IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 29 May 2024, accepted 23 June 2024, date of publication 10 July 2024, date of current version 18 July 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3425836

RESEARCH ARTICLE

Modified Jensen-Bregman LogDet Divergence for Target Detection With Region Covariance Descriptor

XIQIAN FAN[1] AND SHAOZHU YE[2]
[1]Hangzhou Institute for Advanced Research, Chinese Academy of Sciences, Hangzhou, Zhejiang 310024, China
[2]Hangzhou Wolei Intelligent Technology Company Ltd., Hangzhou, Zhejiang 310018, China

Corresponding author: Xiqian Fan (1723285323@qq.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 52372423 and in part by the Major Science and Technology Project of Zhejiang under Grant 2021C01136.

ABSTRACT In this study, we exploit the modified Jensen-Bregman LogDet (MJBLD) divergence to measure the dissimilarity between two region covariance descriptors extracted from an image, and design a target detection method based on this descriptor. In particular, MJBLD divergence, which considers the non-Euclidean geometric structure, is used as the measurement on the symmetric positive-definite (SPD) matrix manifold. The MJBLD divergence is a modified version of the Jensen-Bregman LogDet (JBLD) divergence which has many properties similar to the affine invariant Riemannian metric. Then, the MJBLD divergence is applied for the task of the image target detection where the image region of interest is represented as a covariance descriptor. The covariance descriptor is a SPD matrix which is constructed by the first and second gradients of intensity and the three-dimensional color information. Since the SPD matrix naturally resides on the non-Euclidean Riemannian manifold and the MJBLD divergence can be treated as a manifold metric, applying the non-Euclidean distance to SPD matrices can yield a better performance in comparison with the Euclidean distance. Experimental results show that our proposed method outperforms the state-of-the-art method.

INDEX TERMS Riemannian manifold, symmetric positive-definite matrix, image target detection, modified Jensen-Bregman LogDet divergence, region covariance descriptor.

I. INTRODUCTION

Target detection involves finding the target of interest from a two-dimensional image and then dividing it into many known types. Target detection is a significant problem in the fields of computer vision and image processing because it is closely related to applications in robotics [1], [2], [3], surveillance [4], registration, manipulation [5], and signal processing [6], [7], [8]. Several methods have been implemented for target detection. However, it still remains an ongoing research area. A successful example is to extract efficient image features to complete the task of the target detection. Typically, features used for target detection can be classified into two classes, local description features and global description features. Usually, global features characterize the whole region of a target, while local features are represented by some characters extracted from the part region [9]. Global feature-based target detection often employs an exhaustive strategy to search the image at various sizes and scales to discover the target of interest. In contrast to the exhaustive search using the local feature, the global feature is more expensive and sensitive to the change in rotation and scale. Extracting local features typically contains two steps. The salient region invariant to affine transformations is obtained first. Then, a descriptor of the detected region is established to make them discriminative. Classical feature-based approaches have achieved good performance in many applications due to their robustness to

The associate editor coordinating the review of this manuscript and approving it for publication was Shashikant Patil.

rotation, scale, illumination, and occlusions [9]. However, detecting target in the presence of varying appearance and the wide range of poses is a challenging task, and the detection performance needs to be improved.

Classical image features often rely on pixel information such as the intensity, image color, and its gradients, and have been widely used for many image processing tasks for a long time, for example, [10], [11], [12]. Unfortunately, the application of these features for image processing can be easily affected by illumination changes and nonrigid motion. A natural way to extend the pixel-information-based feature is to model the classical image feature as a histogram derived by a nonparametric estimation of an image region. In [13], histogram features extracted from the image region are used for target detection with a nonrigid motion. In [14] and [15], a fast histogram construction method is explored and used for searching the global match. The histogram feature has been used for target tracking [16], texture representation [17], and others. Moreover, different features can be combined with a joint representation via a histogram, but this representation has an expensive computational cost.

Recently, one of the most successful histogram features, the histograms of oriented gradients (HOG) [18], [19], [20], has attracted increasing attention. The HOG feature originates from the scale-invariant feature transformation (SIFT), and it can be viewed as a dense version of SIFT. The HOG feature mainly describes the contrast between contours and the background. Typical applications of the HOG feature are provided in [21] and [22], the HOG feature is used together with the support vector machine algorithm to detect target in the presence of varying appearance. This approach is robust under various conditions including illumination, distortion and environmental noise. Another example is given in [23], the authors exploit the HOG feature together with the local binary pattern (LBP) feature to implement target detection. However, the joint representation of the HOG feature and the other feature requires high computational cost. Nowadays, covariance matrices constructed by the image information have been used for feature description in the field of computer vision, and provide a compact framework of fusing different kinds of features. In contrast to the vector-form feature, covariance matrices can describe the correlation of the second-order feature information of data, and have been shown to provide a powerful representation for many tasks in contexts of image processing and computer vision, including texture categorization [24], [25], joint covariance descriptors for action recognition [26], [27], diffusion tensors-based medical image processing [28], and region covariance descriptors for pedestrian detection [29], [30].

In this paper, we establish a region covariance matrix from an image region via the color information, the coordination, and the first and second gradients of intensity. In particular, the MJBLD divergence that takes into account the non-Euclidean Riemannian structure is exploited to measure the dissimilarity of two region covariance matrices. A coarse-to-fine target detection method is presented and applied to target detection on the Inria person, the Fashion-MNIST and the Pascal VOC datasets. At the coarse detection stage, many similar regions are localized through the covariance matrix extracted from the whole region. To achieve fine detection, four covariance matrices are obtained from left, right, bottom, and top parts of an image region, and used to select the most similar region of the target. Numerical experiments are provided to demonstrate the superiority of this method.

The remainder of this paper can be organized as follows: Section II introduces how to construct the region covariance descriptor; the Riemannian geometry of symmetric positive-definite matrix is presented in Section III; the modified Jensen-Bregman LogDet divergence and its properties are detailed in Section IV; Section V implements numerical examples for target detection on three datasets; and conclusions are provided in Section VI.

Notation: In order to facilitate the description of this paper, some notations should be given. The math italic x, the lowercase bold $\mathbf{x}$, and the uppercase bold $\mathbf{X}$ denote the scalar, vector, and matrix, respectively. Symbols $\mathbf{X}^T$ stands for the transpose of the matrix $\mathbf{X}$. Symbol $|\mathbf{X}|$ denotes the determinant of matrix $\mathbf{X}$. $\mathbf{I}$ is the identity matrix. $\frac{\partial f(x)}{\partial x}$ represents the derivative of function $f(x)$ with respect to x. The conjugate of a complex data y is denoted by $\bar{y}$. Symbols $\|\mathbf{X}\|_F$ and $tr(\mathbf{X})$ denote the F-norm and the trace of the matrix $\mathbf{X}$, respectively.

II. REGION COVARIANCE DESCRIPTOR

A region in the image often contains much information, such as the image intensity, the color information, the coordinates of a pixel, and so on. Let I be a three-dimensional color image, $I(a, b)$ denotes the intensity of pixel (a, b). $R(a, b)$, $B(a, b)$, $G(a, b)$ denote three color values of the location (a, b), respectively. The norm of the first and second order derivatives of the intensities with respect to a and b are $|\frac{\partial I(a,b)}{\partial a}|$, $|\frac{\partial I(a,b)}{\partial b}|$, $|\frac{\partial^2 I(a,b)}{\partial a^2}|$, and $|\frac{\partial^2 I(a,b)}{\partial b^2}|$, respectively. A straightforward way to combine these features in the location (x, y) is to be represented as a nine-dimensional vector $\mathbf{z} = [a \quad b \quad R(a, b) \quad G(a, b) \quad B(a, b) \quad |\frac{\partial I(a,b)}{\partial a}| \quad |\frac{\partial I(a,b)}{\partial b}| \quad |\frac{\partial^2 I(a,b)}{\partial a^2}| \quad |\frac{\partial^2 I(a,b)}{\partial b^2}|]^T$. For an image region Q, suppose that the width is W, the height is H, and there are $n = W \times H$ pixels in this region, each element in the region is a nine-dimensional vector, and the i-th vector can be represented as $\mathbf{z}(i)$.

There are several advantages to using the covariance matrix as the region descriptor. The covariance matrix feature, extracted from an image region, can effectively represent the region across different views and poses. If the two distributions vary only in their covariance matrices, the covariance matrix contains all the sample information needed to discriminate between different distributions. Further, the covariance matrix can be used to fuse several kinds of correlate features. The autocorrelation of each feature is represented by the diagonal entries, and the correlation of multiple features is noted by the non-diagonal entries. It is worth noting that noise in any one feature can result in a large entry in the covariance matrix. Contrast to other region descriptors, such as the histogram, the covariance matrix is low dimensional. The covariance matrix is $(d^2 + d)/2$-dimensional, where d is the dimensionality of the features, and the joint feature histogram is q^d-dimensional, where q represents the number of histogram bins.

Suppose that a region R contains n pixels, each pixel is modeled as a d-dimensional vector $\mathbf{z}$, and $\{\mathbf{z}_j\}_{j=1}^{n}$ denote n feature points inside the region R. Then, the region R can be represented by a $d \times d$ covariance matrix $\mathbf{C}_R$ constructed as follows,

$$\mathbf{C}_R = \frac{1}{n-1}\sum_{k=1}^{n}(\mathbf{z}_k - \bar{\mathbf{z}})(\mathbf{z}_k - \bar{\mathbf{z}})^T, \tag{1}$$

where $\mathbf{C}_R$ is the covariance matrix estimated by the image information of the region R, and $\bar{\mathbf{z}}$ denotes the arithmetic mean of n vectors.

On the basis of this representation, each region in the image can be expressed as a covariance matrix. The (i, j)-th element of matrix $\mathbf{C}_R$ can be given by

$$\mathbf{C}_R(i,j) = \frac{1}{n-1}\sum_{k=1}^{n}(z_k(i) - \bar{z}(i))(z_k(j) - \bar{z}(j)). \tag{2}$$

Expand the mean $\bar{z}$ and rearrange the terms, then we can obtain

$$\mathbf{C}_R(i,j) = \frac{1}{n-1}[\sum_{k=1}^{n} z_k(i)z_k(j) - \frac{1}{n}\sum_{k=1}^{n} z_k(i)\sum_{k=1}^{n} z_k(j)]. \tag{3}$$

To obtain the covariance matrix in a given region R, the summation of the terms $\{\mathbf{z}_j\}_{j=1}^{n}$ and the summation of the multiplication $\{\mathbf{z}_j\}_{j=1}^{n}$ are needed to compute. Suppose R is a rectangular region, (w_1, h_1) is the upper left coordinate and (w_2, h_2) is the lower right coordinate, let G be the $d \times H \times W$ dimensional image feature obtained from the image I, then, the $d \times H \times W$ tensor P can be given as

$$P(x_1, y_1, m) = \sum_{a<w_1, b<h_1} G(a, b, m), m = 1, \ldots, d. \tag{4}$$

Additionally, the $d \times d \times H \times W$ tensor of the second order image feature U is given as

$$U(w_1, h_1, m, n) = \sum_{a<w_1, b<h_1} G(a, b, m)G(a, b, n), \quad m, n = 1, \ldots, d. \tag{5}$$

Assume that $\mathbf{P}_{a,b}$ is a d-dimensional vector and $\mathbf{U}_{a,b}$ is a $d \times d$-dimensional matrix, then, $\mathbf{P}_{a,b}$ and $\mathbf{U}_{a,b}$ can be formulated as

$$\mathbf{P}_{a,b} = [P(a, b, 1) \ldots P(a, b, d)]^T,$$
$$\mathbf{U}_{a,b} = \begin{pmatrix} U(a, b, 1, 1) & \cdots & U(a, b, 1, d) \\ \vdots & \vdots & \vdots \\ U(a, b, d, 1) & \cdots & U(a, b, d, d) \end{pmatrix}. \tag{6}$$

Note that $\mathbf{U}_{a,b}$ is a symmetric matrix with $(d^2 + d)/2$ dimensions and $\mathbf{P}_{a,b}$ is a d-dimensional vector. The computational complexity of computing $\mathbf{P}$ and $\mathbf{U}$ are $O(d^2WH)$. Let $R(w_1, h_1; w_2, h_2)$ be the rectangular region, the covariance matrix of the region bounded by $(1, 1)$ and (w_1, h_1) is estimated as

$$\mathbf{C}_{R(1,1;w_1,h_1)} = \frac{1}{n-1}[\mathbf{U}_{w_1,h_1} - \frac{1}{n}\mathbf{P}_{w_1,h_1}\mathbf{P}^T_{w_1,h_1}], \tag{7}$$

where $n = w_1h_1$ is the number of points in the region R. Through a serial of manipulations, the covariance matrix of the region $R(w_1, h_1; w_2, h_2)$ can be computed as

$$\begin{aligned}&\mathbf{C}_{R(w_1,h_1;w_2,h_2)}\\ &= \frac{1}{n-1}[\mathbf{U}_{w_1,h_1} + \mathbf{U}_{w_2,h_2} - \mathbf{U}_{w_2,h_1} - \mathbf{U}_{w_1,h_2}\\ &\quad - \frac{1}{n}(\mathbf{P}_{w_1,h_1} + \mathbf{P}_{w_2,h_2} - \mathbf{P}_{w_1,h_2} - \mathbf{P}_{w_2,h_1})(\mathbf{P}_{w_1,h_1} + \mathbf{P}_{w_2,h_2}\\ &\quad - \mathbf{P}_{w_1,h_2} - \mathbf{P}_{w_2,h_1})^T],\end{aligned} \tag{8}$$

where $n = (w_2 - w_1)(h_2 - h_1)$. According to Eq.(8), the region covariance descriptor can be established. Given a region R, its corresponding covariance matrix is invariant to the scale and rotation in different images as it does not contain any information about the number of points and the order of the matrix. However, if the information regarding the gradient (scale) with respect to the location is included in the covariance matrix, the covariance matrix is sensitive to the rotation (scale). The region covariance matrix estimated by Eq.(8) is a symmetric positive-definite (SPD) matrix. SPD matrices naturally lie on the non-Euclidean Riemannian manifold, which will be introduced in the subsequent text.

III. RIEMANNNIAN GEOMETRY OF SYMMETRIC POSITIVE-DEFINITE MATRIX MANIFOLD

A Riemannian manifold is a non-linear mathematical space, where a point x on the manifold has a local neighbourhood that is differentiable homeomorphism with its tangent space $T_x\mathcal{M}$ (Euclidean space). The tangent space of a point on the Riemannian manifold defines an inner product, which induces the norm $\|y\|_x^2 = \langle y, y\rangle_x$. As stated in [31] and [32] that the geometric structure of the Riemannian manifold is determined by a Riemannian metric, which can reflect the powerful framework to work on the manifold. SPD matrices form a connected Riemannian manifold Sym_d^+ constructed by a set of SPD matrices, where each point on this manifold is a SPD matrix. A Riemannian metric can be defined as [33]

$$\langle \mathbf{Y}, \mathbf{Z}\rangle_{\mathbf{X}} = tr(\mathbf{X}^{-\frac{1}{2}}\mathbf{Y}\mathbf{X}^{-1}\mathbf{Z}\mathbf{X}^{-\frac{1}{2}}). \tag{9}$$

As is known that by choosing a point $\mathbf{X}$ on the SPD manifold and a vector $\vec{xy}$ on the tangent space $T_{\mathbf{X}}\mathcal{M}$ of the point $\mathbf{X}$, only one geodesic starting from $\mathbf{X}$ with the tangent vector. A geodesic is the shortest curve connected two points on the manifold. The exponential mapping, which maps a point on the tangent space $T_{\mathbf{X}}\mathcal{M}$ to the manifold, is defined as

$$exp_{\mathbf{X}}\mathbf{Y} = \mathbf{X}^{\frac{1}{2}}exp(\mathbf{X}^{-\frac{1}{2}}\mathbf{Y}\mathbf{X}^{-\frac{1}{2}})\mathbf{X}^{\frac{1}{2}}. \tag{10}$$

The exponential mapping is a function that is defined on the tangent space $T_X\mathcal{M}$. The mapping is not a global diffeomorphism but only a local one, as the one-to-one mapping is meet on the local neighbourhood of the point $\mathbf{X}$. Therefore, the inverse function, that is the logarithm mapping, is defined as

$$\log_{\mathbf{X}} \mathbf{Y} = \mathbf{X}^{\frac{1}{2}} \log(\mathbf{X}^{-\frac{1}{2}}\mathbf{Y}\mathbf{X}^{-\frac{1}{2}})\mathbf{X}^{\frac{1}{2}}. \tag{11}$$

For the case of SPD matrix, the matrix can be computed easily by eigenvalue decomposition, as

$$\Sigma = \mathbf{U}\mathbf{D}\mathbf{U}^T = \mathbf{U}diag(\lambda_i)\mathbf{U}^T. \tag{12}$$

The exponential and logarithm mapping are given as

$$\begin{aligned} exp(\Sigma) &= \mathbf{U}diag(exp(\lambda_i))\mathbf{U}^T, \\ \log(\Sigma) &= \mathbf{U}diag(\log(\lambda_i))\mathbf{U}^T. \end{aligned} \tag{13}$$

Based on this metric and the exponential and logarithmic mapping, a geodesic distance is given as follows,

$$\begin{aligned} d_R^2(\mathbf{X}, \mathbf{Y}) &= \langle \log_{\mathbf{X}} \mathbf{Y}, \log_{\mathbf{X}} \mathbf{Y} \rangle_{\mathbf{X}} \\ &= tr(\mathbf{X}^{-\frac{1}{2}}\mathbf{X}^{\frac{1}{2}} \log(\mathbf{X}^{-\frac{1}{2}}\mathbf{Y}\mathbf{X}^{-\frac{1}{2}})\mathbf{X}^{\frac{1}{2}}\mathbf{X}^{-1}\mathbf{X}^{\frac{1}{2}} \\ &\quad \times \log(\mathbf{X}^{-\frac{1}{2}}\mathbf{Y}\mathbf{X}^{-\frac{1}{2}})\mathbf{X}^{\frac{1}{2}}\mathbf{X}^{-\frac{1}{2}}) \\ &= tr(\log^2(\mathbf{X}^{-\frac{1}{2}}\mathbf{Y}\mathbf{X}^{-\frac{1}{2}})) \\ &= tr(\log^2(\mathbf{X}^{-1}\mathbf{Y})). \end{aligned} \tag{14}$$

Thus, the geodesic distance which is also called the Affine Invariant Riemannian Metric (AIRM) can be formulated as [34]

$$d_R(\mathbf{X}, \mathbf{Y}) = \| \log(\mathbf{X}^{-1}\mathbf{Y})\|_F. \tag{15}$$

In addition to the geodesic distance, lots of divergences can be exploited to measure the dissimilarity between two points on the SPD manifold. In the next section, the modified Jensen-Bregman LogDet (MJBLD) divergence is formally discussed.

IV. MODIFIED JENSEN-BREGMAN LOGDET DIVERGENCE

A Riemannian manifold can be endowed with different information divergences. Different divergences reflect different geometric structures of the Riemannian manifold. Here, we introduce a new divergence, that is a modified version of the JBLD divergence [35]. In the following, the JBLD divergence is introduced first, and then the MJBLD divergence is presented.

Given two vectors $\mathbf{u}$ and $\mathbf{v}$, the Bregman divergence $d_\varphi : L \times relint(L) \rightarrow [0, \infty)$ is defined as

$$d_\varphi(\mathbf{u}, \mathbf{v}) = \varphi(\mathbf{u}) - \varphi(\mathbf{v}) - \langle \mathbf{u} - \mathbf{v}, \nabla\varphi(\mathbf{v})\rangle, \tag{16}$$

where $\varphi : L \subseteq \mathbb{R}^d \rightarrow \mathbb{R}$ denotes a Legendre type function and strictly convex on int(dom L). ∇ denotes the differentiation of a function. As the Bregman divergence is asymmetric, it loses many useful properties. As a consequence, a substantial interest is focused on a symmetrized version, the so-called JBLD divergence, which is given by [36]

$$J_\varphi(\mathbf{u}, \mathbf{v}) = \frac{1}{2}(d_\varphi(\mathbf{u}, \mathbf{s}) + d_\varphi(\mathbf{s}, \mathbf{v})), \tag{17}$$

where $\mathbf{s} = (\mathbf{u} + \mathbf{v})/2$.

Eq.(16) and Eq.(17) can be naturally extended to the case of SPD matrix by instead using the eigenvalue map λ of the convex function φ, and by substituting the trace for the inner product used in Eq.(16). Thus, the Bregman divergence between the given SPD matrices $\mathbf{U}$ and $\mathbf{V}$, is defined as

$$B_\varphi(\mathbf{U}, \mathbf{V}) = \varphi(\mathbf{U}) - \varphi(\mathbf{V}) - \langle \mathbf{U} - \mathbf{V}, \nabla\varphi(\mathbf{V})\rangle. \tag{18}$$

Similarly, for two SPD matrices $\mathbf{U}$ and $\mathbf{V}$, the JBLD divergence between them can be derived by employing $\varphi(\mathbf{U}) = -\log|\mathbf{Y}|$ as the seed function, as follows,

$$J_{ld} = \frac{1}{2}(B_\varphi(\mathbf{U}, \mathbf{S}) + B_\varphi(\mathbf{S}, \mathbf{V})), \mathbf{S} = (\mathbf{U} + \mathbf{V})/2. \tag{19}$$

Substitute Eq.(18) into Eq.(19), and we can obtain

$$\begin{aligned} J_{ld} &= \frac{1}{2}(\varphi(\mathbf{U}) - \varphi(\mathbf{S}) - \langle \mathbf{U} - \mathbf{S}, \nabla\varphi(\mathbf{S})\rangle + \varphi(\mathbf{S}) - \varphi(\mathbf{V}) \\ &\quad - \langle \mathbf{S} - \mathbf{V}, \nabla\varphi(\mathbf{V})\rangle) \\ &= \frac{1}{2}(\varphi(\mathbf{U}) - \varphi(\mathbf{V}) - \langle \mathbf{U} - \mathbf{S}, \nabla\varphi(\mathbf{S})\rangle \\ &\quad - \langle \mathbf{S} - \mathbf{V}, \nabla\varphi(\mathbf{V})\rangle). \end{aligned} \tag{20}$$

Plug $\varphi(\mathbf{U}) = -\ln|\mathbf{U}|$ and $\mathbf{S} = (\mathbf{U} + \mathbf{V})/2$ into Eq.(20), and we have

$$J_{ld} = \ln\left|\frac{\mathbf{U} + \mathbf{V}}{2}\right| - \frac{1}{2}\ln|\mathbf{U}\mathbf{V}|. \tag{21}$$

It can be noted from Eq.(21) that the JBLD divergence $J_{ld}(\mathbf{x}, \mathbf{y})$ is symmetric, nonnegative, and definite [37]. Many properties about the JBLD divergence can be summarized as follows:

1. nonnegativity: $J_{ld}(\mathbf{U}, \mathbf{V}) \geq 0$,
2. definiteness: $J_{ld}(\mathbf{U}, \mathbf{V}) = 0 \quad \text{iff} \quad \mathbf{U} = \mathbf{V}$,
3. symmetry: $J_{ld}(\mathbf{U}, \mathbf{V}) = J_{ld}(\mathbf{V}, \mathbf{U})$,
4. triangle inequality: $\sqrt{J_{ld}(\mathbf{U}, \mathbf{V})} \leq \sqrt{J_{ld}(\mathbf{U}, \mathbf{Z})} + \sqrt{J_{ld}(\mathbf{Z}, \mathbf{V})}$,
5. affine invariance: $J_{ld}(\mathbf{AUB}, \mathbf{AVB}) = J_{ld}(\mathbf{U}, \mathbf{V})$ for invertible matrices $\mathbf{A}$ and $\mathbf{B}$,
6. invariance to inversion: $J_{ld}(\mathbf{U}^{-1}, \mathbf{V}^{-1}) = J_{ld}(\mathbf{U}, \mathbf{V})$.

Additionally, J_{ld} is commonly used as a proxy for the AIRM D_R^2 due to its close relation to the Riemannian metric that is given in Theorem 1.

Theorem 1: Let $\mathbf{U}, \mathbf{V} \in \mathcal{S}_{++}^d$. Then,

$$J_{ld}(\mathbf{U}, \mathbf{V}) \leq D_R^2(\mathbf{U}, \mathbf{V}), \tag{22}$$

and if $M\mathbf{I} \succeq \mathbf{U}, \mathbf{V} \succeq m\mathbf{I} \succ 0$, then,

$$D_R^2(\mathbf{U}, \mathbf{V}) \leq 2\ln(M/m)(J_{ld}(\mathbf{U}, \mathbf{V}) + \gamma), \quad \gamma = d\ln 2. \tag{23}$$

Proof 1 (Proof of Theorem 1): As $\mathbf{U}$ and $\mathbf{V} \in \mathcal{S}_{++}^d$ are positive, the eigenvalues $\lambda(\mathbf{U}\mathbf{V}^{-1}) > 0$. Let $v_i =$

$\lambda_i(\mathbf{U}\mathbf{V}^{-1}) = e^{u_i}$, then, the AIRM between $\mathbf{U}$ and $\mathbf{V}$ can be rewritten as

$$D_R^2(\mathbf{U}, \mathbf{V}) = \|\mathbf{u}\|_2, \tag{24}$$

and the JBLD is given as

$$J_{ld}(\mathbf{U}, \mathbf{V}) = \sum_{i=1}^{d}(\ln(1+e^{u_i}) - u_i/2 - \ln 2). \tag{25}$$

According to affine invariance, we have

$$\begin{aligned} J_{ld}(\mathbf{U}, \mathbf{V}) &= J_{ld}(\mathbf{U}\mathbf{V}^{-1}, \mathbf{V}\mathbf{V}^{-1}) \\ &= J_{ld}(\mathbf{U}\mathbf{V}^{-1}, \mathbf{I}) \\ &= J_{ld}(\mathbf{I}, \mathbf{U}\mathbf{V}^{-1}) \\ &= \ln|\mathbf{I} + \mathbf{U}\mathbf{V}^{-1}| - \frac{1}{2}\ln|\mathbf{U}\mathbf{V}^{-1}| - \ln 2^d. \end{aligned} \tag{26}$$

Let

$$f(x) = x^2 - \ln(1+e^x) + x/2 + \ln 2, \tag{27}$$

the second derivative of $f(x)$ can be given as

$$f''(x) = 2 - \frac{e^x}{(1+e^x)^2}. \tag{28}$$

It is clear that the function $f(x)$ is convex as $f''(x) > 0$. Moreover, the first derivative of $f(x)$ is given as

$$f'(x) = 2x - e^x/(1+e^x) + 1/2. \tag{29}$$

Set $f'(x) = 0$, and we have $x^* = 0$. Thus, for all $x \in \mathbb{R}$, the formulation $f(x) \geq f(x^*) = 0$ holds, this means that

$$\sum_{i=1}^{d} f(u_i) = D_R^2(\mathbf{U}, \mathbf{V}) - J_{ld}(\mathbf{U}, \mathbf{V}) \geq 0. \tag{30}$$

To prove the inequality Eq.(23), let's first note that

$$\sum_{i=1}^{d}(|u_i|/2 - \log 2) \leq \sum_{i=1}^{d}(-u_i/2 - \ln 2 + \ln(1+e^{u_i})), \tag{31}$$

Then, we have the bound

$$J_{ld}(\mathbf{X}, \mathbf{Y}) + d\ln 2 \geq \frac{1}{2}\|u\|_1. \tag{32}$$

Substitute the Holder's inequality $u^T n \leq \|u\|_\infty \|u\|_1$ into Eq.(32), and we can obtain the bound

$$2\|u\|_\infty (J_{ld} + d\ln 2) \geq \|u\|_2^2 = D_R^2(\mathbf{U}, \mathbf{V}). \tag{33}$$

In addition, $M\mathbf{I} \succeq \mathbf{U}, \mathbf{V} \succeq m\mathbf{I}$ implies that $\ln(M/m) \geq \|u\|_\infty$.

Based on the definition of the JBLD divergence (21), we define a divergence function by replacing the matrix determinant with the matrix trace, as

$$(\mathbf{U}, \mathbf{V}) \mapsto tr(\frac{\mathbf{U}+\mathbf{V}}{2}) - tr(\mathbf{U}^{\frac{1}{2}}\mathbf{V}^{\frac{1}{2}}). \tag{34}$$

It is obvious from (34) that this function is positive and equals to zero if $\mathbf{U} = \mathbf{V}$. Indeed, (34) is a divergence function which is called the modified JBLD divergence.

Proposition 1: The modified JBLD divergence is a symmetric function on the SPD manifold.

Proof 2 (Proof of Proposition 1): The modified JBLD divergence function can be rewritten as

$$D_{MJ}(\mathbf{U}, \mathbf{V}) = \frac{1}{2}tr((\mathbf{U}^{\frac{1}{2}} - \mathbf{V}^{\frac{1}{2}})^2) = \frac{1}{2}\|\mathbf{U}^{\frac{1}{2}} - \mathbf{V}^{\frac{1}{2}}\|^2. \tag{35}$$

It is clear that (35) is a symmetric function. When matrices $\mathbf{U}$ and $\mathbf{V}$ commute, $D_{MJ}(\mathbf{U}, \mathbf{V}) = D_{MJ}(\mathbf{V}, \mathbf{U})$.

Proposition 2: The modified JBLD divergence satisfies the following triangle inequality,

$$D_{MJ}(\mathbf{U}, \mathbf{V}) \leq 2(D_{MJ}(\mathbf{U}, \mathbf{W}) + D_{MJ}(\mathbf{W}, \mathbf{V})). \tag{36}$$

Proof 3 (Proof of Proposition 2): According to (35), the following triangle inequality can be given,

$$\sqrt{D_{MJ}(\mathbf{U}, \mathbf{V})} \leq \sqrt{D_{MJ}(\mathbf{U}, \mathbf{W})} + \sqrt{D_{MJ}(\mathbf{W}, \mathbf{V})}. \tag{37}$$

Square left and right of the inequality, and we have

$$\begin{aligned} D_{MJ}(\mathbf{U}, \mathbf{V}) \leq\ & D_{MJ}(\mathbf{U}, \mathbf{W}) + D_{MJ}(\mathbf{W}, \mathbf{V}) \\ & + 2\sqrt{D_{MJ}(\mathbf{U}, \mathbf{W}) D_{MJ}(\mathbf{W}, \mathbf{V})}. \end{aligned} \tag{38}$$

Then we can obtain the following inequality

$$\sqrt{D_{MJ}(\mathbf{U}, \mathbf{W}) D_{MJ}(\mathbf{W}, \mathbf{V})} \leq \frac{1}{2}(D_{MJ}(\mathbf{U}, \mathbf{W}) + D_{MJ}(\mathbf{W}, \mathbf{V})). \tag{39}$$

Proposition 3: The modified JBLD divergence is invariant under congruence transformations.

$$D_{MJ}(\mathbf{W}\mathbf{U}\mathbf{W}^T, \mathbf{W}\mathbf{V}\mathbf{W}^T) = D_{MJ}(\mathbf{U}, \mathbf{V}). \tag{40}$$

It is noted that the modified JBLD divergence is not invariant under the inversion, as $D_{MJ}(\mathbf{U}, \mathbf{V}) = D_{MJ}(\mathbf{U}^{-1}, \mathbf{V}^{-1})$.

An outstanding advantage of the MJBLD and the JBLD divergences against the Riemannian distance is its computational complexity. Specifically, D_{MJ} can be computed by a matrix multiplication $\mathbf{U}$ and $\mathbf{V}$, and it requires $(1/2)d^3$ flops. J_{ld} need only computation of determinants, which is completed through three Cholesky factorizations ($\mathbf{U}+\mathbf{V}$, $\mathbf{U}$ and $\mathbf{V}$), and each factorization requires $(1/3)d^3$ flops. However, the Riemannian distance D_R is computed via eigenvalues, and it can be derived for SPD matrices in approximately $4d^3$ flops. Therefore, D_{MJ} and J_{ld} are much faster than D_R. Lots of trials are provided in Table 1.

V. EXPERIMENTS

Detecting targets in images is challenging due to varying appearances and a wide range of poses. Many methods have been implemented for this purpose but it still remains as an ongoing area of research. In this Section, we aim to implement and evaluate our proposed method. The region covariance descriptor is applied to extract the image feature for target detection, and the MJBLD divergence is used as the dissimilarity measure between two covariance descriptors. Given an image, the goal of target detection is to find the location of the target in a given image. In this paper,

TABLE 1. Time consuming (seconds) of calculating function values over 5 ∼ 1000 trials.

d	5	10	20	50	100	200	500	1000
D_R	0.03	0.041	0.088	0.36	1.321	8.263	77.611	499.635
J_{ld}	0.026	0.038	0.068	0.131	0.401	2.359	22.763	119.903
D_{MJ}	0.023	0.033	0.061	0.125	0.326	2.101	20.090	114.663

we exploit the pixel coordinates (a, b), the color information (RGB) and the first and second gradients of the intensity with respect to a and b coordination. Then, each location of the image can be represented as a 9-dimensional feature vector, as $[a \quad b \quad R(a,b) \quad G(a,b) \quad B(a,b) \quad |\frac{\partial I(a,b)}{\partial a}| \quad |\frac{\partial I(a,b)}{\partial b}| \quad |\frac{\partial^2 I(a,b)}{\partial a^2}| \quad |\frac{\partial^2 I(a,b)}{\partial b^2}|]^T$, where the first and second order derivative can be obtained through the filters $[-1 \quad 2 \quad -1]^T$ and $[-1 \quad 0 \quad 1]^T$, respectively. For an image region, the covariance descriptor is a 9×9 covariance matrix and can be derived as (1).

To conduct the target detection experiment, the image target is represented by five covariance matrices. As illustrated in Figure 1, these five covariance matrices are computed inside the overlapping image region. Specifically, $\mathbf{C}_i, i = 1, \ldots, 5$ are derived in the whole, left part, right part, top part, and bottom part of the region. A two-stage scheme is employed to detect the image target from coarse to fine steps. At the first stage, covariance matrices derived from the whole region with nine different scales are used to search a region where covariance matrices of all locations are close, and the dissimilarity of any two matrices is measured by the MJBLD divergence. We do not change the scale of the target image but the size of the search window. During the process, there is a 0.15 scaling factor in different search windows.

At the second stage, 1000 best matching locations are selected, and we repeat to search for these 1000 locations by exploiting these five SPD matrices $\mathbf{C}_i, i = 1, \ldots, 5$. Given a target model, the dissimilarity between region covariance descriptors in the search region and the target model is calculated by

$$d(T, R) = \underset{\min}{j} \left[\sum_{i=1}^{5} d(\mathbf{C}_i^T, \mathbf{C}_i^R) - d(\mathbf{C}_j^T, \mathbf{C}_j^R)\right], \tag{41}$$

where $\mathbf{C}_i^T$ and $\mathbf{C}_i^R$ are covariance descriptors of the target model and the search region, respectively. The matching region is regarded as the region with the smallest dissimilarity with respect to the covariance descriptor of the target model. This method is robust to the possible occlusion and large illumination change, as five covariance matrices in different regions are used to compute the dissimilarity.

To obtain a statistically meaningful conclusion, a variety of experiments are given to validate the superiority of our proposed target detection method that is the covariance descriptor with the MJBLD (CD-MJBLD) divergence on the Inria person [38], the Fashion-MNIST [39] and the Pascal VOC [40] datasets. Detection performances are compared with the covariance descriptor with the JBLD (CD-JBLD) divergence, the AIRM (CD-AIRM) [41] and the histogram of oriented gradients (HOG) method [42], which is one of the most successful target descriptors. The HOG descriptor was created to allow the human form in images to be discriminated

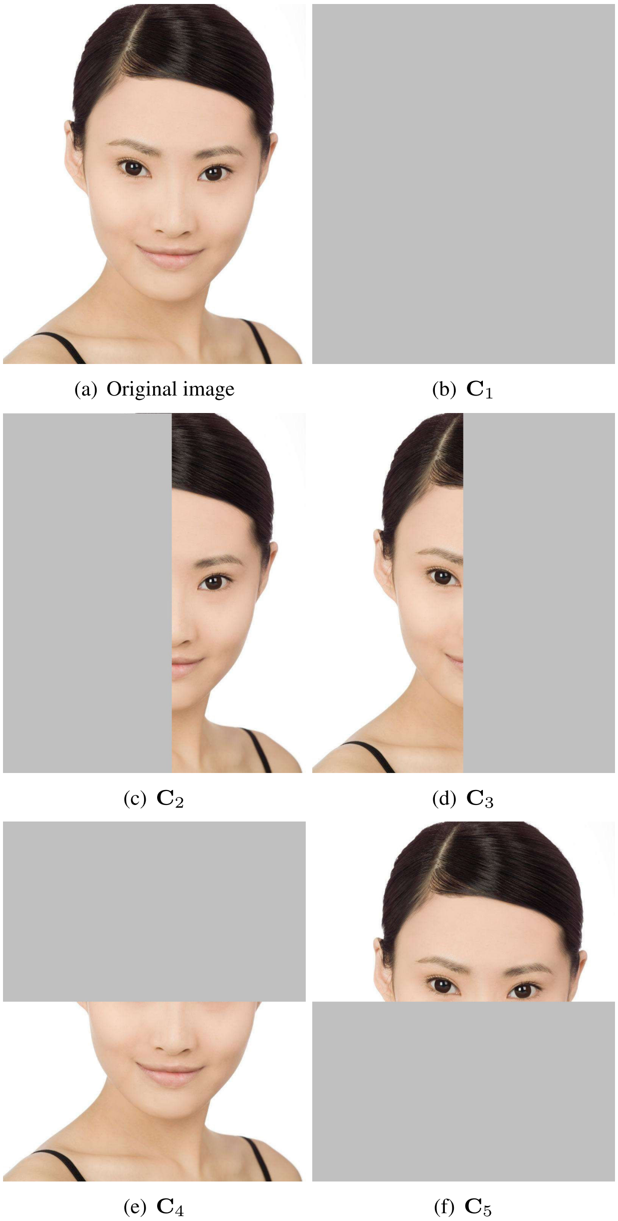
(a) Original image
(b) $\mathbf{C}_1$
(c) $\mathbf{C}_2$
(d) $\mathbf{C}_3$
(e) $\mathbf{C}_4$
(f) $\mathbf{C}_5$

FIGURE 1. Target representation. Five covariance matrices are derived from overlapping regions of the feature image. (a) Original image. (b) Covariance matrix from the whole region. (c) Covariance matrix from the left region. (d) Covariance matrix from the right region. (e) Covariance matrix from the top region. (f) Covariance matrix from the bottom region.

FIGURE 2. Sample images from the inria person dataset.

clearly at first then applied to other problem domains as well. The three datasets are presented as follows:

Inria Person: The Inria person dataset is collected from video and images with different postures of people. This dataset was divided in two versions: (a) positive images in normalized format, and (b) original images with corresponding annotation files. The obtained positive images have been cropped to highlight persons by the high resolution images. The people in these images are bystanders taken from the image backgrounds, and then there is no particular bias in their pose. See Figure 2 for sample images.

Fashion-MNIST: The Fashion-MNIST dataset consists of 7000 products with 10 categories. Each image is a 28 × 28 grayscale image. These products have several different groups, including women, men, neutral and kids. Particularly, we have randomly selected 6000 images from each class. Labels and images are contained in the same file format, and the file stores the vector-form and the matrix-form data. This dataset has been used for benchmarking many machine learning tasks.

TABLE 2. Perfromance comparison (%) on inria person dataset.

Method	CD-MJBLD	CD-JBLD	CD-AIRM	HOG
Average accuracy	96.23	95.71	90.63	85.44
Recall	87.13	86.69	84.78	83.94

Pascal VOC: The PASCAL VOC 2007 dataset is collected from everyday scenes with 20 classes, including bicycle, boat, bus, cat, dog, person, train, dining table, sheep, motorbike, potted plant, bird, chair, horse, train, aeroplane, bottle, car, cow, sofa, and tv monitor. We select 5000 images with 12000 annotated instances for test.

Detection results on these three datasets are provided in Table 2, Table 3, and Table 4. It can be noted from Table 2, 3, and 4 that the CD-MJBLD has the best detection performance on these datasets. Both CD-JBLD and CD-AIRM

TABLE 3. Perfromance comparison (%) on fashion-MNIST dataset.

Method	CD-MJBLD	CD-JBLD	CD-AIRM	HOG
Average accuracy	95.53	92.28	89.52	83.81
Recall	86.92	84.87	84.07	82.93

TABLE 4. Average accuracy (%) on pascal VOC dataset.

Method	CD-MJBLD	CD-JBLD	CD-AIRM	HOG
Average precision	98.07	97.16	95.01	90.90
Recall	90.17	89.36	87.55	86.74

outperform the HOG method. Region covariance descriptor can characterize the region of interest accurately while many regions found by the HOG descriptor are mismatched. Even among the correctly detected regions with these three methods, it is clear that the covariance descriptor can better localize the target region. The detection task is challenging as there are nonrigid motion, illumination changes, and large scale. The result also indicates the robustness of the proposed approach. It can be concluded that the region covariance descriptor is very effective and discriminative. The average accuracy for CD-MJBLD has a more than 3% $\sim$ 6% performance improvement with respect to the CD-AIRM method.

VI. CONCLUSION

In this paper, a region covariance descriptor-based target detection method has been proposed via the MJBLD divergence. In particular, the MJBLD divergence has been employed as the dissimilarity measurement between two region covariance descriptors. We have analyzed the properties of the MJBLD divergence with the detailed mathematical foundation. The MJBLD divergence is a modified version of the JBLD divergence which has been proven to be a proxy of AIRM metric since many intrinsic attractive properties such as the affine invariance. In the assessment stage, a variety of experiments have been provided to verify the superiority of our proposed target detection method in comparison with the region covariance descriptor with the JBLD divergence, the AIRM metric and the HOG descriptor. The results have shown that the region covariance descriptor with the MJBLD divergence has better detection performance than the region covariance descriptor with the JBLD divergence and the AIRM metric, followed by the HOG descriptor.

REFERENCES

[1] A. Kanezaki, T. Suzuki, T. Harada, and Y. Kuniyoshi, "Fast object detection for robots in a cluttered indoor environment using integral 3D feature table," in *Proc. IEEE Int. Conf. Robot. Autom.*, May 2011, pp. 4026–4033, doi: 10.1109/ICRA.2011.5980129.

[2] R. C. Luo and C. C. Lai, "Multisensor fusion-based concurrent environment mapping and moving object detection for intelligent service robotics," *IEEE Trans. Ind. Electron.*, vol. 61, no. 8, pp. 4043–4051, Aug. 2014, doi: 10.1109/TIE.2013.2288199.

[3] A. Coates and A. Y. Ng, "Multi-camera object detection for robotics," in *Proc. IEEE Int. Conf. Robot. Autom.*, May 2010, pp. 412–419, doi: 10.1109/ROBOT.2010.5509644.

[4] W. Zhang, F. Chen, W. Xu, and E. Zhang, "Real-time video intelligent surveillance system," in *Proc. IEEE Int. Conf. Multimedia Expo.*, Jul. 2006, pp. 1021–1024, doi: 10.1109/ICME.2006.262707.

[5] S. Hamidreza Kasaei, M. Oliveira, G. H. Lim, L. Seabra Lopes, and A. M. Tomé, "Towards lifelong assistive robotics: A tight coupling between object perception and manipulation," *Neurocomputing*, vol. 291, pp. 151–166, May 2018.

[6] X. Hua, L. Peng, W. Liu, Y. Cheng, H. Wang, H. Sun, and Z. Wang, "LDA-MIG detectors for maritime targets in nonhomogeneous sea clutter," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5101815, doi: 10.1109/TGRS.2023.3250990.

[7] X. Hua, Y. Ono, L. Peng, and Y. Xu, "Unsupervised learning discriminative MIG detectors in nonhomogeneous clutter," *IEEE Trans. Commun.*, vol. 70, no. 6, pp. 4107–4120, Jun. 2022.

[8] X. Hua, Y. Ono, L. Peng, Y. Cheng, and H. Wang, "Target detection within nonhomogeneous clutter via total Bregman divergence-based matrix information geometry detectors," *IEEE Trans. Signal Process.*, vol. 69, pp. 4326–4340, 2021.

[9] E. Ohn-Bar and M. M. Trivedi, "Fast and robust object detection using visual subcategories," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2014, pp. 179–184, doi: 10.1109/CVPRW.2014.32.

[10] A. M. Martinez and S. Du, "A model of the perception of facial expressions of emotion by humans: Research overview and perspectives," in *Gesture Recognition*. Cham, Switzerland: Springer, 2017, pp. 183–202.

[11] R. Brunelli and T. Poggio, "Face recognition: Features versus templates," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 15, no. 10, pp. 1042–1052, Oct. 1993, doi: 10.1109/34.254061.

[12] A. Kumar, A. Kaur, and M. Kumar, "Face detection techniques: A review," *Artif. Intell. Rev.*, vol. 52, no. 2, pp. 927–948, Aug. 2019, doi: 10.1007/s10462-018-9650-2.

[13] V. P. Kshirsagar, M. R. Baviskar, and M. E. Gaikwad, "Face recognition using eigenfaces," in *Proc. 3rd Int. Conf. Comput. Res. Develop.*, vol. 2, Mar. 2011, pp. 302–306, doi: 10.1109/ICCRD.2011.5764137.

[14] B. Zhang, Y. Yang, C. Chen, L. Yang, J. Han, and L. Shao, "Action recognition using 3D histograms of texture and a multi-class boosting classifier," *IEEE Trans. Image Process.*, vol. 26, no. 10, pp. 4648–4660, Oct. 2017, doi: 10.1109/TIP.2017.2718189.

[15] F. Porikli, "Integral histogram: A fast way to extract histograms in Cartesian spaces," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Nov. 2005, pp. 829–836, doi: 10.1109/CVPR.2005.188.

[16] P. Bilinski, M. Koperski, S. Bak, and F. Bremond, "Representing visual appearance by video Brownian covariance descriptor for human action recognition," in *Proc. 11th IEEE Int. Conf. Adv. Video Signal Based Surveill. (AVSS)*, Aug. 2014, pp. 87–92, doi: 10.1109/AVSS.2014.6918649.

[17] F. Palmieri and U. Fiore, "A nonlinear, recurrence-based approach to traffic classification," *Comput. Netw.*, vol. 53, no. 6, pp. 761–773, Apr. 2009.

[18] O. Déniz, G. Bueno, J. Salido, and F. De la Torre, "Face recognition using histograms of oriented gradients," *Pattern Recognit. Lett.*, vol. 32, no. 12, pp. 1598–1603, Sep. 2011, doi: 10.1016/j.patrec.2011.01.004.

[19] J. Arróspide, L. Salgado, and M. Camplani, "Image-based on-road vehicle detection using cost-effective histograms of oriented gradients," *J. Vis. Commun. Image Represent.*, vol. 24, no. 7, pp. 1182–1190, Oct. 2013, doi: 10.1016/j.jvcir.2013.08.001.

[20] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Oct. 2005, pp. 886–893, doi: 10.1109/CVPR.2005.177.

[21] V.-D. Hoang, M.-H. Le, and K.-H. Jo, "Hybrid cascade boosting machine using variant scale blocks based HOG features for pedestrian detection," *Neurocomputing*, vol. 135, pp. 357–366, Jul. 2014, doi: 10.1016/j.neucom.2013.12.017.

[22] M. Khalid, M. M. Yousaf, K. Murtaza, and S. M. Sarwar, "Image defencing using histograms of oriented gradients," *Signal, Image Video Process.*, vol. 12, no. 6, pp. 1173–1180, Sep. 2018.

[23] X. Wang, T. X. Han, and S. Yan, "An HOG-LBP human detector with partial occlusion handling," in *Proc. IEEE 12th Int. Conf. Comput. Vis.*, Sep. 2009, pp. 32–39, doi: 10.1109/ICCV.2009.5459207.

[24] M. T. Harandi, C. Sanderson, R. Hartley, and B. C. Lovell, "Sparse coding and dictionary learning for symmetric positive definite matrices: A kernel approach," in *Proc. ECCV*, 2012, pp. 216–229.

[25] S. Jayasumana, R. Hartley, M. Salzmann, H. Li, and M. Harandi, "Kernel methods on the Riemannian manifold of symmetric positive definite matrices," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 73–80, doi: 10.1109/CVPR.2013.17.

[26] E. Zhang, W. Chen, Z. Zhang, and Y. Zhang, "Local surface geometric feature for 3D human action recognition," *Neurocomputing*, vol. 208, pp. 281–289, Oct. 2016, doi: 10.1016/j.neucom.2015.12.122.

[27] I. Kviatkovsky, E. Rivlin, and I. Shimshoni, "Online action recognition using covariance of shape and motion," *Comput. Vis. Image Understand.*, vol. 129, pp. 15–26, Dec. 2014, doi: 10.1016/j.cviu.2014.08.001.

[28] D. Mulfari, A. Longo Minnolo, and A. Puliafito, "Building TensorFlow applications in smart city scenarios," in *Proc. IEEE Int. Conf. Smart Comput.*, May 2017, pp. 1–5, doi: 10.1109/SMARTCOMP.2017.7946991.

[29] V. Eiselein, G. Sternharz, T. Senst, I. Keller, and T. Sikora, "Person re-identification using region covariance in a multi-feature approach," in *Image Analysis and Recognition*. Springer, 2014, pp. 77–84.

[30] A. Brunetti, D. Buongiorno, G. F. Trotta, and V. Bevilacqua, "Computer vision and deep learning techniques for pedestrian detection and tracking: A survey," *Neurocomputing*, vol. 300, pp. 17–33, Jul. 2018, doi: 10.1016/j.neucom.2018.01.092.

[31] R. G. Sanfelice and L. Praly, "Convergence of nonlinear observers on BBR^n with a Riemannian metric (Part I)," *IEEE Trans. Autom. Control*, vol. 57, no. 7, pp. 1709–1722, Jul. 2012.

[32] K. C. Gurijala, R. Shi, W. Zeng, X. Gu, and A. Kaufman, "Colon flattening using heat diffusion Riemannian metric," *IEEE Trans. Vis. Comput. Graph.*, vol. 19, no. 12, pp. 2848–2857, Dec. 2013, doi: 10.1109/TVCG.2013.139.

[33] M. Moakher, "On the averaging of symmetric positive-definite tensors," *J. Elasticity*, vol. 82, no. 3, pp. 273–296, Mar. 2006, doi: 10.1007/s10659-005-9035-z.

[34] X. Pennec, P. Fillard, and N. Ayache, "A Riemannian framework for tensor computing," *Int. J. Comput. Vis.*, vol. 66, no. 1, pp. 41–66, Jan. 2006, doi: 10.1007/s11263-005-3222-z.

[35] A. Fischer, "Quantization and clustering with Bregman divergences," *J. Multivariate Anal.*, vol. 101, no. 9, pp. 2207–2221, Oct. 2010, doi: 10.1016/j.jmva.2010.05.008.

[36] F. Nielsen and R. Nock, "Jensen-bregman Voronoi diagrams and centroidal tessellations," in *Proc. Int. Symp. Voronoi Diagrams Sci. Eng.*, Jun. 2010, pp. 56–65, doi: 10.1109/ISVD.2010.17.

[37] A. Cherian, S. Sra, A. Banerjee, and N. Papanikolopoulos, "Efficient similarity search for covariance matrices via the Jensen-Bregman LogDet divergence," in *Proc. Int. Conf. Comput. Vis.*, Nov. 2011, pp. 2399–2406, doi: 10.1109/ICCV.2011.6126523.

[38] J. Marín, D. Vázquez, A. M. López, J. Amores, and L. I. Kuncheva, "Occlusion handling via random subspace classifiers for human detection," *IEEE Trans. Cybern.*, vol. 44, no. 3, pp. 342–354, Mar. 2014, doi: 10.1109/TCYB.2013.2255271.

[39] H. Xiao, K. Rasul, and R. Vollgraf, "Fashion-MNIST: A novel image dataset for benchmarking machine learning algorithms," 2017, *arXiv:1708.07747*.

[40] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The Pascal visual object classes (VOC) challenge," *Int. J. Comput. Vis.*, vol. 88, no. 2, pp. 303–338, Jun. 2010.

[41] O. Tuzel, F. Porikli, and P. Meer, "Region covariance: A fast descriptor for detection and classification," in *Proc. Eur. Conf. Comput. Vis.*, 2006, pp. 589–600.

[42] D. Sangeetha and P. Deepa, "Efficient scale invariant human detection using histogram of oriented gradients for IoT services," in *Proc. 30th Int. Conf. VLSI Design 16th Int. Conf. Embedded Syst. (VLSID)*, Jan. 2017, pp. 61–66, doi: 10.1109/VLSID.2017.60.

XIQIAN FAN received the B.S. degree in telecommunication engineering from Harbin University of Science and Technology, in 2021. He is currently pursuing the M.S. degree with Hangzhou Institute for Advanced Research, Chinese Academy of Sciences, Hangzhou, China. His research interests include digital signal processing, computer vision, and target detection.

SHAOZHU YE received the B.S. degree in measurement and control technology and instrument from China Jiliang University, Hangzhou, China, in 2011. His research interests include the development of image processing and visual inspection techniques.