.27%	95.59%	99.02%	99.61%
6	97.92%	96.45%	99.14%	99.39%
7	98.86%	97.06%	99.67%	99.84%
8	98.77%	97.30%	99.75%	100.00%
9	99.02%	98.04%	99.51%	100.00%

We highlighted in the previous sections the complementarity of the two solutions proposed in this paper, i.e., the detection of multilevel keypoints which focuses on the elasticity and shape changes on the original hand-dorsa surface and DoG applied to its OGMs which simulate the response of complex cells in the visual cortex in highlighting these details through gradients at different orientations. A natural alternative to further improve the accuracy of the approach is to combine these scores of the two solutions to account for both their descriptive power. We adopt the weighted sum rule as defined in (12) for fusion, and final performance is significantly improved. The entire system reports a rank-one recognition rate up to 91.29% (as shown in Fig. 13 and Table IV) with only one image of each subject in the gallery set. As we can see in Table. IV and Fig. 13, these multisource keypoints are consistently complementary.

These cumulative match characteristic (CMC) curves related to different numbers (from 1 to 9) of enrolled samples in the gallery set of each subject are provided in Fig. 14.

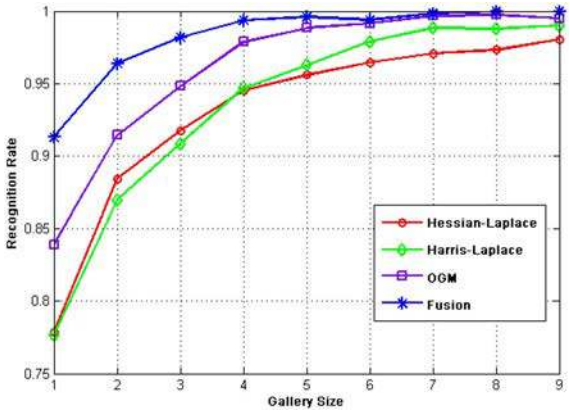

Fig. 13. Accuracy curves based on different sources of keypoint matching with respect to the gallery size of each subject on NCUT Part A.

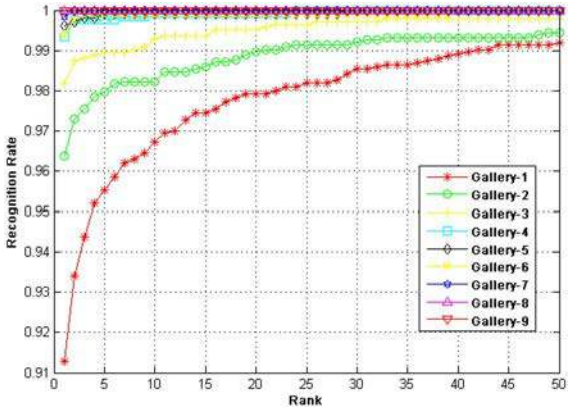

Fig. 14. CMC curves based on multisource keypoint matching of different numbers of gallery samples of each subject on NCUT Part A.

TABLE V
PERFORMANCE IN THE SCENARIO OF VERIFICATION OF THE PROPOSED METHOD ON THE NCUT PART A DATASET

Methods	FAR	VR	EER
Multi-level Keypoint	0.001	91.99%	1.91%
OGM based keypoints	0.001	93.63%	1.36%
Multi-source Keypoints	0.001	96.35%	0.81%

D. Verification Validation

We also performed experiments in the scenario of verification with the three modalities, i.e., multilevel keypoint detection on the original images, DoG based keypoint detection on these OGMs of the images, as well as the combination of the previous two modalities. For each subject, the first image was regarded as the gallery and the remaining images were treated as the probe samples to calculate the verification rates (VR) at the false acceptance rate (FAR) of 0.001 and the equal error rate (EER). Table V and Figs. 15 and 16 display these results, from which we can draw similar conclusions as in identification.

E. Comparison with the State of the Art

We compared the proposed method with the state of the art ones on NCUT Part A as illustrated in Table VI. Specifically, Wang *et al.* [29] firstly localized the vein network on the dorsal hand; then represented the detection result as a binary image, and finally applied SIFT for the matching step. Such an approach was originally introduced by Ladoux *et al.* [5] for the purpose of hand-palm vein identification. As we can see from that table, when only the vein regions are used as in [29], the accuracy is only 78.68% with the first four samples in the gallery and the other six as probes, thereby far behind the performance achieved by the proposed approach. This comparison confirms once more the importance of considering the optical properties of the whole hand-dorsa image. Our result is also higher than the best one reported in [29] that was achieved by adopting the relationship of multiple gallery samples of each subject. Wang *et al.* [8] employed an improved version

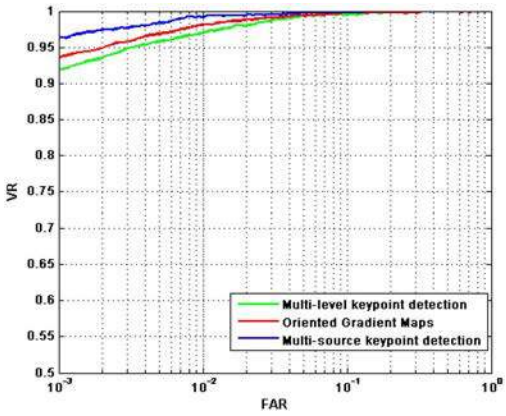

Fig. 15. Receiver operating characteristic (ROC) curves between FAR and VR of the proposed method on the NCUT Part A dataset.

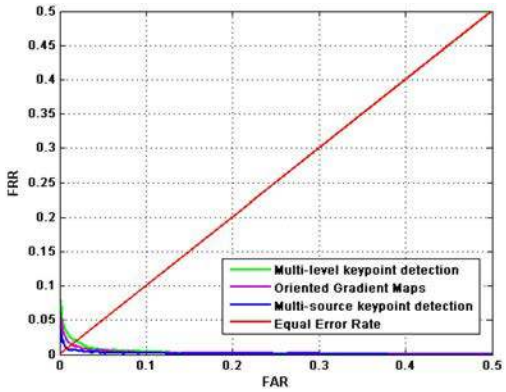

Fig. 16. ROC curves between FAR and false rejection rate (FRR) of the proposed method on the NCUT Part A dataset.

TABLE VI
COMPARISON WITH THE STATE OF THE ART IN RANK-ONE RECOGNITION RATE ON THE NCUT PART A DATASET

Approaches	Class	Gall./Prob.	Results
Multi-source Keypoint+SIFT	204	816/1224	**99.35%**
Binary+SIFT [29]	204	816/1224	78.68%
Best Binary+SIFT [29]	204	816/1224	97.95%
Multi-source Keypoint+SIFT	204	1020/1020	**99.61%**
Multi-level Keypoint + SIFT [20]	204	1020/1020	98.04%
OGMs + SIFT [21]	204	1020/2020	99.02%
CP-LBP [8]	204	1020/1020	90.88%
LBP+Graph [30]	204	1020/1020	96.67%
MLBP [6]	15	150	97.30%

TABLE VII
COMPARATIVE SUMMARY OF RELATED WORK ON DORSAL HAND VEIN BASED IDENTIFICATION AND VERIFICATION ON DIFFERENT DATABASES

Reference	Approach	Imaging	Database	Identification (G:P)	Verification (FAR; FRR; G:P)
1995 Cross and Smith [12]	Sequential correlation on vein signatures	NIR	20 Hands (5 Images)	–	0.00%; 5.00%; (3:2)
2004 Tanaka and Kubo [31]	FFT phase correlation and template matching	NIR	25 Hands (Unknown)	–	0.73%; 4.00%; (Unknown)
2004 Lin and Fan [3]	Multi-resolution filter representation	FIR	32 Hands (20 Images)	–	2.30%; 2.30%; (5:15)
2005 Wang and Leedham [32]	Hausdorff distance based line segment matching	FIR	12 Hands (9 Images)	–	0.00%; 0.00%; (3:6)
2005 Ding et al. [33]	Distances between crossing and end points	NIR	48 Hands (5 Images)	–	0.00%; 0.90%; (1120 Matches)
2006 Wang et al. [13]	Multi-supplemental features and multi-classifier fusion	NIR	100 Hands (5 Images)	–	0.00%; 0.50%; (3:2)
2008 Wang et al. [34]	Modified Hausdorff distance on segmented vein area	FIR	47 Hands (3 Images)	–	0.00%; 0.00%; (Unknown)
2008 Wang et al. [35]	Line Edge Mapping (LEM)/Gabor	NIR	100 Hands (5 Images)	66%/80%; (1:4)	–
2009 Kumar and Prathyusha [7]	Shape feature and triangulation	NIR	100 Hands (3 Images)	–	1.14%; 1.14%; (2:1) 3 Times
2010 Yuksel et al. [36]	Independent Component Analysis (ICA)	NIR	100 Hands (3 Images)	94.16%; (1:2)	2.47%; 2.47%; (1:2)
2012 Hsu et al. [37]	Block based 2D PCA and 2D LDA	NIR	214 Hands (20 Images)	98.55%; (10:10)	–
This paper	Multi-source keypoint based SIFT matching	NIR	204 Hands (10 Images)	99.61%; (5:5)	0.80%; 0.80%; (1:9)

TABLE VIII
RESULTS OF LEFT HAND ONLY, RIGHT HAND ONLY, AND THEIR FUSION USING DIFFERENT NUMBERS OF GALLERY SAMPLES ON THE NCUT PART A DATASET

Gallery Num.	Left Hand	Right Hand	Fusion
1	91.72%	91.18%	98.15%
2	96.94%	95.59%	99.51%
3	98.46%	97.76%	100.00%
4	99.35%	99.51%	100.00%
5	99.61%	99.41%	100.00%
6	99.51%	99.26%	100.00%
7	99.67%	100.00%	100.00%
8	100.00%	100.00%	100.00%
9	100.00%	100.00%	100.00%

TABLE IX
RANK-ONE RECOGNITION RATES OF KEYPOINT MATCHING IN EACH SOURCE AS WELL AS THEIR COMBINATION, i.e., MULTIPLE SOURCES, WITH RESPECT TO THE GALLERY SIZE OF EACH SUBJECT ON NCUT PART B

$GalleryNum.$	$Harris$	$Hessian$	$OGMs$	$Fusion$
1	70.02%	73.82%	79.26%	89.77%
2	82.98%	82.98%	90.90%	94.06%
3	85.01%	86.14%	94.27%	95.76%
4	88.45%	87.95%	95.54%	96.78%
5	91.78%	90.20%	96.93%	97.82%
6	94.55%	91.46%	98.02%	98.14%
7	95.05%	93.07%	98.42%	98.51%
8	96.53%	93.81%	99.01%	99.13%
9	97.52%	95.54%	99.01%	99.50%

TABLE X
PERFORMANCE IN THE SCENARIO OF VERIFICATION OF THE PROPOSED METHOD ON THE NCUT PART B DATASET

Methods	FAR	VR	EER
Multi-level Keypoint	0.001	90.29%	3.21%
OGM based keypoints	0.001	91.38%	3.08%
Multi-source Keypoints	0.001	96.31%	1.12%

of the local binary patterns (LBP), namely circular partition local binary patterns (CP-LBP), and achieved a recognition rate of 90.88% with five hand vein images in the gallery set and the remaining 5 ones used as probes. With this protocol, Zhu and Huang [30] evaluated their approach via hierarchically combining the LBP based texture features and graph matching based geometric features, and a rank-one recognition rate of 97.67% was reported. In our case, a better result is obtained by using such an experimental setup. A comparable performance was generated in [6], but in their experiments only a subset of the dataset (150 gallery and probe images of 15 persons) was exploited for evaluation. These facts clearly demonstrate the effectiveness of the proposed approach for dorsal hand vein recognition.

For further information and comparison, Table VII summarizes major state of the art approaches for the issue of dorsal hand vein based people identification and verification. As we can see, the proposed method achieves competitive results in both the scenarios of identification and verification while using a more comprehensive dataset.

F. Complementarity of Left and Right Hands

Since vein patterns are different to some level for both hands of the same individual [27], intuitively, the left and right hands of one person should possess complementary information for recognition. In the experiment, we further investigated such an answer to this problem by fusing the similarity measurement of each hand using the weighted sum rule as the other fusion steps in this paper. We can see from Table VIII that the accuracy based on the fusion of both hands (in the third column) always outperforms that based on either of single hand (in the first and second columns), and we can achieve a rank-one recognition rate of 98.15% even using one enrolled sample per hand for each subject. These results thus suggest that the use of left and right hands can further reinforce the robustness and the performance of the proposed hand vein based biometric system.

G. Evaluation on Generalization Ability

The previous experimental results suggest that the proposed method achieves very good performance on the NCUT Part A dataset. A key question, nevertheless, is whether it generalizes to other databases. We aim to answer this question using the novel NCUT Part B dataset. It roughly keeps the same size as NCUT Part A, but the hand vein images are more noisy since they were captured using a low cost device. We conducted both scenarios of identification as well as verification, and adopted the same protocols as in Tables IV and V, respectively.

We can observe from Tables IX and X that the conclusions achieved on NCUT Part B are consistent with the ones on Part A, i.e., the joint utilization of multiple keypoints to characterize these optical attributes of entire dorsal hand images improves the performance in comparison with those local feature-based approaches using single ones.

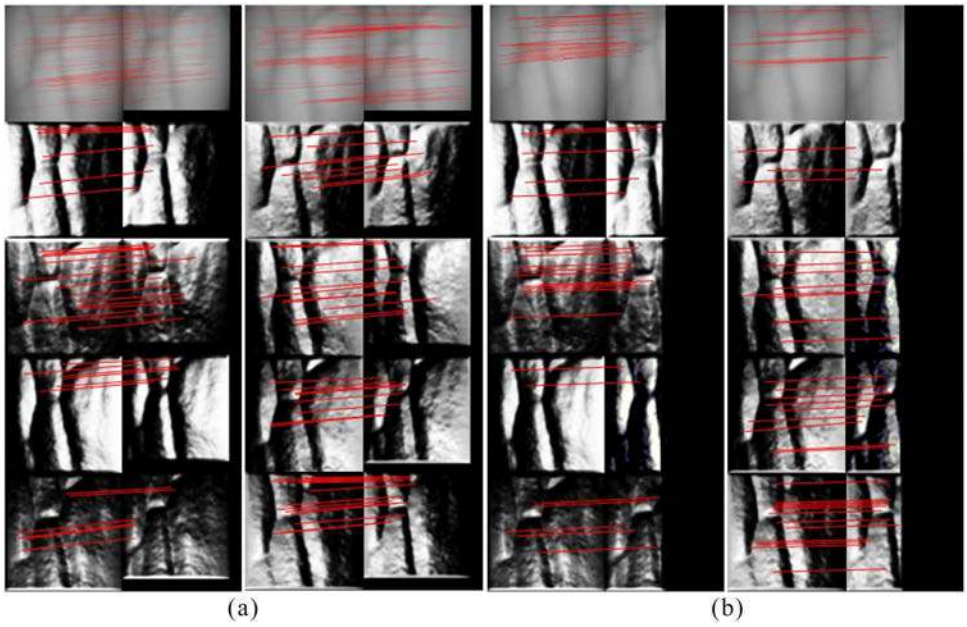

Fig. 17. Matching the keypoints across (a) scale variations and (b) translations (in both sub-figures, left column from top to bottom: Harris–Laplace, DoG on OGM1, DoG on OGM3, DoG on OGM5, and DoG on OGM7 and right column from top to bottom: Hessian–Laplace, DoG on OGM2, DoG on OGM4, DoG on OGM6, and DoG on OGM8).

TABLE XI
AVERAGE CONSUMED TIME OF EACH COMPONENT OF THE DORSAL HAND VEIN RECOGNITION SYSTEM

System Component		Average Consumed Time
Keypoint Detection	Harris-Laplace	$46ms$
	Hessian-Laplace	$72ms$
	DoG on OGMs	$66ms$
SIFT Feature Extraction		$52ms$
Matching		$4ms$

Furthermore, when we compare the accuracies on NCUT Part A and NCUT Part B (i.e., Table IV versus Table IX and Table V versus Table X), it can be seen that the proposed method reports competitive accuracies on NCUT Part B as well, quite close to the ones achieved on NCUT Part A. These results thereby suggest a quite good generalization skill of the proposed approach to dorsal hand vein recognition which further displays the robustness to noise caused by the cost decrease in the device of data acquisition.

H. Complexity Analysis

Real-life applications require a fast runtime in field deployment. The major cost centers in our system lie in multisource keypoint detection, SIFT feature extraction, and matching. We therefore focused our attention on optimizing these procedures. Primarily, we optimally coded them using C++. Given the fact that each process on a single type of keypoints, i.e., the ones provided by Harris–Laplace, Hessian–Laplace, and OGMs, is separated, we then utilized multithreading implementation of the entire system so that they can operate in parallel. In each process, the matching stage was further optimized and made 20 times faster in efficiency through GPU. Finally, the overall time cost is the maximum of the individual processes (i.e., the Hessian–Laplace based process). Table XI shows the details.

As we can see in Table XI, the method currently costs about 128 ms to achieve a 1-to-1 verification, including multisource keypoint detection (72 ms for the Hessian–Laplace based step) and SIFT feature extraction (52 ms) on the given probe and its matching with the gallery (4 ms), by using a machine equipped with two Intel (R) Xeon E5-2620 v2 CPUs (12-core, 2.6 GHz), 16 GB RAM, and a GTX 780 graphics card. When dealing with recognition, keypoint detection and local feature extraction on the probe are conducted online only once, while the time cost in matching is multiplied by the number of the gallery samples (1-to-N matching), leading to the computation cost of 524 ms (72 + 52 + 100 $\times$ 4 ms) of a 1-to-100 system.

I. Discussion

According to the experimental results, the proposed method outperforms its counterparts, thus proving more discriminative to distinguish NIR dorsal hand vein images, which is supported by two principal theoretical foundations. On the one hand, it depends on recent investigations in optical health science [14], presenting that people identification using hand vein images should not only focus on the vein network but also make use of these optical properties of the surrounding skin regions whose spectral response conveys subject-dependent inhomogeneous composition of blood and various chromophores and pigments. On the other hand, based on the progress achieved in psycho-visual studies [15], we interpret dorsal hand vein images as surfaces whose geometric characteristics, i.e., plateau, cliffs, ridges, valleys, etc., capture the subject-dependent variations of absorption and scattering attributes and could be perfectly characterized through curvature related quantities in differential geometry.

From the experimental viewpoint, local feature-based methods, see [8], [20], [21], [29], outperform holistic techniques,

such as PCA and LDA based subspace analysis [28], by a gap reaching more than 15 points on the NCUT Part A database. When we only focus on local feature-based methods, because we interpret hand vein images as surfaces to characterize their optical properties of hand skin as well as subcutaneous tissue in addition to vein network, the proposed approach employs multiorder (first, second, third) differential quantities closely related to differential geometry and hence provides a more comprehensive description of geometric properties in comparison with several existing local based methods that only exploit single type of features. The performance is hence better than that of SIFT [29], CP-LBP (an LBP variant) [8], or even a hybrid one by combining local and global features [30].

Additionally, the proposed approach employs the SIFT-like matching framework, and inherits the reputed robustness to in-plane rotation, scale changes, and translations, hence showing the potential to be competent in more difficult and complicated scenarios. Unfortunately, to the best of our knowledge, there is no publicly available database which contains these challenges. As such, we illustrate the robustness using artificial examples. For instance, Fig. 17 depicts the matching results across scale variation and translation. The two dorsal hand vein images are from the same subject in NCUT Part A. In Fig. 17(a), the left is of the original size while the right is resized to 90%. In Fig. 17(b), the left is fixed while half of the right is occluded due to translation. From the figure, we can see that even if in different scales or with large translations, the points detected on two images can still be correctly associated.

VII. Conclusion

This paper proposed a novel local feature-based approach to hand-dorsa vein recognition via matching keypoints localized through quantities of first to third order gradients closely related to differential geometry. In contrast to the state of the art work that only concentrates on the vein area, we demonstrated a key finding that the discrimination of a person by the dorsal hand vein image should focus not only on the vein network but also on the surrounding subcutaneous tissue whose optical properties are subject dependant as well. Furthermore, we interpreted the dorsal hand vein images and their distinctiveness enhanced representation, i.e., OGMs, as landscapes or surfaces, and generated a comprehensive description of their optical attributes by adopting geometric characteristics, modeled by Harris–Laplace, Hessian–Laplace and DoG based detectors and SIFT features. Extensive experiments conducted on NCUT Part A illustrated the effectiveness of the proposed approach, reaching the best performance so far reported on this database, both in identification and verification. Additional experimental results achieved on NCUT Part B also highlighted its robustness to low quality data.

In future work, we will investigate this issue in going one step further through a deep geometric approach which casts the matching of local features into a problem of surface registration, making use of full geometrical and topological properties of two dorsal hand vein surfaces. Furthermore, we will apply more powerful techniques [38] to improve the fusion of the matching results of multisource keypoints or even to combine with other hand biometrics (see [39]). In addition, we will also dedicate to building a representative dataset of the dorsal hand vein images containing significant variations in rotation, scale, illumination, etc., and experimentally test the robustness of the method to these factors.

References

[1] A. Kumar, M. Hanmandlu, and H. M. Gupta, "Online biometric authentication using hand vein patterns," in *Proc. IEEE Symp. Comput. Intell. Secur. Defence Appl.*, Ottawa, ON, Canada, 2009, pp. 1–7.

[2] P. MacGregor and R. Welford, "Imaging for security and personnel identification," *Adv. Imag.*, vol. 6, no. 7, pp. 52–56, 1991.

[3] C. Lin and K. Fan, "Biometric verification using thermal images of palm-dorsa vein patterns," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, no. 2, pp. 199–213, Feb. 2004.

[4] S. Malki and L. Spaanenburg, "Hand veins feature extraction using DT-CNNs," *Proc. SPIE*, vol. 6590, May 2007.

[5] P. Ladoux, C. Rosenberger, and B. Dorizzi, "Palm vein verification system based on SIFT matching," in *Proc. IEEE Int. Conf. Biometrics*, Alghero, Italy, 2009, pp. 1290–1298.

[6] S. Zhao, Y. Wang, and Y. Wang, "Biometric identification based on low quality hand vein pattern images," in *Proc. IEEE Int. Conf. Mach. Learn. Cybern.*, Kunming, China, 2008, pp. 1172–1177.

[7] A. Kumar and K. Prathyusha, "Personal authentication using hand vein triangulation and knuckle shape," *IEEE Trans. Image Process.*, vol. 18, no. 9, pp. 2127–2136, Sep. 2009.

[8] Y. Wang, K. Li, J. Cui, L. Shark, and M. Varley, "Study of hand-dorsa vein recognition," in *Proc. Int. Conf. Intell. Comput.*, Changsha, China, 2010, pp. 490–498.

[9] N. Miura, A. Nagasaka, and T. Miyatake, "Feature extraction of finger-vein pattern based on repeated line tracking and its application to personal identification," *Mach. Vis. Appl.*, vol. 15, no. 4, pp. 194–203, 2004.

[10] M. Heenaye-Mamode Khan, R. K. Subramanian, and N. Ali Mamode Khan, "Representation of hand dorsal vein features using a low dimensional representation integrating Cholesky decomposition," in *Proc. Int. Congr. Image Signal Process.*, Tianjin, China, 2009, pp. 1–6.

[11] J. Liu, D. Xue, J. Cui, and X. Jia, "Palm-dorsa vein recognition based on kernel principal component analysis and fisher linear discriminant," *J. Comput. Inf. Syst.*, vol. 8, no. 4, pp. 1545–1552, 2012.

[12] J. Cross and C. Smith, "Thermographic imaging of the subcutaneous vascular network of the back of the hand for biometric identification," in *Proc. Int. Carnahan Conf. Secur. Technol.*, Sanderstead, U.K., 1995, pp. 20–35.

[13] K. Wang, Y. Yuan, Z. Zhang, and D. Zhuang, "Hand vein recognition based on multi-supplemental features of multi-classifier fusion decision," in *Proc. Int. Conf. Mechatronics Autom.*, Luoyang, China, 2006, pp. 1790–1795.

[14] A. N. Bashkatov, E. A. Genina, and V. V. Tuchin, "Optical properties of skin, subcutaneous, and muscle tissues: A review," *J. Innovative Opt. Health Sci.*, vol. 4, no. 7, pp. 9–38, 2011.

[15] M. J. Morgan, "Features and the primal sketch," *Vis. Res.*, vol. 51, no. 7, pp. 738–753, 2011.

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

[17] K. Mikolajczyk and C. Schmid, "An affine invariant interest point detector," in *Proc. Eur. Conf. Comput. Vis.*, Copenhagen, Denmark, 2002, pp. 128–142.

[18] D. Huang, M. Ardabilian, Y. Wang, and L. Chen, "Oriented gradient maps based automatic asymmetric 3D-2D face recognition," in *Proc. IEEE/IAPR Int. Conf. Biometrics*, New Delhi, India, 2012, pp. 125–131.

[19] D. G. Lowe, "Distinctive image features from scale-invariant key-points," *Int. J. Comput. Vis.*, vol. 60, no. 4, pp. 91–110, 2004.

[20] Y. Tang, D. Huang, and Y. Wang, "Hand-dorsa vein recognition based on multi-level keypoint detection and local feature matching," in *Proc. Int. Conf. Pattern Recognit.*, Tsukuba, Japan, 2012, pp. 2837–2840.

[21] D. Huang, Y. Tang, Y. Wang, L. Chen, and Y. Wang, "Hand vein recognition based on oriented gradient maps and local feature matching," in *Proc. Asian Conf. Comput. Vis.*, Daejeon, Korea, 2012, pp. 430–444.

[22] P. M. Roth and M. Winter, "Survey of appearance-based methods for object recognition," Inst. Comput. Graph. Vis., Graz Univ. Technol., Graz, Austria, Tech. Rep., ICGTR-01/08, 2008.

[23] W. Kuhnel, *Differential Geometry: Curves, Surfaces, Manifolds*. Providence, RI, USA: American Mathematical Society, 2005.
[24] D. Huang, W. B. Soltana, and M. Ardabilian, "Textured 3D face recognition using biological vision-based facial representation and optimized weighted sum fusion," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit. Workshop Biometrics*, Colorado Springs, CO, USA, 2011, pp. 1–8.
[25] D. H. Hubel and T. N. Wiesel, "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex," *J. Physiol.*, vol. 160, pp. 106–154, Jan. 1962.
[26] A. S. Mian, M. Bennamoun, and R. Owens, "Keypoint detection and local feature matching for textured 3D face recognition," *Int. J. Comput. Vis.*, vol. 79, no. 1, pp. 1–12, 2008.
[27] A. M. Badawi, "Hand vein biometric verification prototype: A testing performance and patterns similarity," in *Proc. Int. Conf. Image Process. Comput. Vis. Pattern Recognit.*, 2006, pp. 3–9.
[28] L. Chen, D. Huang, H. Chiheb, and C. Ben Amar, "Increasing the distinctiveness of hand vein images by oriented gradient maps," in *Proc. Int. Conf. Special Interest Group Biometrics Electron. Signatures*, 2011, pp. 227–234.
[29] Y. Wang *et al*., "Hand vein recognition based on multiple keypoints sets," in *Proc. Int. Conf. Biometrics*, New Delhi, India, 2012, pp. 367–371.
[30] X. Zhu and D. Huang, "Hand dorsal vein recognition based on hierarchically structured texture and geometry features," in *Proc. Chin. Conf. Biometric Recognit.*, Guangzhou, China, 2012, pp. 157–164.
[31] T. Tanaka and N. Kubo, "Biometric authentication by hand vein patterns," in *Proc. SICE Annu. Conf.*, Sapporo, Japan, vol. 1. 2004, pp. 249–253.
[32] L. Wang and G. Leedham, "A thermal hand vein pattern verification system," in *Proc. Int. Conf. Adv. Pattern Recognit.*, Bath, U.K., 2005, pp. 58–65.
[33] Y. Ding, D. Zhuang, and K. Wang, "A study of hand vein recognition method," in *Proc. Int. Conf. Mechatronics Autom.*, 2005, pp. 2106–2110.
[34] L. Wang, G. Leedham, and D. S.-Y. Cho, "Minutiae feature analysis for infrared hand vein pattern biometrics," *Pattern Recognit.*, vol. 41, pp. 920–929, Mar. 2008.
[35] Z. Wang, B. Zhang, W. Chen, and Y. Gao, "A performance evaluation of shape and texture based methods for vein recognition," in *Proc. Congr. Image Signal Process.*, vol. 2. Sanya, China, 2008, pp. 659–661.
[36] A. Yuksel, L. Akarun, and B. Sankur, "Hand vein biometry based on geometry and appearance methods," *IET Comput. Vis.*, vol. 5, no. 6, pp. 398–406, 2011.
[37] C.-B. Hsu, J.-C. Lee, P.-Y. Kuei, and K.-C. Chan, "Combining local and global features based on the eigenspace for vein recognition," in *Proc. Int. Symp. Intell. Signal Process. Commun. Syst.*, New Taipei, Taiwan, 2012, pp. 401–405.
[38] X.-Z. Wang, R. Wang, H.-M. Feng, and H.-C. Wang, "A new approach to classifier fusion based on upper integral," *IEEE Trans. Cybern.*, vol. 44, no. 5, pp. 620–635, May 2014.
[39] R. Cappelli, M. Ferrara, and D. Maio, "A fast and accurate palmprint recognition system based on minutiae," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 42, no. 3, pp. 956–962, Jun. 2012.