This article has been accepted for publication in IEEE Transactions on Information Forensics and Security. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIFS.2023.3333681

Similarity Distance Learning on SPD Manifold for Writer Independent Offline Signature Verification

Elias N. Zois, Dimitrios Tsourounis, and Dimitrios Kalivas

Abstract— **Identifying the existence or approval of a human in a number of past, recent and present day activities with the use of a handwritten signature is a captivating biometric challenge. Several engineering branches such as computer vision, pattern recognition and quite recently data-driven machine learning algorithms are combined in a multi-disciplined signature verification framework in order to deliver an equivalent and efficient e-assistance to manually executed duties, which usually demand knowledge and skills. In this work, we propose, for the first time, the use of a learnable Symmetric Positive Definite manifold distance framework in offline signature verification literature in order to build a global writer-independent signature verification classifier. The key building block of the framework relies on the use of regional covariance matrices of handwritten signature images as visual descriptors, which maps them into the Symmetric Positive Definite manifold. The learning and verification protocol explores both blind intra and blind inter transfer learning frameworks with the use of four popular signature datasets of Western and Asian origin. Experiments strongly indicate that the learnable SPD manifold similarity distance can be highly efficient for offline writer independent signature verification.**

Index Terms— **Manifold Optimization, Symmetric Positive Definite Matrices, Spatial pyramid segmentation, Writer Independent Off-line Signature Verification.**

I. INTRODUCTION

Perhaps the most widespread handwritten attribute used, in order to express our endorsement or manifestation, is the handwritten signature. The authentication of the handwritten signature by means of computer engineering is a captivating e-society challenge [1], [2] with numerous testimonials to emphasize its use and importance. Automated signature verification (SV or ASV) is the engineering branch that provides technical advances and research areas in the fields of: the examination of Forensic Handwriting Documents [3], health, cybersecurity of private, public and/or government acts (e.g. verifying ballots in elections) [4], [5]. It is also an applied field for commercial and business solutions toward an number of applications ranging from essential economic transactions to mail ballots and therapeutic actions approval [6], [7].

The formation of the signature silhouette combines the learned scripting customs as well as the individual and specific brain motoric process [8]-[11] of a person. In case that the signature is drawn on a sheet of paper, its static counterpart is acquired with the use of a scanner while in the case of using an electronic device one could acquire, besides a digital image, a time indexed multivariable sequence. Earlier [12] as well as newly published research papers and reviews in SV [13], [14] initially categorize the SV methods either as dynamic-online (signal vs. time) [1], [11], [15]-[20] or as static-offline (image) [21]-[25]. Another categorization of SV methods classifies them into Writer Dependent (WD) or Independent (WI) according to the verification strategy followed [26]-[32]. The WD approach is the most frequently encountered in the literature in which a dedicated classifier is trained for each signatory with his/her reference samples [24], [33]-[36].

A. Writer independent signature verification

In a more challenging approach, the WI-SV protocols learn a universal classifier in order to discriminate between two types of distributions: a) the positive (ω^+), expressed by the genuine-to-genuine (similar) pairs of a learning set of signatures and b) the negative (ω^-), expressed by the genuine-to-forgeries (dissimilar) pairs [27], [28], [37]-[39]. This is usually attained by employing the dichotomy transformation [40], in which the feature space $F \in \mathbb{R}^K$, which contains any two signature pairs (F_i, F_j), is transformed to a distance space $\left|F_i - F_j\right| \in \mathbb{R}_+^K$ denoted hereafter as the (dis)similarity distance space. Recently, WI verifiers have been proposed with a number of Deep learning methods like attention Siamese networks [41], inverse discriminative networks [42], static-dynamic interaction networks [43], self-supervised attention-guided reconstruction [44] and capsule neural networks [45] that do not necessary follow the dichotomy transform.

Metric learning appears to be an attractive way for WI-SV research. In the literature one may find additional research efforts pointing mainly to deep metric learning. For example, Rantzsch et al. [46] proposed a comparison between triplets of two genuine and one forged signature, in order to embed signatures in a high-dimensional space, with Euclidean distance acting as the similarity measure. Soleimani et al. [38] combined the use of Histograms of Oriented Gradients (HOG), as signature descriptors, and a Deep Multitask Metric Learning (DMML) approach in order to learn a set of hierarchical nonlinear transformations with a deep neural network. Maergner et al. [47] proposed the combined usage of

Elias N. Zois & Dimitrios Kalivas are with the Telsip research laboratory, University of West Attica, Aigaleo, 12241, Greece. (e-mail: ezois@uniwa.gr dikal@uniwa.gr).

Dimitrios Tsourounis is with the Electronics Laboratory, University of Patras, Patras, 26504, Greece (e-mail: dtsourounis@upatras.gr).

Supplementary material includes additional matlab figures.

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

This article has been accepted for publication in IEEE Transactions on Information Forensics and Security. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIFS.2023.3333681

a structural approach based on graph edit distance with a statistical approach based on deep triplet networks in order to address a keypoint graph-based dissimilarity computation. They also proposed the use of a graph edit distance in order to learn a convolutional neural network using a triplet loss function. Zhu et al. [48] proposed a point-to-set similarity based deep feature learning by dividing a training batch into a support set and a query set. Lai et al. [49] proposed an oriented feature extractor by combining a classification loss and a metric learning loss on a WD basis. Chattopadhyay et al. [44] proposed also a two-stage deep learning framework that leverages both self-supervised representation and metric learning. Liu et al. [50] proposed a Mutual Signature DenseNet (MSDN) to extract features and learn the similarity measure from local regions instead of whole signature images. Lin et al. [51] proposed a 2-Channel-2-Logit network whose output measures the dissimilarity between reference and query signatures and avoid overfitting. Hanif et al., [52] used a Mahalanobis metric learning approach on HOG and LBP descriptors computed at interest points. Ji et al. [53], proposed a paired contrastive transformation (PCF) of similar and dissimilar signature pairs with rejection and top-rank learning for highly reliable signature verification.

B. Concerns regarding WI systems

An issue that one should acknowledge at the WI-SV design and learning (i.e. training & validation) stages stems from the fact that the (ω^-) pairs can be of different types. For example, a dissimilar pair can be formed by pairing a genuine sample of a person with a genuine sample of another person (i.e. *Random forgeries)*, or ii) a genuine sample of a person with a simulated-or-skilled sample of the same person. For unbiased operation of any WI verifier, neither intra and inter class pairs that learn and validate the dissimilarity space should be independent, or blind, that is they must not be a part of the claimed identity. This type of independency allow us to utilize the ω^- class of genuine-to-forgeries pairs with diverse quality. Thus, the ω^- class can be formed by ratios of a) genuine-to-random or b) genuine-to-skilled-simulated forgeries.

Ideally, a SV system should be able to effectively cope with the fundamental question [54]: “Given a group of reference signatures, does a questioned signature belong to them?” Broadly, a low dimensional, efficient as well as descriptive visual representation of images was introduced in [55], with the utilization of the region covariance descriptors of image feature stacks (or maps). This representation introduced the principles of non-Euclidean geometry in computer vision algorithms with a notable representative among others, the symmetric positive definite (SPD) manifold space with numerous applications like fine-grained image classification [56],[57], generic/imageNet image classification [58], [59], action/video classification [60], [61], person re-identification [62]-[64] domain adaptation [65], few-shot learning [66], meteorology [67], medical imaging [68], brain-computer interface analysis [69], [70], etc.

C. The proposed symmetric positive definite WI-SV system

For the first time in the offline SV literature, a WD-SV SPD mapping of a signature with a corresponding region covariance matrix [35] was proposed recently. Therefore, we felt that it is reasonable to ask ourselves in what way we can explore a common ground between the region covariance/SPD matrices and the domain of WI-SV by employing an appropriate metric learning for establishing a similarity measure. We commence with the following facts: a) similarity-based algorithms are agnostic to the geometry of the feature space and mainly reside in the idea that they need only (dis)similarities of any (ω^+, ω^-) pairs [57], b) SPD geometry does not have a Euclidean nature which often makes the design of SPD based classifiers quite challenging. Popular SPD based classification methods have to convert an SPD point into a Euclidean-style feature vector by means of tangent approximations [35], [71], [72], the kernel and/or the coding methods [73]-[75]. These methods, however, perturb the intrinsic matrix properties and might provide mediocre results, and c) manifold geometry is closely related to the notion of distance measure. Similarity between two SPD points can be measured with several mathematical entities like a) the affine-invariant metric (AIM) [68], b) the log-Euclidean metric (LEM) [76], c) the Stein divergence [77], d) the Burg matrix divergence [78] and e) the alpha-beta divergence [79], [80]. Metric learning in the SPD manifold has also been proposed in order to model heterogeneous applications in which a distance measure from the data under examination [81] is transferred to the application domain. It has been reported [82] that there are three categories of SPD manifold metric learning methods. The first category comprises methods that learn a distance metric in the Euclidean tangent space. The second category consist of methods that learn the distance metric in the kernel space, usually in reproducing kernel Hilbert space (RKHS) [83] which generalizes the LEM between two SPD points and infinite-dimensional covariance matrices by Hilbert-Schmidt operators. The third category proposes to preserve the global SPD structure [81] by: a) projecting an initial high-dimensional SPD point into another SPD one, typically with fewer dimensions and b) learn a corresponding metric. A local extension of the third category has been proposed in [57] in which a projected SPD point is partitioned in combinations of discovered visual information represented by local SPD matrices.

A literature search reveals that, to the author’s best knowledge no prior work has been presented which models the handwritten signature with the use of SPD matrices and corresponding metric learning for WI-SV. The literature research provided in section I.A always imply an underlying Euclidian nature. This is an important issue because, frequently, machine learning algorithms assume that there is an underlying Euclidean nature of the input signature descriptor. But, up to now, no one has assumed that they are immersed into a non-Euclidean vector space. This geometrical constraint regarding the intrinsic structure of the signature descriptor may provide suboptimal results regarding the

This article has been accepted for publication in IEEE Transactions on Information Forensics and Security. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIFS.2023.3333681

verification efficiency. Our work contemplates the non-Euclidean nature of the signature descriptors and introduce a SPD metric learning framework. The novelty characteristics of the proposed approach are:

1. We address the WI-SV problem as a learnable SPD distance problem in which pairs of similar/dissimilar signatures are employed in its learning stage for the creation of a similarity distance between input pairs [84] rather than the original input space. Contrary to the design of a Euclidean learning model, the use of a non-Euclidean manifold makes the task of classification complicated as well as demanding. This is due to the fact that the SPD manifold is not a Euclidean vector space.

2. We explore, for the first time in the for WI-SV literature, the SPD manifold geometry with a theoretical framework presented in [57]. The key idea is to map a set of initial SPD input points into another SPD manifold so that similar points in the original space are mapped together on the new manifold while dissimilar point are mapped apart [85]. For this purpose, (ω^+, ω^-) SPD pairs are employed in order to learn a set of three parameters $\Theta = \{\boldsymbol{W}, \boldsymbol{A}, \mathbf{M}\}$ which represent a joint learning algorithm with a point-to-set transformation $\boldsymbol{W}$, a set-to-set distance measure $\boldsymbol{A}$ and a merging factor $\mathbf{M}$.

The rest of the paper is organized as follows: Section II present the preliminaries of the proposed system architecture and introduces the necessary steps for the creation of the signature global covariance matrix. Section III reviews the elements and mathematical tools of the SPD Riemannian manifold. Section IV provides details regarding the proposed SPD distance learning method. Section V describes the experimental methods and provides the results. Finally, section VI provides the conclusion.

II. Proposed System & Signature Covariance Matrix

Figure 1 illustrates conceptually a toy example of the proposed learning procedure. Section II.A describes the way that, handwritten signature images are converted to $\mathbb{R}^{n\times n}$ covariance matrices. Now, let us denote the outcome of the SPD metric learning model with the similarity distance $\Delta(\boldsymbol{W}, \boldsymbol{A}, \mathbf{M})$. This has the following learning parameters: a) a projection matrix $\boldsymbol{W}$, which maps the initial covariance matrices to another covariance $\in \mathbb{R}^{mp\times mp}$ with $mp \leq n$ and at the same time selects m-numbered non-overlapping block-diagonal covariance matrices $\mathbb{R}^{p\times p}$, b) the $\boldsymbol{A} \in \mathbb{R}^{m\times 2}$ set of learnable parameters which characterize each of the m-numbered alpha-beta divergences $D_{i=1,\dots,m}$, and c) the matrix $\mathbf{M} \in \mathbb{R}^{m\times m}$ which weights and sum up any alpha-beta divergences. Following the completion of the learning stage, the testing stage uses the learned model with i) one questioned sample Q and ii) a reference population $\mathcal{G}$ of $\mathcal{G}_{NREF}$ samples, $\Delta_Q^{\mathcal{G}}(\boldsymbol{W}, \boldsymbol{A}, \mathbf{M})$. All these procedures will be elaborated further.

A. Signature covariance matrix

Let $I \in \mathbb{R}^{w\times h}$ be a digital image of w-columns and h-rows and $F \in \mathbb{R}^{w\times h\times n}$ be a corresponding image stack with n image planes, evaluated from I with the use of a number of n-filters: $F(x,y,i) = \Phi_i(I,x,y), i = 1{:}n$. The function Φ can be any type of a stack of mapping functions such as intensity, gradients, pixel locations, filter mappings, etc. Given a rectangular image region $\mathcal{R} \subset F$, let $\boldsymbol{f} = [\boldsymbol{f}_i]_{i=1,2,\dots S} \in \mathbb{R}^{n\times S}$ be a local feature map of S total pixels that reside in $\mathcal{R}$. Then, the region $\mathcal{R}$ is modelled by its region covariance matrix $\boldsymbol{C}_{\mathcal{R}} \in \mathbb{R}^{n\times n}$ of the $\boldsymbol{f}_i \in \mathbb{R}^n$ points which is evaluated as:

$$\boldsymbol{C}_{\mathcal{R}} = \frac{1}{S-1}\sum_{i=1}^{S}(\boldsymbol{f}_i - \boldsymbol{\mu})(\boldsymbol{f}_i - \boldsymbol{\mu})^T \tag{1}$$

where $\boldsymbol{\mu} \in \mathbb{R}^n$ represents the column mean vector of the $\boldsymbol{f}_i$ points and T denotes the transpose operator. For every signature image a sequence of typical image processing steps which involves: thresholding with Otsu's method and thinning. The pruning level of thinning utilizes an automated algorithm originally proposed in [22]. For the addressed

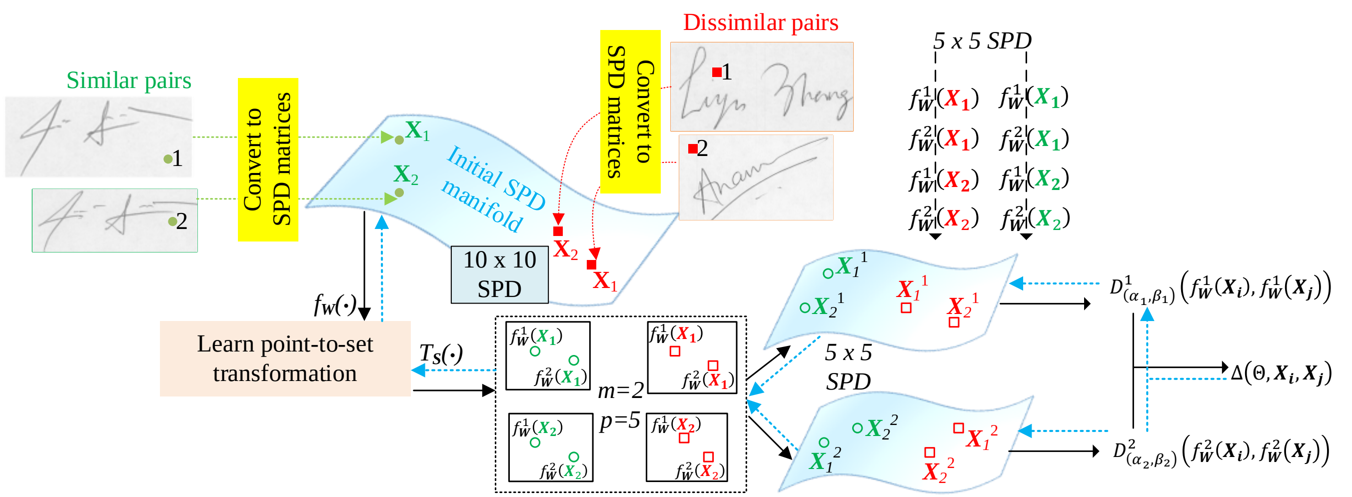

Fig. 1. Toy example of the proposed SPD metric learning framework. Similar (green dots) and dissimilar (red squares) pairs of signature images are converted into SPD matrices of 10×10 size. A joint optimization procedure (with parameter $\Theta = \mathbf{W}, \mathbf{A}, \mathbf{M}$), is represented by forward black and backpropagation blue arrows. i) Map the initial SPD space into another SPD space with the projection matrix $\mathbf{W}$ and at the same time select two (m=2, p=5) $p\times p$ block diagonal SPD matrices (green circles and red boxes), ii) For each new SPD space, the $\mathbf{A} = (\alpha_k, \beta_k)_{k=1}^{2}$ parameters of the local distances $D_{(\alpha_k,\beta_k)}^{k}$ are learned and c) learn a weight matrix $\mathbf{M}$ which merges the $D_{\mathbf{A}}^{k}$ into one score $\Delta(\Theta, \boldsymbol{X}_i, \boldsymbol{X}_j)$.

This article has been accepted for publication in IEEE Transactions on Information Forensics and Security. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIFS.2023.3333681

offline SV problem we define the following mapping $\Phi_i(I,x,y)$ of a raw signature image $I_{raw}(x,y)$ as:

$$\left[I, I_x, I_y, I_{xx}, I_{xy}, I_{yy}, \sqrt{{I_x}^2 + {I_y}^2}, \tan^{-1}\left(I_y/I_x\right), x_n, y_n,\right] \quad (2)$$

in which, I is the grayscale image after the preprocessing step, $I_x, I_y, I_{xx}, I_{xy}, I_{yy}$ are image derivatives of $I(x,y)$, x_n, y_n are the signature pixel coordinates, normalized by their maximum number of rows and columns of the image bounding box and $\tan^{-1}\left(I_y/I_x\right)$ is the gradient direction, normalized into radians with range varying from [-π, π). The corresponding signature covariance matrix $\boldsymbol{C}_{SCM}$ of I is evaluated with eqs. (1)-(2) but under the constraint that only the pixels that are part of the signature trace of the preprocessing step contribute to the computation of $\boldsymbol{C}_{SCM}$. Therefore, any signature image results in a $\boldsymbol{C}_{SCM} \in \mathbb{R}^{10\times 10}$ point of the corresponding SPD manifold. The covariance matrix can be evaluated for different parts of the image; therefore for each signature image we evaluate its corresponding $\boldsymbol{C}_{SCM}$ not only for the global image $\boldsymbol{C}_{SCM}^{1\times 1}$, but also for a spatial pyramid segmentation scenario which provides also four $\boldsymbol{C}_{SCM}^{2\times 2}$, and nine $\boldsymbol{C}_{SCM}^{3\times 3}$ covariance matrices of the equi-mass sub-regions [22].

III. The Symmetric Positive Definite Manifold

Unless otherwise specified, from now on, vectors are labeled by bold lowercase letters (e.g. $\mathbf{x}$) while bold capital letters (e.g. $\mathbf{X}$) denote matrices. Also, $\boldsymbol{I}_n$ is the $n \times n$ identity matrix, $GL(n)$ denotes the space of the real invertible $n \times n$ matrices, and $Sym(n)$ is the space of real $n \times n$ symmetric matrices. The S_{++}^n denotes the SPD manifold, properly defined in the following paragraphs. Finally, $diag(\lambda_1, \lambda_2, \ldots \lambda_n)$ is a diagonal matrix whose elements are the real values $\left\{\lambda_\mu\right\}_{\mu=1}^{n}$.

A. The Riemannian Manifold of SPD Matrices

According to the terminology found in [35] and some seminal references [57], [81], [86], let us define a topological manifold (or manifold) as a topological space that is locally homomorphic to the *n*-dimensional Euclidean space $\mathbb{R}^n$. A differentiable manifold is a manifold equipped with a globally defined differential structure which allows to define the derivatives of curves on the manifold. The derivatives at any point $\mathbf{X}$ on the manifold lie in the tangent space $T_{\boldsymbol{X}}$ of $\mathbf{X}$, which is a vector space expressed by symmetric matrices. A Riemannian manifold $\mathcal{M}$ [68] is a differentiable manifold equipped with a smoothly varying inner product $\langle,\rangle_{\boldsymbol{X}\in\mathcal{M}}$. This is also defined as the Riemannian metric (or norm) of a tangent vector $\boldsymbol{Y} \in T_{\boldsymbol{X}}$ such that $\|\boldsymbol{Y}\|_{\boldsymbol{X}}^2 = \langle \boldsymbol{Y}, \boldsymbol{Y}\rangle_{\boldsymbol{X}\in\mathcal{M}}$. Given a point $\boldsymbol{X} \in \mathcal{M}$ and a tangent vector $\boldsymbol{Y} \in T_{\boldsymbol{X}}$ a unique geodesic curve $\boldsymbol{\Gamma}(t)$ exists with $\boldsymbol{\Gamma}(0) = \boldsymbol{X}$ and $\dot{\boldsymbol{\Gamma}}(t) = \boldsymbol{Y}$. In addition, the Riemannian exponential map $exp_{\boldsymbol{X}\in\mathcal{M}}: T_{\boldsymbol{X}} \to \mathcal{M}$ is defined by $exp_{\boldsymbol{X}\in\mathcal{M}}(\boldsymbol{Y}) = \boldsymbol{\Gamma}(1)$. In this work, we consider only Riemannian manifolds which a) verify the identity $d\big(\boldsymbol{X} \in \mathcal{M}, exp_{\boldsymbol{X}\in\mathcal{M}}(\boldsymbol{Y})\big) = \|\boldsymbol{Y}\|_{\boldsymbol{X}\in\mathcal{M}}$ and b) have a well-defined logarithm map: $log_{\boldsymbol{X}\in\mathcal{M}} = exp_{\boldsymbol{X}}{}^{-1}: \mathcal{M} \to T_{\boldsymbol{X}}$. Then, the following relation $d(\boldsymbol{X} \in \mathcal{M}, \boldsymbol{Y} \in \mathcal{M}) = \|log_{\boldsymbol{X}}(\boldsymbol{Y})\|_{\boldsymbol{X}}$ holds, which ensures both the existence and a closed-form expression for any the distance between the manifold points $\boldsymbol{X}, \boldsymbol{Y} \in \mathcal{M}$.

The SPD manifold $P_n \equiv S_{++}^n$ is the space of all $n \times n$ real matrices $\mathbf{Z}$ which are symmetric $\boldsymbol{Z}^{\boldsymbol{T}} - \boldsymbol{Z} = \boldsymbol{0}$ and strictly positive definite: $\boldsymbol{v}^{\boldsymbol{T}}\boldsymbol{Z}\boldsymbol{v} > \boldsymbol{0}, \forall \boldsymbol{v} \in \mathbb{R}^n - \boldsymbol{0}_n$ i.e. the $\mathbb{R}^n$ space with the exclusion of the zero origin. Intuitively, the points that belong to the SPD manifold lie in the interior of a convex cone in a $n(n+1)/2$-dimensional Euclidean space. Given an SPD matrix $\boldsymbol{X} \in P_n$, its associated matrix exponential and logarithm functions have the following mathematical expressions:

$$expm(\boldsymbol{X}) = \boldsymbol{U}diag\big(\exp(\lambda_\mu)\big)\boldsymbol{U}^T \quad (3)$$

$$logm(\boldsymbol{X}) = \boldsymbol{U}diag\big(\log(\lambda_\mu)\big)\boldsymbol{U}^T \quad (4)$$

In the above equations, λ_μ is the μ-th eigenvalue derived from the eigenvalue analysis of $\boldsymbol{X} = \boldsymbol{U}diag\big(\lambda_\mu\big)\boldsymbol{U}^T$.

B. Geometric Perspective of SPD manifolds

The geometry of an SPD manifold is frequently formed with the use of the related Affine Invariant Riemannian Metric (AIRM) defined for $\boldsymbol{X} \in P_n$ and $\boldsymbol{Y}, \boldsymbol{W} \in T_{P_n}$ as [68]:

$$\begin{aligned}\langle \boldsymbol{Y}, \boldsymbol{W}\rangle_{\boldsymbol{X}} &\triangleq \langle \boldsymbol{X}^{-1/2}\boldsymbol{Y}\boldsymbol{X}^{-1/2}, \boldsymbol{X}^{-1/2}\boldsymbol{W}\boldsymbol{X}^{-1/2}\rangle \\ &= Tr(\boldsymbol{X}^{-1}\boldsymbol{Y}\boldsymbol{X}^{-1}\boldsymbol{W}) \end{aligned} \quad (3)$$

which induces the notion of a distance, formally termed geodesic distance, between the manifold points $\boldsymbol{X}, \boldsymbol{Z} \in P_n$ as:

$$\delta_R(\boldsymbol{X}, \boldsymbol{Z}) = \left\|\boldsymbol{X}^{-1/2}\boldsymbol{Z}\boldsymbol{X}^{-1/2}\right\|_F \quad (4)$$

Another, SPD matrix metric has been proposed in [76] formally termed as Log-Euclidean-Metric (LEM) which projects the $\boldsymbol{X}, \boldsymbol{Z} \in P_n$ points onto the common pole tangent space $T_{\boldsymbol{I}_{n\times n}}$ with:

$$D^{LEM}(\boldsymbol{X}, \boldsymbol{Z}) = \|logm(\boldsymbol{X}) - logm(\boldsymbol{Z})\|_F \quad (5)$$

Several other measures for matrix (dis)similarity have been also proposed, such as the Stein, Jeffrey Kullback-Leibler and Burg matrix divergences [57]. A proposed approach for measuring distances in the S_{++}^n is the learnable alpha-beta divergence [87] which was found to be adaptive under any underlying distribution. Therefore, for the $\boldsymbol{X}, \boldsymbol{Z} \in P_n$ the alpha-beta divergence is defined according to:

$$D^{(\alpha,\beta)}(\boldsymbol{X}\|\boldsymbol{Z}) = \frac{1}{\alpha\beta}\log\left(\det\left(\frac{\alpha(\boldsymbol{X}\boldsymbol{Z}^{-1})^{\beta} + \beta(\boldsymbol{X}\boldsymbol{Z}^{-1})^{-\alpha}}{\alpha+\beta}\right)\right)$$

$$\frac{1}{\alpha\beta}\sum_{\mu=1}^{n}\log\left(\frac{\alpha(\lambda_\mu)^{\beta} + \beta(\lambda_\mu)^{-\alpha}}{\alpha+\beta}\right) \quad (6)$$

with $\alpha \neq 0, \beta \neq 0$ and $\alpha + \beta \neq 0$. The pair (α, β) represents the learnable parameters of the divergence while $\det(\cdot)$ is the determinant and λ_μ is the μ-th eigenvalue of the product $\boldsymbol{X}\boldsymbol{Z}^{-\boldsymbol{1}}$. With respect to the (α, β) parameters, eq.(6) is considered to be smooth and several SPD distances can be expressed with specific (α, β) value assignments [79]. Furthermore, it has an invariance property under any affine transformation i.e. for any real and invertible matrix $\boldsymbol{B} \in GL(n)$, we have $D^{(\alpha,\beta)}(\boldsymbol{X}\|\boldsymbol{Z}) = D^{(\alpha,\beta)}(\boldsymbol{B}^T\boldsymbol{X}\boldsymbol{B}\|\boldsymbol{B}^T\boldsymbol{Z}\boldsymbol{B})$ [57].

IV. The Proposed SPD WI-SV Similarity Distance

This section provides the necessary theory as well the algorithmic steps for the proposed WI-SV (dis)similarity distance. As a prelude, we state that the proposed objective target is to create a modular SPD distance model $\Delta(\Theta)$

This article has been accepted for publication in IEEE Transactions on Information Forensics and Security. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIFS.2023.3333681

parametrized by the set of parameters $\Theta = \{\boldsymbol{W}, \boldsymbol{A}, \mathbf{M}\}$. Following a learnable mapping $\boldsymbol{W}$ from the original SPD manifold P_n to another SPD manifold $P_{m\cdot p}$, the model $\Delta(\Theta)$ explores, with the help of $\boldsymbol{A}$ and $\mathbf{M}$ parameters, diverse visual information which is stored in the $m-$individual block diagonal matrices $\in \mathbb{R}^{p\times p}$. We begin the analysis by getting familiarized with the mathematical formulation of the problem, initially introduced in [57]. To this end, let us declare a loss function $\mathcal{L}(\Theta, \mathrm{S}, \mathrm{D}, \mathrm{Y})$ in which $\mathrm{S}, \mathrm{D}, \mathrm{Y}$ are the set of similar and dissimilar signature pairs, along with their corresponding set labels y_{ij}=1 for $\boldsymbol{X_i}$ similar to $\boldsymbol{X_j}$ and y_{ij}=0 otherwise. Some useful mathematical notations follows also:

i. The projection $f_{\boldsymbol{W}}(\cdot)$: it maps any initial signature covariance matrix $\boldsymbol{C_{SCM}} \equiv \boldsymbol{X_i} \in P_n$ in a new, covariance matrix $\boldsymbol{Z_i} \in P_{m\cdot p}$, $m \cdot p \leq n$ by learning a orthogonal projection matrix $\boldsymbol{W}$.

ii. The point-to-set transformation $T_S(\cdot)$: it partitions the result of $f_{\boldsymbol{W}}(\cdot)$ in a set $\mathcal{X}_i$ of m, non-overlapping block diagonal SPD matrices $\{\boldsymbol{Z}_i^k\}_{k=1}^m \in P_p$. Both (i) and (ii) steps formulate in a notation $\{\boldsymbol{Z}_i^k\} = \{f_{\boldsymbol{W}}^k(\boldsymbol{X_i})\}_{k=1}^m$ which describes the derived set of m-numbered SPD matrices.

iii. The set of sub-distances $D_{\boldsymbol{A}}(\cdot,\cdot)$: it is a measure between corresponding pairs $\boldsymbol{Z}_i^k, \boldsymbol{Z}_j^k$ for each m-numbered SPD manifold returned by steps (i), (ii). The learnable parameter $\boldsymbol{A} \in \mathbb{R}^{m\times 2}$ comprises of the $(\alpha_k, \beta_k)_{k=1}^m$ parameters for each m-numbered SPD manifold. Fig. 1 again presents a toy example, in which the initial SPD signature covariance matrices $\boldsymbol{X_i}, \boldsymbol{X_j} \in P_n$, then, for the parameters m=2 and p=5, then we learn the set $(\alpha_k, \beta_k)_{k=1,2}$: $D^1_{(\alpha_1,\beta_1)}\left(f_{\boldsymbol{W}}^1(\boldsymbol{X_i}), f_{\boldsymbol{W}}^1(\boldsymbol{X_j})\right)$ and $D^2_{(\alpha_2,\beta_2)}\left(f_{\boldsymbol{W}}^2(\boldsymbol{X_i}), f_{\boldsymbol{W}}^2(\boldsymbol{X_j})\right)$.

iv. The integration function $H_{\boldsymbol{M}}(\cdot)$: it combines the m-numbered $\left\{D^k_{(\alpha_k,\beta_k)}\right\}_{k=1}^m$ sub-distances into one with the use of the learnable $\boldsymbol{M}$ parameter.

v. The set-to-set distance $D_{s2s}(\cdot,\cdot)$: computes the final distance between two SPD sets by combining the $D^k_{(\alpha_k,\beta_k)}$ with the use of the integration function $H_{\boldsymbol{M}}(\cdot)$.

To proceed, let us consider a pair of $\boldsymbol{X_i}$ and $\boldsymbol{X_j}$ SPD points. At first, we use steps (i), (ii), which results in a set of m-numbered low dimensional SPD sub-matrices $\{f_{\boldsymbol{W}}^k(\cdot)\}_{k=1}^m$ i.e. the $\boldsymbol{X_i}$, $\boldsymbol{X_j}$ points are assigned to their equivalent sets: $\mathcal{X}_i = T_S(\boldsymbol{X_i}) = \{f_{\boldsymbol{W}}^k(\boldsymbol{X_i})\}_{k=1}^m$and $\mathcal{X}_j = T_S(\boldsymbol{X_j}) = \{f_{\boldsymbol{W}}^k(\boldsymbol{X_j})\}_{k=1}^m$. Now, rather than using only one distance $\delta_R(\boldsymbol{X_i}, \boldsymbol{X_j})$ or $D^{LEM}(\boldsymbol{X_i}, \boldsymbol{X_j})$ between the original $\boldsymbol{X_i}$, $\boldsymbol{X_j}$ SPD points we propose to learn a family of m-numbered alpha-beta divergences $D_A^k\left(f_W^k(\boldsymbol{X_i}), f_W^k(\boldsymbol{X_j})\right)_{k=1}^m$. Next, the set-to-set distance $D_{s2s}(\mathcal{X}_i, \mathcal{X}_j) = D_{s2s}(\{f_W^k(\boldsymbol{X_i})\}_{k=1}^m, \{f_W^k(\boldsymbol{X_j})\}_{k=1}^m)$ will weight the m–individual instances of $\{D_{\boldsymbol{A}}^k(\cdot,\cdot)\}_{k=1}^m$ into one $\Delta(\Theta)$ with the use of the $H_{\boldsymbol{M}}(\cdot)$ integration function. The proposed WI-SV method is elaborated below by declaring the point-to-point distance $D^{\Theta}(\cdot,\cdot)$ on P_n as a set-to-set distance $D_{s2s}(\cdot,\cdot)$:

$$\Delta(\Theta, \boldsymbol{X_i}, \boldsymbol{X_j}) = D_{s2s}\left(T_S(\boldsymbol{X_i}), T_S(\boldsymbol{X_j})\right) = D_{s2s}(\mathcal{X}_i, \mathcal{X}_j)$$
$$= D_{s2s}(\{f_W^k(\boldsymbol{X_i})\}_{k=1}^m, \{f_W^k(\boldsymbol{X_j})\}_{k=1}^m)$$
$$= H_{\boldsymbol{M}}\left(D_A^1\left(f_W^1(\boldsymbol{X_i}), f_W^1(\boldsymbol{X_j})\right), \dots, D_A^m\left(f_W^m(\boldsymbol{X_i}), f_W^m(\boldsymbol{X_j})\right)\right) \quad (7)$$

In order to build the learning protocol, the analytical parametric form of the objective contrastive loss function $\mathcal{L}(\Theta, \mathrm{S}, \mathrm{D}, \mathrm{Y})$ is:

$$\mathcal{L}(\Theta, \mathrm{S}, \mathrm{D}, \mathrm{Y}) = \frac{1}{|S|}\sum_{i,j\in S} y_{ij} \cdot \max\left(D^{\Theta}(\boldsymbol{X_i}, \boldsymbol{X_j}) - \zeta_S, 0\right)^2$$
$$+\frac{1}{|D|}\sum_{i,j\in D}(1 - y_{ij}) \cdot \max\left(\zeta_D - D^{\Theta}(\boldsymbol{X_i}, \boldsymbol{X_j}), 0\right)^2 + \xi \cdot \gamma(\boldsymbol{M})$$
$$s.t. \boldsymbol{M} \in S_{++}^m, \boldsymbol{W} \in St(mp, n) \quad (8)$$

where $|S|$ and $|D|$ are the cardinality number of the similar ω^+ and dissimilar ω^- SPD signature covariance pairs. The intuition behind eq.(8) is that the distance between two similar signature covariance matrices should be smaller than a threshold ζ_S while the distance between two dissimilar signature covariance matrices should be larger than ζ_D. In the above formulation, we impose a) the learnable matrix $\boldsymbol{W} \in St(mp, n)$ with an orthogonality constraint (or belonging to the Stiefel manifold [88]) and b) the learnable matrix $\boldsymbol{M} \in S_{++}^m$ with a SPD constraint, in order to provide a robust distance measure. The regularization term of the loss function in (8) is implemented by the Burgman matrix divergence [78] between $\boldsymbol{M}$ and a prior matrix $\boldsymbol{M_0}$ i.e. $\gamma(\boldsymbol{M}) = tr(\boldsymbol{M}\boldsymbol{M}_0^{-1}) - logdet(\boldsymbol{M}\boldsymbol{M}_0^{-1}) - m$. The manifold nature of the learnable constraints $\boldsymbol{W}$ and $\boldsymbol{M}$ clearly suggests that Riemannian optimization methods can be applied. In the subsequent section, we discuss, in more detail, the implementation of the point-to-set and the set-to-set distances.

A. Point-to-Set Transformation

The proposed point-to-set transformation maps any original manifold points $\boldsymbol{X_i} \in \mathbb{R}^{n\times n}$ to m-numbered individual manifolds $\boldsymbol{X}_i^{1\dots m} \in \mathbb{R}^{p\times p}$:

$$\boldsymbol{X}_i^1 = f_{\boldsymbol{W}}^1(\boldsymbol{X}_i) = \boldsymbol{W}_1^T\boldsymbol{X}_i\boldsymbol{W}_1, \dots, \boldsymbol{X}_i^m = f_{\boldsymbol{W}}^m(\boldsymbol{X}_i) = \boldsymbol{W}_m^T\boldsymbol{X}_i\boldsymbol{W}_m \quad (9)$$

with $\boldsymbol{W}_k \in \mathbb{R}^{n\times p}$ to be the k-th mapping kernel in one low-dimensional SPD matrix $\boldsymbol{X}_i^k \in \mathbb{R}^{p\times p}$. We hypothesize that each one of the new $\boldsymbol{X}_i^k$ matrices can also provide discriminative visual information. Following, an SPD set $\mathcal{X}_i$ is formed by assembling the set of $\{\boldsymbol{X}_i^k\}_{k=1}^m = \{f_W^k(\boldsymbol{X_i})\}_{k=1}^m$ SPD matrices. Each of the $\boldsymbol{W}_k$ matrices must be full-column rank in order to ensure that the derived $\boldsymbol{X}_i^k$ matrices are SPD. Evidence is provided that this is achieved by imposing $\boldsymbol{W}_k$ on the subsequent orthogonality constraint: $\boldsymbol{W}_k^T\boldsymbol{W}_{l\neq k} = \boldsymbol{0} \in \mathbb{R}^{p\times p}$. We begin our analysis by referring to a theorem which states that: given two matrices $\boldsymbol{X}, \boldsymbol{Z} \in P_n$ and a projection matrix $\boldsymbol{W} \in \mathbb{R}^{n\times p}, p \leq n$, the objective function $\mathcal{L}\left(D_{(\alpha,\beta)}(\boldsymbol{W^T X W} \| \boldsymbol{W^T Z W})\right)$ can be optimized without loss of generality if we impose the orthogonality constraint [57]. Any $D^k_{(\alpha_k,\beta_k)}$ is equipped with the affine invariance property, that is, for an $\boldsymbol{R} \in GL(p)$ orthogonal matrix we have: $D^k_{(\alpha_k,\beta_k)}(\boldsymbol{W}_k^T\boldsymbol{X}\boldsymbol{W}_k \| \boldsymbol{W}_k^T\boldsymbol{Z}\boldsymbol{W}_k) = D^k_{(\alpha_k,\beta_k)}(\boldsymbol{R^T}\boldsymbol{W}_k^T\boldsymbol{X}\boldsymbol{W}_k\boldsymbol{R} \| \boldsymbol{R^T}\boldsymbol{W}_k^T\boldsymbol{Z}\boldsymbol{W}_k\boldsymbol{R})$, which results in the invariance of the objective function under the right action of any $\boldsymbol{W}_k$ orthogonal matrix. Therefore, the $\boldsymbol{W}_k$ lies in a Grassman manifold [88]. To continue, if $\boldsymbol{W}_k$ and $\boldsymbol{W}_{l\neq k}$ belong to the Grassman manifold, then it is known that the projection distance between them is provided by the relation $p - \|\boldsymbol{W}_k^T \boldsymbol{W}_l\|_F$. Thus, the projection distance is maximized when the orthogonality constraint: $\boldsymbol{W}_k^T \boldsymbol{W}_{l\neq k} = \boldsymbol{0} \in \mathbb{R}^{p\times p}$ is applied; as a consequence, we expect to derive as much as possible discriminative visual information from the individuals $\boldsymbol{W}_k$ and $\boldsymbol{W}_{l\neq k}$. An efficient way to compute the $\boldsymbol{X}_i^k$ is by observing that the set of $\{\boldsymbol{W}_k \in \mathbb{R}^{n\times p}\}_{k=1}^m$ matrices generates a total projection matrix $\boldsymbol{W} = [\boldsymbol{W}_1, \boldsymbol{W}_2, \dots, \boldsymbol{W}_m] \in \mathbb{R}^{n\times mp}$. For an original signature covariance matrix $\boldsymbol{X_i} \in P_n$ we can apply a diagonal block binary mask on the matrix $\boldsymbol{W^T X_i W}$ as:

$$\begin{aligned} \boldsymbol{Z_i} &= mask(\boldsymbol{W^T X_i W}) \\ &= mask\left(\begin{bmatrix} \boldsymbol{W}_1^T\boldsymbol{X_i}\boldsymbol{W}_1 & \cdots & \boldsymbol{W}_1^T\boldsymbol{X_i}\boldsymbol{W}_m \\ \vdots & \vdots & \vdots \\ \boldsymbol{W}_m^T\boldsymbol{X_i}\boldsymbol{W}_1 & \cdots & \boldsymbol{W}_m^T\boldsymbol{X_i}\boldsymbol{W}_m \end{bmatrix}\right) \\ &= \begin{bmatrix} \boldsymbol{X_i^1} & \cdots & 0 \\ \vdots & \boldsymbol{X_i^2} & \vdots \\ 0 & \cdots & \boldsymbol{X_i^m} \end{bmatrix} \end{aligned} \tag{10}$$

in order to obtain a corresponding diagonal block matrix $\boldsymbol{Z_i} \in P_{mp}$ which is equivalent to the SPD set $\mathcal{X}_i$. The advantages of this form are a) the efficient use of the eigen-decomposition and matrix inversion algorithms as well as b) the diagonal form of $\boldsymbol{Z_i}$ which elaborates our analysis and focuses on the essential visual information.

B. Set-to-Set Distance

Following the definition and the implementation details of the point-to-set transformation, we now provide details regarding the implementation of the set-to-set distance and its relation to the $H_{\boldsymbol{M}}(\cdot)$ integration function. Specifically, we express the $D_{s2s}(\cdot,\cdot)$ as the learnable combination function $H_{\boldsymbol{M}}$ of the family of local alpha-beta divergences $\{D_{\boldsymbol{A}}^k(\cdot,\cdot)\}_{k=1}^m$, with $\boldsymbol{A} \in \mathbb{R}^{m\times 2}$ the set of learnable parameters $\{(\alpha_k, \beta_k)\}_{k=1}^{\mu}$. For any arbitrary k-th low dimensional manifold and one pair of locally projected SPD points $\{\boldsymbol{X}_i^k, \boldsymbol{X}_j^k\}$ the $D_{\boldsymbol{A}}^k(\boldsymbol{X}_i^k, \boldsymbol{X}_j^k)$ is expressed by:

$$\begin{aligned} d_{ij}^k &= D_{\boldsymbol{A}}^k(\boldsymbol{X}_i^k, \boldsymbol{X}_j^k) = D_{(\alpha_k,\beta_k)}^k(\boldsymbol{X}_i^k \| \boldsymbol{X}_j^k) \\ &= \frac{1}{\alpha_k \cdot \beta_k} \sum_{u=1}^{p} log\left(\frac{\alpha_k (\lambda_{iju}^k)^{\beta_k} + \beta_k (\lambda_{iju}^k)^{-\alpha_k}}{\alpha_k + \beta_k}\right) \end{aligned} \tag{11}$$

in which, λ_{iju}^k is the u-th eigenvalue of the matrix $\boldsymbol{X}_i^k (\boldsymbol{X}_j^k)^{\boldsymbol{-1}}$. As a result, the $D_{s2s}(\mathcal{X}_i, \mathcal{X}_j)$ provide a vector comprised of any local distances $\boldsymbol{d}_{ij} = [d_{ij}^1, \dots d_{ij}^m] \in \mathbb{R}^m$. Given the fact that these contribute to the final $D_{s2s}(\cdot,\cdot)$ we elaborate the point-to-point distance of eq. (7) as to its final form:

$$\begin{aligned} \Delta(\Theta, \boldsymbol{X_i}, \boldsymbol{X_j}) &= D_{s2s}(\mathcal{X}_i, \mathcal{X}_j) = H_{\boldsymbol{M}}([d_{ij}^1, \dots d_{ij}^m]) = \boldsymbol{d}_{ij}^T \boldsymbol{M} \boldsymbol{d}_{ij} \\ &= \sum_{k=1}^{m}\sum_{l=1}^{m} (d_{ij}^k M_{kl} d_{ij}^l) \end{aligned} \tag{12}$$

where $\boldsymbol{M}$ is the learnable integration parameter and M_{kl} is the corresponding k-row and l-column element of $\boldsymbol{M}$. A physical interpretation of M_{kl} is that it provides a measure of any potential influence of the locally mapped visual information between $\boldsymbol{X_i}, \boldsymbol{X_j}$ along with their correlation. An examination of the relation (12) reveals the following: a) if $\boldsymbol{X_i} = \boldsymbol{X_j}$ then $\Delta(\Theta, \boldsymbol{X_i}, \boldsymbol{X_j}) = 0$, b) if $\boldsymbol{X_i} \neq \boldsymbol{X_j}$ then due to their SPD nature, they have positive eigenvalues; therefore $\boldsymbol{d}_{ij} \in \mathbb{R}_+^m$ and the corresponding $\Delta(\Theta, \boldsymbol{X_i}, \boldsymbol{X_j})$ is positive. The above observations result in the profound constraint of the learnable parameter $\boldsymbol{M}$ to the SPD manifold P_m. An exploration of the resemblance between the well-known Mahalanobis distance and the proposed mathematical framework with the use of $\Delta(\Theta)$ indicates that the projected local manifolds and the corresponding block-diagonal PSD matrices can be regarded as an intuitive generalization of the Euclidean vectors which comprise the Mahalanobis distance.

C. Learning Protocol and Optimization Algorithm

This section links the mathematical model of the proposed SPD distance $\Delta(\Theta)$ with the proposed WI-SV framework. The learning procedure, as formulated in eq. (8), commences by setting the hyper-parameters ξ, ζ_S and ζ_D to fixed values. Then, iteratively the algorithm updates the learnable parameters $\Theta = \{\mathbf{W}, \mathbf{A}, \mathbf{M}\}$ by using a mini-batch of similar $(\boldsymbol{X_i}, \boldsymbol{X_j})_S$ and dissimilar $(\boldsymbol{X_i}, \boldsymbol{X_j})_D$ pairs in order to optimize the objective loss function $\mathcal{L}(\Theta)$. Assembling the mini-batch is discussed here by assuming a template signature dataset $\mathcal{D}$ comprised by a total of $\mathbb{N}_{\mathcal{D}}$ writers. Each writer is represented by his/her positive and negative classes (Ω^+), (Ω^-) with $\mathbb{N}_{\Omega^+}$ genuine and $\mathbb{N}_{\Omega^-}$ simulated (or skilled forgery) signature samples. All signature samples are converted to corresponding covariance matrices according to the material exposed in section II.A.

A development subset $\mathbb{L}_{\mathbb{DW}}$ of total $\mathbb{N}_{\mathbb{DW}}$ writers is randomly selected for the learning stage of the proposed SPD distance $\Delta(\Theta)$ while the remaining $\mathbb{N}_{\mathbb{T}_{\mathbb{s}}}$ writers are employed in the testing stage $\mathbb{T}_{\mathbb{s}}$. The development set is comprised by the training set $\mathbb{T}_{\mathbb{R}}$ and the validation set $\mathbb{V}$. The training set $\mathbb{T}_{\mathbb{R}} = \{\mathbb{T}_{\mathbb{R}+}, \mathbb{T}_{\mathbb{R}-}\}$ is being built as follows: For each one of the $\mathbb{N}_{\mathbb{DW}}$ writers, seventy percent of the $\mathbb{N}_{\Omega^+}$ genuine signatures, denoted now as $\mathbb{N}_{\omega^+} = 0.7\mathbb{N}_{\Omega^+}$ are selected. Then, similar covariance pairs $(\boldsymbol{X_i}, \boldsymbol{X_j})_S$ are formed in order to create the corresponding set $\omega^{\mathbb{T}_{\mathbb{R}+}}$ with its cardinality $|S|$ equals to a theoretical total of $\mathbb{N}_{\mathbb{DW}} \times$ (*#training-pairs/writer*). According to the discussion exposed in section I.B., the dissimilar genuine-forgery pairs $(\boldsymbol{X_i}, \boldsymbol{X_j})_D$ which form the $\omega^{\mathbb{T}_{\mathbb{R}-}}$ set can be of different types. Therefore, we explore two distinct setups for the formation of the second half of the $(\boldsymbol{X_i}, \boldsymbol{X_j})_D$ pair. In the first setup named as $\omega_{100\%RF}^{\mathbb{T}_{\mathbb{R}-}}$, we use exclusive random forgeries for pairing with genuine samples. In the second one, defined as $\omega_{0\%RF}^{\mathbb{T}_{\mathbb{R}-}}$ only skilled forgeries shall be utilized. For this purpose, we also select a number of seventy percent $\mathbb{N}_{\omega^-} = 0.7\mathbb{N}_{\Omega^-}$ for the negative class training stage. Whenever possible, the cardinality number $|D|$ of the dissimilar set $\omega^{\mathbb{T}_{\mathbb{R}-}}$ has also been set equal to $|S|$. The aforementioned elements are displayed in a tabulated form in Table I for a total of the four datasets that were used. Additional details regarding the datasets will be provided later.

This article has been accepted for publication in IEEE Transactions on Information Forensics and Security. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIFS.2023.3333681

Algorithm 1: Learn the parameters $\Theta = \{\mathbf{W}, \mathbf{A}, \mathbf{M}\}$ of the $\Delta(\Theta)$ SPD distance.

Require: A mini-batch of similar $(\boldsymbol{X_i}, \boldsymbol{X_j})_S$ and dissimilar $(\boldsymbol{X_i}, \boldsymbol{X_j})_D$ pairs, the loss function $\mathcal{L}(\Theta, \mathrm{S}, \mathrm{D}, \mathrm{Y})$ of (8).

BEGIN

1: **SET:** Learnable parameters $\mathbf{W_0}, \boldsymbol{A}_0, \mathbf{M_0}$ to arbitrary values.

2: **SET:** Iteration t_{stop} parameter, learning rate η, validation trigger *epoch.*

3: **FOR** $t = 1{:}\, t_{stop}$

4: **EVALUATE**: Loss function $\mathcal{L}(\Theta, \mathrm{S}, \mathrm{D}, \mathrm{Y})$ (8)

5: **UPDATE** the parameter $\boldsymbol{A}$ with: $\boldsymbol{A}_t = \boldsymbol{A}_{t-1} - \eta \left(\frac{\partial \mathcal{L}}{\partial \boldsymbol{A}_{t-1}}\right)$ (13)

6: **UPDATE** the $\mathbf{W}_t$ parameter according to the following rules:

6a: Convert the Euclidean gradient $\left(\frac{\partial \mathcal{L}}{\partial \boldsymbol{W}}\right)$ *to its Riemannian counterpart* $\left(\frac{\partial \mathcal{L}}{\partial \boldsymbol{W^R}}\right)$ *according to:* $\frac{\partial \mathcal{L}}{\partial \boldsymbol{W}^R_{t-1}} = \frac{\partial \mathcal{L}}{\partial \boldsymbol{W}_{t-1}} - \boldsymbol{W}_t \frac{1}{2}\left(\mathbf{W}^{\boldsymbol{T}} \frac{\partial \mathcal{L}}{\partial \boldsymbol{W}_{t-1}} + \frac{\partial \mathcal{L}}{\partial \boldsymbol{W}_{t-1}}^T \boldsymbol{W}\right)$ (14)

6b: The Riemannian gradient of $\boldsymbol{W}$ *belongs to a Stiefel tangent (vector) space,* ***update W*** *by applying the retraction operation* $q(\cdot)$ *on the Stiefel manifold which projects the tangent vector back to the Riemannian manifold:*

$$\boldsymbol{W}_t = q\left(\boldsymbol{W}_{t-1} - \eta \frac{\partial \mathcal{L}}{\partial \boldsymbol{W}^R_{t-1}}\right) \quad (15)$$

7: **UPDATE** the $\mathbf{M}_t$ parameter according to the following rules:

7a: Convert the Euclidean gradient $\left(\frac{\partial \mathcal{L}}{\partial \boldsymbol{M}}\right)$ *to its Reimannian counterpart* $\left(\frac{\partial \mathcal{L}}{\partial \boldsymbol{M^R}}\right)$ *according to* $\frac{\partial \mathcal{L}}{\partial \boldsymbol{M}^R_{t-1}} = \boldsymbol{M}_{t-1} \frac{1}{2}\left(\frac{\partial \mathcal{L}}{\partial \boldsymbol{M}_{t-1}} + \frac{\partial \mathcal{L}}{\partial \boldsymbol{M}_{t}\mathbf{-1}}^T\right) \boldsymbol{M}_{t-1}$ (16)

7b: The Riemannian gradient of $\boldsymbol{M}$ *belongs to a SPD tangent (vector) space,* ***update M*** *by applying the retraction operation* $expm(\cdot)$ *on the SPD manifold which projects the tangent vector back to the Riemannian manifold:* $\boldsymbol{M}_t = \mathbf{M}^{1/2}_{t-1} expm(-\eta \mathbf{M}^{-1/2}_{t-1} \frac{\partial \mathcal{L}}{\partial \boldsymbol{M}^R_{t-1}} \mathbf{M}^{-1/2}_{t-1}) \mathbf{M}^{1/2}_{t-1}$ (17)

8: **IF** t **EQUAL** to *epoch*

9: **RUN** VALIDATION

10: **end_IF**

11: **end_FOR**

END

As anticipated, the optimization procedure of the objective loss function is a non-jointly convex function of its learning parameters. We also note here that the learnable parameters $\mathbf{W}, \mathbf{M}$ lie in the Grassman and SPD manifolds respectively, which makes the overall optimization of the loss function a challenging problem. The Θ set parameters are optimized by the stochastic gradient descent (SGD) algorithm. Given $\boldsymbol{A} \in \mathbb{R}^{m \times 2}$, its update stage will rely on the Euclidean gradient $\frac{\partial \mathcal{L}}{\partial \boldsymbol{A}}$ while the Riemannian constraints of $\mathbf{W}, \mathbf{M}$ impose the use of Riemannian gradients $\frac{\partial \mathcal{L}}{\partial \boldsymbol{W^R}}$ and $\frac{\partial \mathcal{L}}{\partial \boldsymbol{M^R}}$ [89]. Algorithm 1, provides a description of the optimization steps. In steps (*6b*) and (*7b*), the retraction operators of $q(\mathbf{W})$ and $expm(\cdot)$ are defined by a) the $\mathbf{Q}$-part of the QR decomposition of $\mathbf{W}$, i.e. $\mathbf{W} = \mathbf{QR}$ and b) the matrix exponential function of (3). The interested reader may find analytical details regarding the implementation of a) the gradients of the $\mathcal{L}(\Theta)$ with respect to the individuals $\boldsymbol{A}_t = (\alpha_k, \beta_k)_t$, $(\mathbf{W}_k)_t$ (with $k = 1, \dots m$), $\mathbf{M}_t$ and $\partial\gamma(\boldsymbol{M}, \boldsymbol{M_0})/\partial \boldsymbol{M}$ at time step t, and b) the properties of the convergence of the procedure at the SPD and Grassman manifolds in pivotal references such as [57], [89], [90]. In all baseline experiments, ξ was set to 0.01 and the prior matrix $\boldsymbol{M}_0$ was set to $\boldsymbol{I}_m$. Additionally, in all our datasets, the thresholds ζ_S and ζ_D were set to 0.5 and 20, the learning rate η was set to 10^{-3} and the size of the mini-batch was set to 400 i.e. 200 similar and 200 dissimilar pairs. When the iteration step t equals an integer multiple of a fixed epoch value (e.g. *epoch*=200), the validation set $\mathbb{V}$ is employed with a stopping condition in order to select and return the optimal $\Delta(\Theta)$. The set $\mathbb{V}$ is comprised by the remaining $\mathbb{N}^{\mathbb{V}}_{\omega^{\pm}} = 0.3\mathbb{N}_{\Omega^{\pm}}$ associated signature covariance matrices. Algorithm 2 provides a description of the validation steps.

TABLE I

LEARNING PARAMETERS FOR THE SIGNATURE DATASETS

Notation	Description	$\mathcal{D}_1$	$\mathcal{D}_2$	$\mathcal{D}_3$	$\mathcal{D}_4$
$\mathbb{N}_{\mathcal{D}}$	*# Writers*	55	75	100	160
$\mathbb{N}_{\Omega^+}, \mathbb{N}_{\Omega^-}$	*# Genuine & # Simulated-skilled forgery samples per writer*	24/24	15/16	24/30	
$\mathbb{N}_{\mathbb{DW}}, \mathbb{N}_{\mathbb{T}_s}$	*# Development & # Testing writers*	28/27	38/37	50/50	80/80
$\mathbb{N}_{\omega^+}$	*# Samples per writer for the* $\mathbb{T}_{\mathbb{R}}$:	17	11	17	
$\omega^{\mathbb{T}_{\mathbb{R}+}}$ & $\lvert S\rvert$	$\mathbb{T}_{\mathbb{R}+}$*: Set of similar covariance pairs & corresponding cardinality*	3808	2090	6800	10880
$\omega^{\mathbb{T}_{\mathbb{R}-}}_{100\%RF}$ & $\lvert D\rvert$	$\mathbb{T}_{\mathbb{R}-}$*: Set of dissimilar covariance pairs & corresponding cardinality*				
$\omega^{\mathbb{T}_{\mathbb{R}-}}_{0\%RF}$ & $\lvert D\rvert$					

Algorithm 2: Validate the parameters $\Theta = \{\mathbf{W}, \mathbf{A}, \mathbf{M}\}$ of the $\Delta(\Theta)$ SPD distance.

Requires: The trained $\Delta(\Theta)$ model of **Algorithm 1,** Minimum value of Area under Curve (AUC) from previous executions of **Algorithm 2**.

BEGIN

1: **CREATE:** $\omega^{\mathbb{V}+}$ set of similar pairs by pairing genuine samples of the validation set.

2: **CREATE:** $\omega^{\mathbb{V}-}$ set of dissimilar pairs by pairing genuine samples with simulated-skilled forgeries only.

3: **EVALUATE**: $\Delta(\Theta)$ with the $\omega^{\mathbb{V}+}$ and $\omega^{\mathbb{V}-}$ sets, and denote their scores by $\Delta(\Theta, \omega^{\mathbb{V}+})$ and $\Delta(\Theta, \omega^{\mathbb{V}-})$.

4: **PERFORM:** Receiver operating characteristic (ROC) analysis with $\Delta(\Theta, \omega^{\mathbb{V}+})$ and $\Delta(\Theta, \omega^{\mathbb{V}-})$.

4a: ***MEASURE***: $AUC_{current}$

5: **IF** $AUC_{current} < min(AUC)$ ***then****:*

5a: **RESUME: Algorithm 1** *of (13)-(17) with new mini-batches*
RETURN: new value of $min(AUC)$.

5*b*: **ELSEIF** $AUC_{current} \geq min(AUC)$ for a number of predetermined number of epochs N_{epochs}, then:
HALT:
RETURN: $\Delta(\Theta)$ which corresponds to the $min(AUC)$.

6: **end_IF**

END

V. EXPERIMENTS

A. The datasets

Four popular offline handwritten signature datasets $\mathcal{D}_{1-4}$ of western and Indo-Aryan origin were employed in order to evaluate the proposed WI-SV SPD distance model. They are: the Western styled a) $\mathcal{D}_1$=CEDAR [91], b) the $\mathcal{D}_2$=MCYT-75 [92] and the Asian oriented c) $\mathcal{D}_3$=BENGALI and d) $\mathcal{D}_4$=HINDI, which are the two sub-sets of the Indo-Aryan BHSig260 database [93]. Table I, provides any essential information regarding the four datasets.

B. Protocol

The proposed WI-SV was explored with two key experimental blind frameworks, denoted as $\mathcal{F}_{intra}$ and $\mathcal{F}_{inter}$. The $\mathcal{F}_{intra}$ follows a typical 5×2 blind fold for each individual dataset. In detail, for each $\mathcal{D}_i$ and according to the protocol described for the template dataset in section IV.c, we alternate the role of the learning (i.e. training & validation) and testing datasets $\mathbb{L}_{\mathbb{DW}}$ and $\mathbb{T}_{\mathbb{s}}$. Specifically, in each one of the five folds, the proposed distance $\Delta(\Theta, \mathbb{L}_{\mathbb{DW}})$ is learned with the $\mathbb{L}_{\mathbb{DW}}$ while its efficiency is evaluated at $\mathbb{T}_{\mathbb{s}}$ and then, the $\mathbb{L}_{\mathbb{DW}}$ and $\mathbb{T}_{\mathbb{s}}$ exchange roles. The second blind $\mathcal{F}_{inter}$ bear a resemblance to a transfer learning protocol which utilizes blind datasets for learning and testing. In detail, the proposed distance $\Delta(\Theta, \mathcal{D}_i)$ is learned in the entire dataset $\mathcal{D}_i$ with $\mathbb{N}_{\mathbb{DW}} = \mathbb{N}_{\mathcal{D}_i}$ and then, it is evaluated in the remaining three datasets $\mathcal{D}_{j \neq i}$.

This article has been accepted for publication in IEEE Transactions on Information Forensics and Security. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIFS.2023.3333681

The results on both $\mathcal{F}_{intra,inter}$ WI-SV frameworks are reported under three modus operandi, denoted hereafter as $\mathfrak{W}_1$ $\mathfrak{W}_2$ and $\mathfrak{W}_3$. The $\mathfrak{W}_1$, addresses a very simple test query without considering the identity of any writer; one reference signature $\boldsymbol{X_{REF}}$ and one questioned $\boldsymbol{X_Q}$ SCM are presented in the learned $\Delta(\Theta)$ model which evaluates a score $sc_{(\boldsymbol{X_{REF}},\boldsymbol{X_Q})} = \Delta(\Theta)^{\boldsymbol{X_Q}}_{\boldsymbol{X_{REF}}}$. The derived scores from all testing pairs are evaluated and conditioned as similar $sc_{\mathbb{T}_{\mathbb{s}}+}$ and dissimilar $sc_{\mathbb{T}_{\mathbb{s}}-}$. Then, a global sliding threshold evaluates the global false acceptance ($\mathrm{FAR_{SF}}$), the false rejection rates (FRR) and eventually reports the global equal error rate $EER_{SF}^{\mathfrak{W}_1-global}$. The $\mathfrak{W}_2$ is similar to $\mathfrak{W}_1$ but it operates and reports the $EER_{SF}^{\mathfrak{W}_2-local}$ at an average local (or user) level. It must be noted here that the reported results of $\mathfrak{W}_1$ and $\mathfrak{W}_2$ are for the case of using only one global $\boldsymbol{C}_{SCM}^{1\times1}$ for each signature image. The notation 1×1 signifies the global covariance matrix derived from all signature pixels.

Regarding the $\mathfrak{W}_3$, a more detailed user defined protocol is explored in which: a) any questioned $\boldsymbol{X_Q}$ SCM is paired alongside a reference set $\mathcal{G}_R$ comprised by five $\mathcal{G}_{NREF}=5$ or ten $\mathcal{G}_{NREF}=10$ genuine reference samples and b) instead of using only one global $\boldsymbol{C}_{SCM}^{1\times1}$ we represented any signature image with a set of four $\boldsymbol{C}_{SCM}^{\{2\times2\}}$ and nine $\boldsymbol{C}_{SCM}^{\{3\times3\}}$ covariance matrices as described in section II.A. Thus, each image is represented by a set of 14 covariance matrices $\boldsymbol{C}^{\{k=1,..14\}}$. Algorithm 3 presents in detail the implementation steps for one user. Summarizing, the score between the $\boldsymbol{X_Q}$ and one reference sample $\boldsymbol{X_{REF}}$ is evaluated in step (5)-(8) by applying $\Delta(\Theta)$ to each pair of segments. These fourteen local scores are sorted and then, in step (10) the *seg_idx* parameter, which denotes the number of participating segments, provide a local score by averaging (step 10c). Finally the minimum value of all scores between $\boldsymbol{X_Q}$ and all reference samples $\boldsymbol{X_{REF}}$ is selected and denotes hereafter the $\boldsymbol{Q}$-score of the $\boldsymbol{X_Q}$, conditioned on the remaining genuine of forgery side of the testing set. The two ends i.e. *seg_idx* =1,14 relate to the error rates when: a) only the global covariance participates and b) all 14 segments are accounted.

Algorithm 3: The basic $\mathfrak{W}_3$ WI-SV protocol for one user

Requires: (a) The learned $\Delta(\Theta)$ model of **Algorithms 1, 2**. (b) the genuine and simulated-skilled forgery signature samples of a user (c) a set of 14 covariance matrices $\boldsymbol{C}_{SCM}^{(k=1,\dots,14)}$ of each signature image.

BEGIN
REPEAT 10 times
1: **CREATE:** The set $\mathcal{G}_R$ comprised of randomly selected $\mathcal{G}_{NREF}$ genuine reference samples.
2: **CREATE:** The $\mathbb{T}_{\mathbb{s}}$ test set comprised from the questioned remaining genuine ($\mathbb{T}_{\mathbb{s}+}$) and simulated-skilled forgery samples ($\mathbb{T}_{\mathbb{s}-}$)
3: **FOR(**i**)**: Each one of the questioned $\boldsymbol{Q}_i \in \{\mathbb{T}_{\mathbb{s}+},\mathbb{T}_{\mathbb{s}-}\}$
4: **FOR(**j**)**: Each one $\boldsymbol{R}_j \in \mathcal{G}_R$ of the $\mathcal{G}_{NREF}$ genuine reference samples.
5: **FOR(**k**)**: each one of the 14 segments
5a: **USE**: $\boldsymbol{Q}_{i}{}_{SCM}^{(k)}$ as the $\boldsymbol{C}_{SCM}^{k}$ of $\boldsymbol{Q}_i$
5b: **USE**: $\boldsymbol{R}_{j}{}_{SCM}^{(k)}$ as the $\boldsymbol{C}_{SCM}^{k}$ of $\boldsymbol{R}_j$
6: **EVALUATE**: temporary scores $tsc_{ij}^{k} = \Delta_{ij}^{k}\left(\Theta, \boldsymbol{Q}_{i}{}_{SCM}^{(k)}, \boldsymbol{R}_{i}{}_{SCM}^{(k)}\right)$.
7: **end_FOR(**k**)**
8: **end_FOR(**j**)**
9: **FOR(***seg_idx***)** equal from 1up to 14 // *#segments encountered to the final score*
10: *IF seg_idx* equals 1 ***then:***
10a: **CREATE**: Scores $sc_{ij} = tsc_{ij}^{1}$, // *i.e. the derived from the* 1×1 *covs.*
Else
10b: **SORT**: $tsc_{ij}^{\{all\ k\}}$ in ascending order with respect to $k{:}\longrightarrow tsc_{ij}^{sorted\ (k)}$
10c: **EVALUATE**: Local scores $sc_{ij}^{seg\_idx} = mean(\{tsc_{ij}^{1:seg\_idx}\})$
10d: end_*IF*
11: **SET** final score of the $\boldsymbol{Q}_{\boldsymbol{i}}^{seg\_idx}$ to be the $min(sc_{ij}^{seg\_idx})$ // *over all j's*
11: **end_FOR(***seg_idx***)**
12: **end_FOR(**i**)**
//We have the scores $sc_{\boldsymbol{Q}_{\boldsymbol{i}}^{seg\_idx}}$ *for all i-samples parametrized on* seg_idx.
13: **FOR(***seg_idx***)** equal from 1up to 14
14: **PERFORM:** ROC analysis with scores from $sc_{\boldsymbol{Q}_{\boldsymbol{i}}^{seg\_idx}\in\mathbb{T}_{\mathbb{s}+}}$, $sc_{\boldsymbol{Q}_{\boldsymbol{i}}^{seg\_idx}\in\mathbb{T}_{\mathbb{s}-}}$
15: **MEASURE:** $EER_{user\_threshold}$ for all seg_idx.
17: **end_FOR(***seg_idx***)**
18: **end_REPEAT**
19: **RETURN:** average $EER_{user\_threshold}$ for one user and all seg_idx.
END

C. Results and Discussion

All experimental protocols were conducted for a number of combinations of the m,p hyper-parameters of $\Delta(\Theta)$, defined at section IV. The m parameter defines the number of block-diagonal SPD matrices $\boldsymbol{X}^{k=1:m,}$ and consequently the number of sub-distances $\{D_{\boldsymbol{A}}^{k}(\cdot,\cdot)\}_{k=1}^{m}$, and the p parameter defines the size of any block diagonal SPD matrix $\boldsymbol{X}^{k} \in \mathbb{R}^{p\times p}$ used in the evaluation of any $D_{\boldsymbol{A}}^{k}$. We recap here that the projection matrix $\boldsymbol{W} \in \mathbb{R}^{n\times m\cdot p}$ maps any initial signature covariance matrix from the $P_{n(=10)}$ manifold to a new $P_{m\cdot p}$ manifold with m-numbered local covariance matrices $\boldsymbol{X}^{k} \in P_p$. The first column of Table II show the values of m,p hyper-parameters that were used in the design of the experimental stage. Due to the relative small initial dimensionality of the P_{10} manifold the case in which the product $m\cdot p$ of the final SPD dimensionality equal 10 was also explored. The initial results for the $\mathcal{F}_{intra}$, the first two modus operandi $\mathfrak{W}_1,\mathfrak{W}_2$ and the two training setups $\mathbb{T}_{\mathbb{R}+}$, $\mathbb{T}_{\mathbb{R}-}$ are presented in Table II. For the case of the $\mathfrak{W}_3$ modus and both training setups of $\mathbb{T}_{\mathbb{R}-}$, (i.e. $\omega_{0\%RF}^{\mathbb{T}_{\mathbb{R}-}}$ and $\omega_{100\%RF}^{\mathbb{T}_{\mathbb{R}-}}$) the corresponding local EER's are

TABLE II
EQUAL ERROR RATES (%): $\mathcal{F}_{intra}$ BLIND FRAMEWORK FOR BOTH TYPES OF DISSIMILAR PAIRS AND $\mathfrak{W}_{1-2}$

m,p	CEDAR		MCYT		BENGALI		HINDI	
	DISSIMILAR PAIRS SETUP = $\omega_{100\%RF}^{\mathbb{T}_{\mathbb{R}-}}$							
	$\mathfrak{W}_1$	$\mathfrak{W}_2$	$\mathfrak{W}_1$	$\mathfrak{W}_2$	$\mathfrak{W}_1$	$\mathfrak{W}_2$	$\mathfrak{W}_1$	$\mathfrak{W}_2$
1,7	8.52	3.93	18.6	11.9	7.70	4.36	15.2	9.34
1,8	8.17	3.96	17.7	11.1	7.55	3.39	15.7	9.57
1,9	8.06	4.05	17.5	10.9	8.39	3.86	15.6	9.54
1,10	8.53	4.65	17.4	10.8	11.1	6.39	15.6	9.43
2,3	12.6	6.66	21.6	14.0	11.4	7.81	20.4	12.8
2,4	11.1	7.04	19.9	12.8	8.81	5.37	17.5	10.5
2,5	8.79	5.70	18.3	11.8	8.15	4.03	15.5	9.92
3,3	11.4	7.28	19.7	12.9	10.2	6.59	19.4	11.4
DISSIMILAR PAIRS SETUP = $\omega_{0\%RF}^{\mathbb{T}_{\mathbb{R}-}}$								
1,7	8.26	4.26	18.2	11.4	7.07	3.82	14.8	9.21
1,8	7.55	3.63	17.9	11.2	7.59	3.43	15.3	9.32
1,9	7.82	3.90	17.6	10.6	7.66	3.40	15.3	9.61
1,10	8.66	4.59	17.6	11.0	11.3	6.54	15.8	9.63
2,3	9.63	5.86	25.5	15.7	11.9	7.29	16.7	11.3
2,4	8.52	4.78	22.0	13.2	7.86	4.61	14.3	9.88
2,5	8.11	5.28	19.5	12.2	7.57	3.56	14.0	9.57
3,3	8.85	5.39	24.1	14.7	7.86	4.65	19.2	11.3

This article has been accepted for publication in IEEE Transactions on Information Forensics and Security. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIFS.2023.3333681

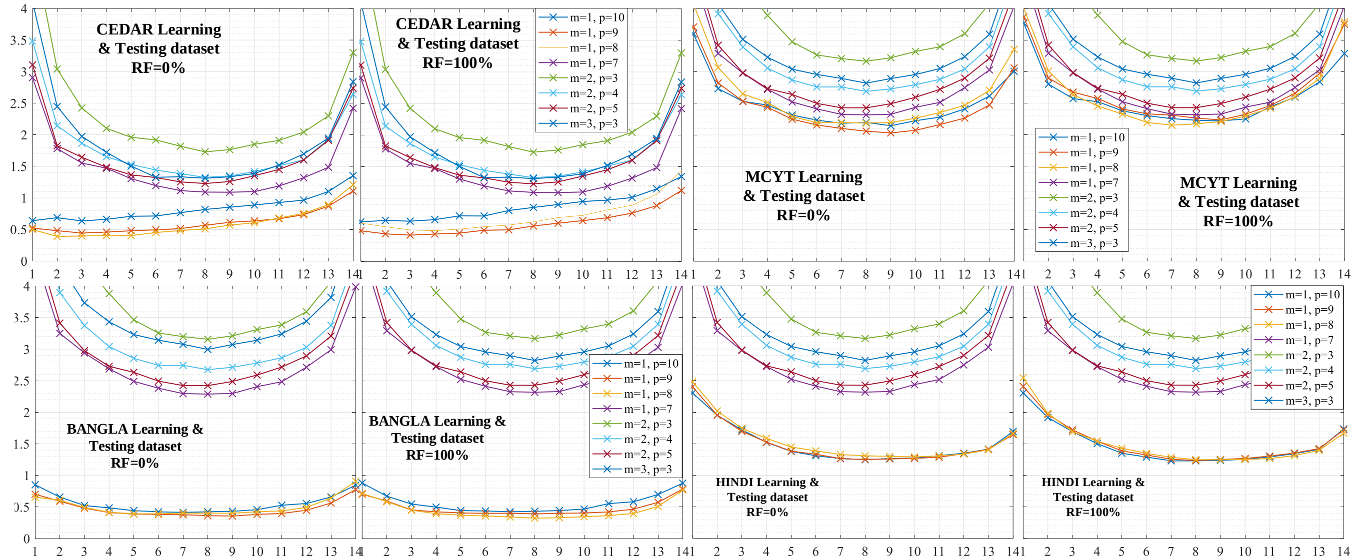

Fig. 2. Local EERs as a function of the *seg_idx=1,…14* parameter of Algorithm 3 for the $\mathcal{F}_{intra}$ blind framework, the $\mathfrak{W}_3$ modus, 5 reference samples and the two variants of the negative class pairs $\omega^{\mathbb{T}\mathbb{R}-}$ a) $\omega^{\mathbb{T}\mathbb{R}-}_{0\%RF}$ and b) $\omega^{\mathbb{T}\mathbb{R}-}_{100\%RF}$.

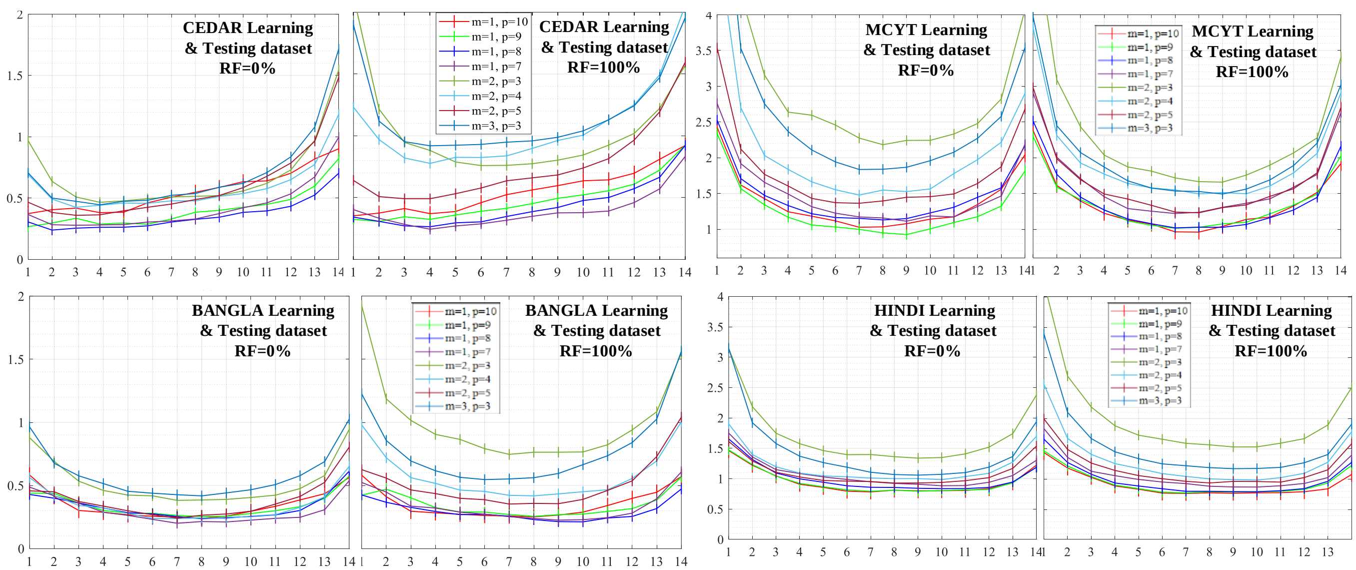

Fig. 3. Local EERs as a function of the *seg_idx=1,…14* parameter of Algorithm 3 for the $\mathcal{F}_{intra}$ blind framework, the $\mathfrak{W}_3$ modus, 10 reference samples and the two variants of the negative class pairs $\omega^{\mathbb{T}\mathbb{R}-}$ a) $\omega^{\mathbb{T}\mathbb{R}-}_{0\%RF}$ and b) $\omega^{\mathbb{T}\mathbb{R}-}_{100\%RF}$.

TABLE III

EQUAL ERROR RATES (%): $\mathcal{F}_2$ BLIND-FRAMEWORK FOR THE TWO TYPES OF DISSIMILAR PAIRS AND $\mathfrak{W}_{1-2}$ MODES.

m,p	Learning Sig. Set -CEDAR						Learning Sig. Set - MCYT					
	MCYT		*BANGLA*		*HINDI*		*CEDAR*		*BANGLA*		*HINDI*	
	DISSIMILAR PAIRS SETUP = $\omega^{\mathbb{T}\mathbb{R}-}_{100\%RF}$						DISSIMILAR PAIRS SETUP = $\omega^{\mathbb{T}\mathbb{R}-}_{100\%RF}$					
	$\mathfrak{W}_1$	$\mathfrak{W}_2$	$\mathfrak{W}_1$	$\mathfrak{W}_2$	$\mathfrak{W}_1$	$\mathfrak{W}_2$	$\mathfrak{W}_1$	$\mathfrak{W}_2$	$\mathfrak{W}_1$	$\mathfrak{W}_2$	$\mathfrak{W}_1$	$\mathfrak{W}_2$
1,8	17.5	9.75	9.01	4.55	15.4	9.62	11.0	4.88	11.1	3.54	15.3	8.04
1,9	17.2	8.92	8.49	3.56	15.6	9.03	9.77	4.44	11.3	4.66	15.6	8.13
1,10	16.8	9.08	9.97	4.07	15.5	8.29	8.70	3.93	10.9	4.62	15.6	8.24
2,5	17.2	10.0	11.2	4.89	16.8	9.90	12.6	6.28	10.2	4.89	16.4	9.08
	DISSIMILAR PAIRS SETUP = $\omega^{\mathbb{T}\mathbb{R}-}_{0\%RF}$						DISSIMILAR PAIRS SETUP = $\omega^{\mathbb{T}\mathbb{R}-}_{0\%RF}$					
1,8	18.9	10.4	7.37	2.44	16.8	10.3	10.0	4.49	12.0	4.53	16.0	8.39
1,9	17.5	9.49	7.79	3.34	16.1	9.38	8.98	3.84	10.9	4.22	15.5	8.17
1,10	17.0	9.39	9.83	4.08	15.5	8.35	8.57	3.81	10.3	4.29	15.5	8.27
2,5	19.4	11.1	11.3	5.58	17.5	9.87	12.6	6.45	13.0	5.55	17.6	9.48

m,p	Learning Sig. Set - BANGLA						Learning Sig. Set - HINDI					
	CEDAR		*MCYT*		*HINDI*		*CEDAR*		*MCYT*		*HINDI*	
	DISSIMILAR PAIRS SETUP = $\omega^{\mathbb{T}\mathbb{R}-}_{100\%RF}$						DISSIMILAR PAIRS SETUP = $\omega^{\mathbb{T}\mathbb{R}-}_{100\%RF}$					
	$\mathfrak{W}_1$	$\mathfrak{W}_2$	$\mathfrak{W}_1$	$\mathfrak{W}_2$	$\mathfrak{W}_1$	$\mathfrak{W}_2$	$\mathfrak{W}_1$	$\mathfrak{W}_2$	$\mathfrak{W}_1$	$\mathfrak{W}_2$	$\mathfrak{W}_1$	$\mathfrak{W}_2$
1,8	10.1	4.89	18.9	10.3	12.8	5.69	9.79	4.92	18.1	10.1	15.3	9.21
1,9	9.67	4.43	18.0	9.85	12.3	5.44	8.29	3.22	17.4	9.42	15.1	8.54
1,10	8.50	3.77	17.2	9.33	10.4	4.34	8.41	3.68	17.0	9.19	14.9	8.25
2,5	13.9	5.55	21.6	11.8	13.6	4.06	11.6	6.01	18.9	10.9	15.0	8.73
	DISSIMILAR PAIRS SETUP = $\omega^{\mathbb{T}\mathbb{R}-}_{0\%RF}$						DISSIMILAR PAIRS SETUP = $\omega^{\mathbb{T}\mathbb{R}-}_{0\%RF}$					
1,8	9.77	4.47	18.4	9.97	13.4	5.84	10.3	5.06	17.4	10.1	15.2	9.18
1,9	8.82	4.05	17.3	9.48	11.5	5.11	7.60	3.16	17.7	9.57	15.8	9.29
1,10	8.59	3.86	17.3	9.41	10.8	4.54	8.41	3.68	17.0	9.20	15.5	8.27
2,5	12.6	6.61	18.4	10.6	12.5	5.26	12.0	4.10	19.3	11.0	16.3	9.98

displayed in figures 2, 3 (for 5 and 10 references) as a function of the *seg_idx* parameter of Algorithm 3.

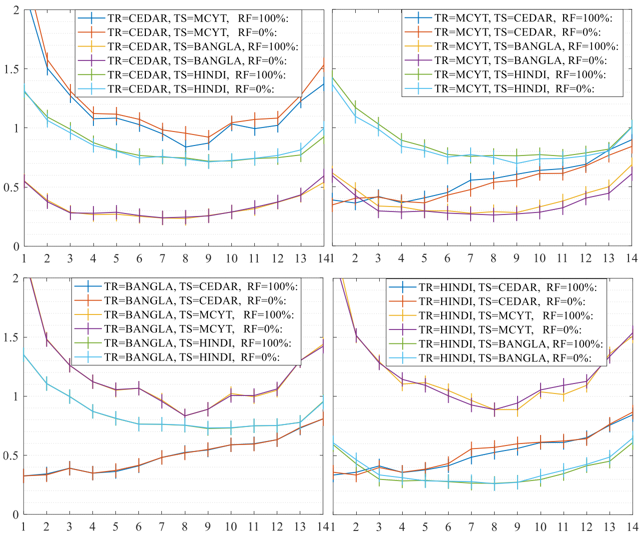

Fig. 4. Local EERs as a function of the *seg_idx* parameter of Algorithm 3 for the blind $\mathcal{F}_{inter}$ framework and the $\mathfrak{W}_3$ modus, with 10 reference samples. $m = 1, p = 10$.

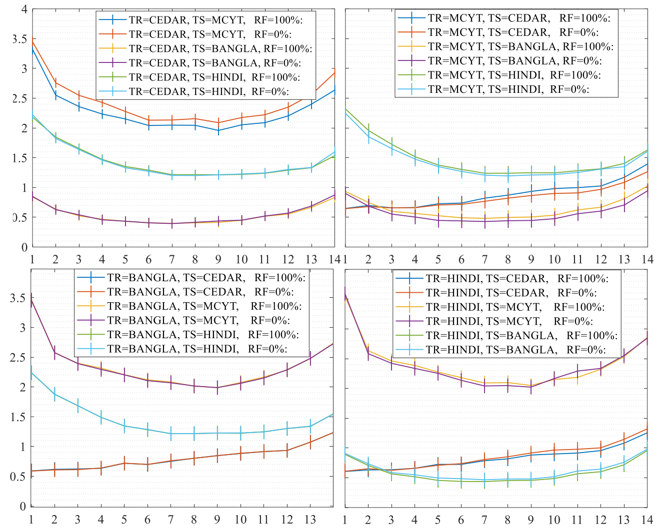

Fig. 5. Local EERs as a function of the *seg_idx* parameter of Algorithm 3 for the blind $\mathcal{F}_{inter}$ framework and the $\mathfrak{W}_3$ modus, with 5 reference samples. $m = 1, p = 10$.

The analogous experiments for the $\mathcal{F}_{blind}$, the first two modus operandi $\mathfrak{W}_1, \mathfrak{W}_2$ and the two training setups of $\mathbb{T}_{\mathbb{R}-}$ are shown in Table III. Figures 4, 5 displays the local $EER_{SF}^{\mathfrak{W}_3-local}$ for the case of the $\mathfrak{W}_3$ modus (10 and 5 references) and the training setups of $\mathbb{T}_{\mathbb{R}+}$, $\mathbb{T}_{\mathbb{R}-}$ as a function of the *seg_idx* parameter. The derived results are presented only for the case of having $m = 1, p = 10$ due to the fact that with this parameter pair we obtain robust results close enough to the lower verification error. Additional figures for several configurations of m, p of $\mathfrak{W}_3$ modus can be visually accessed in the supplementary material.

Commenting on the results for both $\mathcal{F}_{intra}$ and $\mathcal{F}_{blind}$, we initiate our discussion by addressing a number of broad issues. At first, we observe from figures 4 and 5 of the $\mathfrak{W}_3$ modus operandi, the undeniable fact that the best achievable verification error rates are reported when more than one image segment (and corresponding covariance), is involved in the score calculation. So, we state that the two ends of the *seg_idx* parameter of Algorithm 3 are not the ones that attain the lower verification error rates. Intuitively, the use of a) the global image and equivalent global $\boldsymbol{C}_{SCM}^{1\times1}$ covariance (*seg_idx*=1), as well as b) the entire fourteen image segments (*seg_idx*=14), clearly correspond to the two ends of the available granular level of information (i.e. too coarse, too fine) that participates in the decision. A complementary observation is that the optimal number of participating segments that achieve the lowest verification error rates in $\mathcal{D}_1$ (two up to four) in both $\mathcal{F}_{intra}$ and $\mathcal{F}_{blind}$ experiments is smaller when compared to the ones that provide the lowest error rates for $\mathcal{D}_{2-4}$ (seven up to ten). This is probably due to the fact that the signature images of the $\mathcal{D}_1$ dataset contain a smaller amount of signature pixels. Therefore any relative covariance derived from these segments might suffer from degraded properties, such as a much larger number of near zero eigenvalues and might not be as much discriminative as the covariance of the segments of the $\mathcal{D}_{2-4}$ sets.

Secondly, Figures 2, 3 and Table II shows that a direct comparison between the $\omega_{100\%RF}^{\mathbb{T}_{\mathbb{R}-}}$ and $\omega_{0\%RF}^{\mathbb{T}_{\mathbb{R}-}}$ of the $\mathcal{F}_{intra}$, reveals that the use of simulated-or-skilled forgeries $\omega_{0\%RF}^{\mathbb{T}_{\mathbb{R}-}}$ for the $\mathbb{T}_{\mathbb{R}-}$ learning set, leads to more robust results in terms of having more m, p parameter pairs with overall lower verification error rates. This is a somewhat anticipated outcome since each individual dataset has been constructed with the similar acquisition and a-priori conditions. Thus, the learned distance models trained with simulated-or-skilled forgery samples inherently provide the necessary generalization on the learning stage. The leverage of the $\omega_{0\%RF}^{\mathbb{T}_{\mathbb{R}-}}$ of the $\mathcal{F}_{intra}$ models weakens in the case of the $\mathcal{F}_{blind}$ since figures 4, 5 and Table III (as well as the figures provided in the supplementary material) indicate that the models that are learned with the $\omega_{100\%RF}^{\mathbb{T}_{\mathbb{R}-}}$ protocol are more robust when compared to the ones that are learned with the $\omega_{0\%RF}^{\mathbb{T}_{\mathbb{R}-}}$ protocol. The above perspective allows us to assert that the proposed SPD modelling with the learned distances operate efficiently in the WI-SV context even when blind datasets are explored.

Figures 2, 3 inspection provides evidence that the m=1,

TABLE IV

EQUAL ERROR RATES (%) FOR $\mathcal{F}_{intra}$ (BLOCK DIAGONAL) & $\mathcal{F}_{blind}$ FOR BOTH TYPES OF DISSIMILAR PAIRS, $\mathfrak{W}_3$ MODUS, FOR 5 AND 10 REFERENCE SAMPLES. NUMBER OF PARTICIPATING SEGMENTS EQUAL SEVEN (FOUR FOR CEDAR) AND m=1, p=10.

			Testing Datasets			
Learning Datasets	$\mathbb{T}_{\mathbb{R}-}$ Setup	# refs	CEDAR	MCYT	BANGLA	HINDI
CEDAR	$\omega_{0\%RF}^{\mathbb{T}_{\mathbb{R}-}}$	5/10	0.64/0.38	2.13/0.98	0.39/0.24	1.20/0.76
	$\omega_{100\%RF}^{\mathbb{T}_{\mathbb{R}-}}$	5/10	0.66/0.37	2.04/0.95	0.39/0.24	1.22/0.75
MCYT	$\omega_{0\%RF}^{\mathbb{T}_{\mathbb{R}-}}$	5/10	0.66/0.37	2.18/1.02	0.43/0.27	1.21/0.77
	$\omega_{100\%RF}^{\mathbb{T}_{\mathbb{R}-}}$	5/10	0.66/0.36	2.25/0.96	0.48/0.27	1.24/0.75
BANGLA	$\omega_{0\%RF}^{\mathbb{T}_{\mathbb{R}-}}$	5/10	0.64/0.35	2.07/0.96	0.41/0.25	1.22/0.76
	$\omega_{100\%RF}^{\mathbb{T}_{\mathbb{R}-}}$	5/10	0.63/0.35	2.08/0.96	0.42/0.26	1.22/0.75
HINDI	$\omega_{0\%RF}^{\mathbb{T}_{\mathbb{R}-}}$	5/10	0.66/0.35	2.04/0.92	0.47/0.27	1.27/0.78
	$\omega_{100\%RF}^{\mathbb{T}_{\mathbb{R}-}}$	5/10	0.66/0.35	2.09/0.97	0.47/0.27	1.23/0.77

p=10 values of the proposed model form a robust distance for almost all datasets with some minor alterations in $\mathcal{D}_1$ (m=1, p=7, 8, 9) and $\mathcal{D}_2$ (m=1, p=9) which do not pose a loss of generality. A possible explanation comes from the fact that the dimensionality of the original SPD manifold P_{10} is relatively low which does not allow a complete exploitation of the mapped manifolds $P_{m \cdot p}$. In theory, higher dimensional SPD manifolds can be employed with the use of high dimensional keypoint descriptors, although this is out of the scope of this work. Specific details on the results derived by the blind $\mathcal{F}_{intra}$ and $\mathcal{F}_{inter}$ frameworks on all datasets for the $\mathfrak{W}_3$ modus are now provided in Table V by means of the average EERs.

The use of the m=1, p=10 for presentation purposes is

TABLE V
COMPARATIVE SUMMARY OF THE PROPOSED WI-SV METHOD WITH OTHER SV-WI SYSTEMS (%)

Method & [Ref]	**Metric**	**Additional comments**	**Datasets & (#reference samples) (G: Genuine)**			
			CEDAR	**MCYT**	**BANGLA**	**HINDI**
Graph edit distance (MCS) [47]	local EER_{SF}	*(Protocol similar to $\mathfrak{W}_3$)*	5.91(10G)	3.91(10G)	-	-
Surroundedness [94]	AER_{SF} @EER	*Protocol similar to $\mathfrak{W}_1$*	8.33(1G)	-	-	-
Region Deep Metric (MSDN) [50]	(global EER_{SF})	*(Protocol similar to $\mathfrak{W}_1$)*	6.74(1G)	-	-	-
	local EER_{SF}	*(Protocol similar to $\mathfrak{W}_2$)*	4.83(1G)	-	-	-
	local EER_{SF}	*(Protocol similar to $\mathfrak{W}_3$)*	1.75(10G) 1.67(12G)	-	-	-
Point2Set DML [48]	local EER_{SF}	*(Protocol similar to $\mathfrak{W}_3$)*	5.22(5G)	9.86(5G)	-	-
DCCM & Feat. Diss. Thresh. [39]	local AER_{SF}	*(Protocol similar to $\mathfrak{W}_3$ with specific threshold selection)* *D_1: 10 writers for training*	2.10(5G)	-	-	-
Gradient, Structure & Concavity [91]	global EER_{SF}	*Protocol similar to $\mathfrak{W}_1$*	21.9(16G)			
DMML(HOG) [38]	local EER_{SF}	*(Protocol similar to $\mathfrak{W}_3$)*	-	13.4(5G) 9.86(10G)	-	-
Partially ordered sets [28]	local EER_{SF}	*(Protocol exactly to $\mathfrak{W}_3$)* *Direct comparison*	2.90(5G)	3.50(5G)		
CNN-BiLSTM [29] *(ReLU activation)*	local EER_{SF}	*Train: w. skilled forgeries* *Train w. random forgeries*	0.00(N/A) 0.43(N/A)	-	1.76(N/A) 1.96(N/A)	2.23(N/A) 3.08(N/A)
Signet & Dichotomy [27]	local EER_{SF}	*(Protocol similar to $\mathfrak{W}_3$)*	3.32(12G)	2.89(10G)		
P.C.F & Learn w. Rejection [53]	local EER_{SF}	*Train protocol 8:1:1 for train, validation & testing on the entire dataset. Train with skilled forgeries*	-	-	6.10(1G)	8.80(1G)
CBCapsNet [45]	Local AER_{SF}	*60% of all writers for training, 20% for validation, 20% testing* *Train with skilled forgeries*	0.00(N/A)	-	5.70(N/A)	0.00(N/A)
Metric Learning-VLAD [52]	local AER_{SF} @EER	*D_1: 50 writers for training* *D_3: 80 writers for training* *D_4: 100 writers for training* *Skilled forgeries at training*	0.00(N/A)	-	9.62(N/A)	20.2(N/A)
AVN [2]	local EER_{SF}	*$D_{1,3}$: 50 writers for training* *D_4: 100 writers for training* *Skilled forgeries at training*	3.77(N/A)	-	6.14(N/A)	5.65(N/A)
Deep HSV [51]	local EER_{SF}	*D_1: 50 writers for training* *D_3: 80 writers for training* *D_4: 100 writers for training* *Skilled forgeries at training*	0.00(N/A)	-	11.9(N/A)	13.3(N/A)
IDN [42]	local EER_{SF} or AER_{SF}	*D_1: 50 writers for training* *D_3: 80 writers for training* *D_4: 100 writers for training* *Skilled forgeries at training*	3.62(N/A) @EER	-	4.68(N/A) @ AER	6.96(N/A) @ AER
2C2S [95]	local EER_{SF}	*D_1: 50 writers for training* *D_3: 50 writers for training* *D_4: 100 writers for training* *Skilled forgeries at training*	0.00(N/A)	-	6.75(N/A)	9.32(N/A)
SDINet [43]	local EER_{SF} or AER_{SF}	*D_1: 50 writers for training* *D_3: 50 writers for training* *D_4: 100 writers for training* *Skilled forgeries at training*	1.75(N/A) @EER	-	5.58(N/A) @ AER	5.00(N/A) @ AER
SURDS [44]	local AER_{SF}	*D_3: 70 writers for training* *D_4: 112 writers for training* *Skilled forgeries at training*	-	-	12.6(8G)	10.5(8G)
TransOSV [96]	local EER_{SF}	*D_3: 50 writers for training* *D_3: 50 writers for training* *D_4: 100 writers for training* *Skilled forgeries at training*	-	-	9.90(1G) 3.56(1G)	3.24(1G)
Proposed: $\mathcal{F}_{intra}$	global EER_{SF}	$\mathfrak{W}_1$, m=1, p=10, 100%RF	8.53(1G)	17.4(1G)	11.1(1G)	15.6(1G)
Proposed: $\mathcal{F}_{intra}$	global EER_{SF}	$\mathfrak{W}_1$, m=1, p=10, 0%RF	8.66(1G)	17.6(1G)	11.3(1G)	15.8(1G)
Proposed: $\mathcal{F}_{intra}$	local EER_{SF}	$\mathfrak{W}_2$, m=1, p=10, 100%RF	4.59(1G)	11.0(1G)	6.39(1G)	9.43(1G)
Proposed: $\mathcal{F}_{intra}$	(local EER_{SF})	$\mathfrak{W}_2$,m=1, p=10, 0%RF	4.65(1G)	10.8(1G)	6.54(1G)	9.63(1G)
Proposed: $\mathcal{F}_{intra}$ -$\mathcal{F}_{blind}$	local EER_{SF}	$\mathfrak{W}_3$, m=1, p=10, 100%RF *seg_idx* =4 ($\mathcal{D}_1$) or 7 ($\mathcal{D}_{2-4}$)	0.63 – 0.66 (5G)	2.04 – 2.25 (5G)	0.39 – 0.48 (5G)	1.22 – 1.24 (5G)
Proposed: $\mathcal{F}_{intra}$ -$\mathcal{F}_{blind}$	local EER_{SF}	$\mathfrak{W}_3$, m=1, p=10, 0%RF *seg_idx* =4 ($\mathcal{D}_1$) or 7 ($\mathcal{D}_{2-4}$)	0.64 - 0.66 (5G)	2.04 – 2.18 (5G)	0.39 - 0.47 (5G)	1.20 – 1.27 (5G)
Proposed: $\mathcal{F}_{intra}$ -$\mathcal{F}_{blind}$	local EER_{SF}	$\mathfrak{W}_3$, m=1, p=10, 100%RF *seg_idx* =4 ($\mathcal{D}_1$) or 7 ($\mathcal{D}_{2-4}$)	0.35 - 0.37 (10G)	0.95 - 0.97 (10G)	0.24 - 0.27 (10G)	0.75 - 0.77 (10G)
Proposed: $\mathcal{F}_{intra}$ -$\mathcal{F}_{blind}$	local EER_{SF}	$\mathfrak{W}_3$, m=1, p=10, 0%RF *seg_idx* =4 ($\mathcal{D}_1$) or 7 ($\mathcal{D}_{2-4}$)	0.35 - 0.38 (10G)	0.92 - 1.02 (10G)	0.24 -0.27 (10G)	0.76 - 0.78 (10G)

justified in the above paragraph. In addition we report the derived results when the number of participating segments has been set to four for the $\mathcal{D}_1$ dataset with fewer pixels and seven for the remaining $\mathcal{D}_{2-4}$ datasets. While it is a fact that there are other values of the m, p parameters which provide lower error rates, we clearly observe that we obtain robust results for all learning and testing datasets for the specific case presented in Figures 4, 5 and Table IV.

Table V provide, to the best of our abilities, a summary of state-of-the-art (SOTA) results for WI-SV systems reported on the four signature datasets by means of the average error rate (AER_{SF}) or the global (and/or local) versions of the equal error rate (EER_{SF}). The reported EER_{SF} denotes the equal error rate between the FRR - false rejection Rate i.e. similar pairs between similar signatures being classified as dissimilar (Genuine to Skilled-or-simulated forgery) ones and FAR_{SF} - false acceptance rate i.e. dissimilar pairs between dissimilar signatures (Genuine to Skilled-or-simulated forgery) being classified as similar ones. The AER_{SF} can be related a) to the accuracy (i.e. AER(%)=100%-Accuracy(%)) or b) the average of specific values of FRR and FAR_{SF}.

TABLE VI

COMPARATIVE SUMMARY OF THE PROPOSED WI-SV METHOD ON THE $\mathfrak{W}_3$ MODUS WITH OTHER SV-WD SYSTEMS (%)

Method & [Ref]	Local Metric	Datasets & (#reference samples)			
		CEDAR	MCYT	BANGLA	HINDI
SigNet/F/SPP [23]	EER_{SF}	4.63(12G)	2.87(10G)	-	-
		5.87(4G)	3.58(10G)	-	-
VLAD-KAZE [97]	EER_{SF}	1.0(N/A)	6.40(N/A)	-	-
FV&KAZE [98]	EER_{SF}	-	5.47(N/A)	-	-
K-SVD/OMP [22]	EER_{SF}	0.79(10G)	1.37(10G)	1.05(10G)	0.44(10G)
		1.80(5G)	2.82(5G)	-	-
Hybrid Texture [99]	AER_{SF}	9.50(4G)	19.1(4G)	20.1(12G) for BHsig260	
		5.50(12G)	9.26(10G)		
RNN's [21]	EER_{SF}	0.01(12G)	0.34(10G)	0.43(12G)	0.36(12G)
Deformations [100]	EER_{SF}	3.89(12G)	-	9.01(8G)	8.21(8G)
Textr. features [101]	AER_{SF}	-	-	24.4(8G)	33.8(8G)
CNN-CoLL [102]	EER_{SF}	2.03(5G)	2.61(5G)	-	-
		1.66(10G)	1.62(10G)	-	-
Vis. Graphs [25]	EER_{SF}	0.51(10G)		1.02(10G)	0.32(10G)
Cycle GAN [103]	EER_{SF}	4.50(5G)	3.42(5G)	-	-
		3.48(10G)	2.01(10G)	-	-
SPD tang. plane [35]	EER_{SF}	0.49(10G)	-	1.00(10G)	0.27(10G)
TransOSV - SVM [96] WD protocol	EER_{SF}	-	-	5.59(6G)	8.56(6G)
		-	-	4.77(10G)	7.33(10G)
MGnet [104]	EER_{SF}	-	7.00(N/A)	3.5(N/A)	4.5(N/A)
LDerivP - SVM [105]	EER_{SF}	-	11.90(5G)	-	-
Graph ed. dist. & Inkball Mdls [106]	EER_{SF}	-	9.07 (5G) 5.78(10G)	-	-
Par. opt. [24] (20 duplicates / genuine)	EER_{SF}	0.82(3G)	0.01(3G)	-	-
Hybrid CNN & HOG (LSTM) [107]	EER_{SF}	12.5(N/A)	-	-	-
MAML, OC [108]	EER_{SF}	8.27(4G) 7.07(8G)	12.8(5G) 12.4(10G)	-	-
Proposed $\mathfrak{W}_3$, m=1, p=10, 100%RF *seg_idx* =4 ($\mathcal{D}_1$) or 7 ($\mathcal{D}_{2-4}$)	EER_{SF}	0.63-0.66 (5G)	2.04-2.25 (5G)	0.39-0.48 (5G)	1.22-1.24 (5G)
Proposed $\mathfrak{W}_3$, m=1, p=10, 100%RF *seg_idx* =4 ($\mathcal{D}_1$) or 7 ($\mathcal{D}_{2-4}$)	EER_{SF}	0.35-0.37 (10G)	0.95-0.97 (10G)	0.24-0.27 (10G)	0.75-0.77 (10G)

The metric "AER_{SF}" is usually met when the decision threshold is considered a-priori known. Thus, we consider the FAR_{SF} and FRR to be known while the AER_{SF} represents their average value. Sometimes also the reported results are provided by means of the AER, computed at the EER threshold (AER_{SF} @EER). The metric EER (global) has been used for the $\mathfrak{W}_1$ protocol (only 1 reference sample) in which there are only two classes: similar vs dissimilar pairs, irrespective of the origin of the writer. The metric EER_{SF} (local) has been used for the $\mathfrak{W}_2$ or $\mathfrak{W}_3$ protocols.

The contents of Table V clearly indicate that the proposed SPD learnable distance framework operates efficiently in the challenging WI-SV oriented framework. Finally, table VI summarizes the reported results for a number of WD-SV systems found in the literature, along with the proposed SPD method. Its inspection shows that, the proposed SPD-WI framework delivers low error rates even when it is viewed with respect to WD-SV SOTA systems.

VI. CONCLUSION

In this work, offline writer-independent signature verification was addressed by learning a robust similarity distance between pairs of similar and dissimilar signature images and corresponding low dimensional covariance matrices. The similarity measure is comprised of three learnable parts namely, a manifold to manifold projection $\boldsymbol{W}$, a family of alpha-beta divergences $\boldsymbol{A}$ and an integration function $\boldsymbol{M}$. The experiments were conducted with four popular datasets, in blind intra and inter frameworks and the verification errors reported clearly indicate that the proposed similarity distance is effective and worthy of investigation. Our future research agenda includes among others, a comparative analysis with other Riemannian network architectures designed for SPD matrix learning, something that now falls beyond the scope of our current work.

Perhaps the weakest point (i.e. a limitation) in our WI-SPD context is the fact that this descriptor is not generative in the sense that one can compute the Riemannian mean of two covariance matrices but this will not allow us to generate the "mean signature image". Future research will focus towards the design of a WI-SV system with synthetic handwritten signature images, by means of popular duplicators or generative attacking methods, in order to assist a truly agnostic signature verifier in the SPD domain.

ACKNOWLEDGMENT

Part of the over-length page charge will be funded by the Special Account for Research Funds (SARF) of the University of West Attica.

REFERENCES

[1] S. Lai, L. Jin, Y. Zhu, Z. Li, and L. Lin, "SynSig2Vec: Forgery-Free Learning of Dynamic Signature Representations by Sigma Lognormal-Based Synthesis and 1D CNN," *IEEE TPAMI*, vol. 44, pp. 6472-6485, 2022.

[2] H. Li, P. Wei, and P. Hu, "AVN: An Adversarial Variation Network Model for Handwritten Signature Verification," *IEEE Transactions on Multimedia*, vol. 24, pp. 594 - 608, 2021.

[3] M. Deviterne-Lapeyre and S. Ibrahim, "Interpol questioned documents review 2019–2022," *Forensic Science International: Synergy,* vol. 6, 100300, 2023.

[4] Reuters (P. Dave & A. Sullivan). 2020.9.24. *Factbox: U.S. counties using automated signature verification software*. [Online]. Available: Link.

[5] Cybersecurity and Infrastructure Security Agency (CISA). *Signature Verification and Cure Process*, [Online] Available: Link.

[6] Wacom. *Signature Verification Real-time fraud prevention*. [Online]. Available: Link.

[7] Parascript strategic white paper. *How Automated Signature Verification Supports Efficient, reliable vote-by-mail programs.* [Online]. Available: Link.

[8] S. N. Srihari, S.-H. Cha, H. Arora, and S. Lee, "Individuality of handwriting," *Journal of forensic science,* vol. 47, pp. 1-17, 2002.

[9] H.-L.Teulings, "Handwriting Movement Control", in *Handbook of Perception and Action*, vol. 2, pp. 561-613, London: UK, Ac. Press, 1996.

[10] M. Faundez-Zanuy, J. Fierrez, M. A. Ferrer, M. Diaz, R. Tolosana, and R. Plamondon, "Handwriting Biometrics: Applications and Future Trends in e-Security & e-Health," *Cognitive Computation,* vol. 12, pp. 940-953, 2020.

[11] N. Sae-Bae, N. Memon, and P. Sooraksa, "Distinctiveness, complexity, and repeatability of online signature templates," *Pattern Recognition,* vol. 84, pp. 332-344, 2018.

[12] D. Impedovo and G. Pirlo, "Automatic Signature Verification: The State of the Art," *IEEE Transactions on Systems, Man and Cybernetics, Part C: Applications and Reviews,* vol. 38, pp. 609-635, 2008.

[13] Moises Diaz, Miguel A. Ferrer, Donato Impedovo, Muhammad Imran Malik, Giuseppe Pirlo, and R. Plamondon., "A Perspective Analysis of Handwritten Signature Technology," *ACM Computing Surveys,* vol. 51, no.6, 2019, Art. no. 117 doi: 10.1145/3274658.

[14] M. M. Hameed, R. Ahmad, M. L. M. Kiah, and G. Murtaza, "Machine learning-based offline signature verification systems: A systematic review," *Signal Processing: Image Communication,* vol. 93, Art. no. 116139, 2021.

[15] N. Sae-Bae and N. Memon, "Online Signature Verification on Mobile Devices," *IEEE Transactions on Information Forensics and Security,* vol. 9, pp. 933-947, 2014.

[16] V. Venugopal and S. Sundaram, "An improved online writer identification framework using codebook descriptors," *Pattern Recognition,* vol. 78, pp. 318-330, 2018.

[17] M. Okawa, "Online signature verification using single-template matching with time-series averaging and gradient boosting," *Pattern Recognition,* vol. 102, Art. no. 107227, 2020.

[18] M. Zalasiński, K. Łapa, and K. Cpałka, "Prediction of values of the dynamic signature features," *ESWA,* vol. 104, pp. 86-96, 2018.

[19] J. Galbally, M. Diaz-Cabrera, M. A. Ferrer, M. Gomez-Barrero, A. Morales, and J. Fierrez, "On-line signature recognition through the combination of real dynamic data and synthetically generated static data," *Pattern Recognition,* vol. 48, pp. 2921-2934, 2015.

[20] X. Xia, Z. Chen, F. Luan, and X. Song, "Signature alignment based on GMM for on-line signature verification," *Pattern Recognition,* vol. 65, pp. 188-196, 2017.

[21] R. Ghosh, "A Recurrent Neural Network based deep learning model for offline signature verification and recognition system," *ESWA,* vol. 168, p. 114249, 2021.

[22] E. N. Zois, D. Tsourounis, I. Theodorakopoulos, A. L. Kesidis, and G. Economou, "A Comprehensive Study of Sparse Representation Techniques for Offline Signature Verification," *IEEE TBIOM,* vol. 1, pp. 68-81, 2019.

[23] L. G. Hafemann, R. Sabourin, and L. S. Oliveira, "Learning features for offline handwritten signature verification using deep convolutional neural networks," *Pattern Recognition,* vol. 70, pp. 163-176, 2017.

[24] T. M. Maruyama, L. S. Oliveira, A. S. Britto, and R. Sabourin, "Intrapersonal Parameter Optimization for Offline Handwritten Signature Augmentation," *IEEE TIFS,* vol. 16, pp. 1335-1350, 2021.

[25] E. N. Zois, E. Zervas, D. Tsourounis, and G. Economou, "Sequential Motif Profiles and Topological Plots for Offline Signature Verification," in *CVPR*, 2020, pp. 13245-13255.

[26] W. Bouamra, C. Djeddi, B. Nini, M. Diaz, and I. Siddiqi, "Towards the design of an offline signature verifier based on a small number of genuine samples for training," *ESWA,* vol. 107, pp. 182-195, 2018.

[27] V. L. F. Souza, A. L. I. Oliveira, R. M. O. Cruz, and R. Sabourin, "A white-box analysis on the writer-independent dichotomy transformation applied to offline handwritten signature verification," *ESWA,* vol. 154, p. 113397, 2020 2020.

[28] E. N. Zois, A. Alexandridis, and G. Economou, "Writer independent offline signature verification based on asymmetric pixel relations and unrelated training-testing datasets," *ESWA,* vol. 125, pp. 14-32, 2019.

[29] T. Longjam, D. R. Kisku, and P. Gupta, "Writer independent handwritten signature verification on multi-scripted signatures using hybrid CNN-BiLSTM: A novel approach," *ESWA,* vol. 214, p. 119111, 2023.

[30] S. N. Srihari, A. Xu, and M. K. Kalera, "Learning strategies and classification methods for off-line signature verification," in *9th IWFHR,* Tokyo, Japan, 2004, pp. 161-166.

[31] D. Rivard, E. Granger, and R. Sabourin, "Multi-feature extraction and selection in writer-independent off-line signature verification" *IJDAR,* vol. 16, pp. 83-103, 2013.

[32] G. S. Eskander, R. Sabourin, and E. Granger, "Hybrid writer-independent writer-dependent offline signature verification system," *IET Biometrics,* vol. 2, pp. 169-181, 2013.

[33] M. A. Ferrer, M. Diaz-Cabrera, and A. Morales, "Static Signature Synthesis: A Neuromotor Inspired Approach for Biometrics," *IEEE TPAMI,* vol. 37, pp. 667-680, 2015.

[34] M. Diaz, M. A. Ferrer, G. S. Eskander, and R. Sabourin, "Generation of Duplicated Off-line Signature Images for Verification Systems," *IEEE TPAMI* vol. 39, no.5, pp. 951-964, 2016.

[35] E. N. Zois, S. Said, D. Tsourounis, and A. Alexandridis, "Subscripto multiplex: A Riemannian symmetric positive definite strategy for offline signature verification," *Pattern Recognition Letters,* vol. 167, pp. 67-74, 2023.

[36] N. Arab, H. Nemmour, and Y. Chibani, "A new synthetic feature generation scheme based on artificial immune systems for robust offline signature verification," *ESWA,* vol. 213, Art. no. 119306, 2023.

[37] D. Bertolini, L. S. Oliveira, E. Justino, and R. Sabourin, "Reducing forgeries in writer-independent off-line signature verification through ensemble of classifiers," *Pattern Recognition,* vol. 43, pp. 387-396, 2010.

[38] A. Soleimani, B. N. Araabi, and K. Fouladi, "Deep Multitask Metric Learning for Offline Signature Verification," *Pattern Recognition Letters,* vol. 80, pp. 84-90, 2016.

[39] A. Hamadene and Y. Chibani, "One-Class Writer-Independent Offline Signature Verification Using Feature Dissimilarity Thresholding," *IEEE TIFS,* vol. 11, pp. 1226-1238, 2016.

[40] E. Pękalska and R. P. W. Duin, "Dissimilarity representations allow for building good classifiers," *Pattern Recognition Letters,* vol. 23, pp. 943-956, 2002.

[41] Y.-J. Xiong and S.-Y. Cheng, "Attention Based Multiple Siamese Network for Offline Signature Verification," in *ICDAR,* Lausanne, Switzerland, 2021, pp. 337-349.

[42] P. Wei, H. Li, and P. Hu, "Inverse Discriminative Networks for Handwritten Signature Verification," in *CVPR,* Long Beach, 2019, pp. 5764-5772.

[43] H. Li, P. Wei, and P. Hu, "Static-Dynamic Interaction Networks for Offline Signature Verification," in *AAAI Conf. on Artificial Intelligence,* 2021, pp. 1893-1901.

[44] S. Chattopadhyay, S. Manna, S. Bhattacharya, and U. Pal, "SURDS: Self-Supervised Attention-guided Reconstruction and Dual Triplet Loss for Writer Independent Offline Signature Verification," in *ICPR,* Montréal, Canada, 2022, pp. 1600-1606.

[45] E. Parcham, M. Ilbeygi, and M. Amini, "CBCapsNet: A novel writer-independent offline signature verification model using a CNN-based architecture and capsule neural networks," *ESWA,* vol. 185, Art. no. 115649, 2021.

[46] H. Rantzsch, H. Yang, and C. Meinel, "Signature Embedding: Writer Independent Offline Signature Verification with Deep Metric Learning," in *ISVC,* Las Vegas, USA, 2016, doi: 10.1007/978-3-319-50832-0_60.

[47] P. Maergner, V. Pondenkandath, M. Alberti, M. Liwicki, K. Riesen, R. Ingold, and A. Fischer, "Combining graph edit distance and triplet networks for offline signature verification," *Pattern Recognition Letters,* vol. 125, pp. 527-533, 2019.

[48] Y. Zhu, S. Lai, Z. Li, and L. Jin, "Point-to-Set Similarity Based Deep Metric Learning for Offline Signature Verification," in *ICFHR,* Dortmund, Germany, 2020, pp. 282-287.

[49] S. Lai and L. Jin, "Learning Discriminative Feature Hierarchies for Off-Line Signature Verification," in *ICFHR,* 2018, Niagara Falls, USA pp. 175-180.

[50] L. Liu, L. Huang, F. Yin, and Y. Chen, "Offline signature verification using a region based deep metric learning network," *Pattern Recognition*, vol. 118, Art. no. 108009, 2021.

[51] F. Lin, C. Li, Z. Wang, G. Yu, L. Yuan, and H. Wang, "DeepHSV: User-Independent Offline Signature Verification Using Two-Channel CNN," in *ICDAR*, Sydney, Australia, 2019, pp. 166-171.

[52] M. S. Hanif and M. Bilal, "A Metric Learning Approach for Offline Writer Independent Signature Verification," *Pattern Recognition and Image Analysis*, vol. 30, pp. 795-804, 2020.

[53] X. ji, D. Suehiro, and S. Uchida, "Paired Contrastive Feature for Highly Reliable Offline Signature Verification," *Pattern Recognition*, 109816, 2023.

[54] D. S. Guru, K. S. Manjunatha, S. Manjunath, and M. T. Somashekara, "Interval valued symbolic representation of writer dependent features for online signature verification," *ESWA*, vol. 80, pp. 232-243, 2017.

[55] O. Tuzel, F. Porikli, and P. Meer, "Region Covariance: A Fast Descriptor for Detection and Classification," in *ECCV 2006*, Berlin, Heidelberg, Germany, 2006, pp. 589-600.

[56] P. Koniusz, F. Yan, P. Gosselin, and K. Mikolajczyk, "Higher-Order Occurrence Pooling for Bags-of-Words: Visual Concept Detection," *IEEE TPAMI*, vol. 39, pp. 313-326, 2017.

[57] Z. Gao, Y. Wu, M. Harandi, and Y. Jia, "A Robust Distance Measure for Similarity-Based Classification on the SPD Manifold," *IEEE TNNLS*, vol. 31, pp. 3230-3244, 2020.

[58] Z. Gao, J. Xie, Q. Wang, and P. Li, "Global Second-Order Pooling Convolutional Networks," in *CVPR*, Long Beach, 2019, pp. 3019-3028.

[59] Q. Wang, P. Li, Q. Hu, P. Zhu, and W. Zuo, "Deep Global Generalized Gaussian Networks," in *CVPR*, Long Beach, 2019, pp. 5075-5083.

[60] P. Koniusz, L. Wang, and A. Cherian, "Tensor Representations for Action Recognition," *IEEE TPAMI*, Vol. 44, no.2, pp. 648 - 665, 2021.

[61] Y. Zhang, S. Tang, K. Muandet, C. Jarvers, and H. Neumann, "Local Temporal Bilinear Pooling for Fine-Grained Action Parsing," in *CVPR*, Long Beach, 2019, pp. 11997-12007.

[62] G. Wang, S. Yang, H. Liu, Z. Wang, Y. Yang, S. Wang, Y. Yang, S. Wang, G. Yu, E. Zhou and J. Sun, "High-Order Information Matters: Learning Relation and Topology for Occluded Person Re-Identification," in *CVPR*, 2020, pp. 6448-6457.

[63] B. Bryan, Y. Gong, Y. Zhang, and C. Poellabauer, "Second-Order Non-Local Attention Networks for Person Re-Identification," in *ICCV*, Seoul, Korea, 2019, pp. 3759-3768.

[64] B. Chen, W. Deng, and J. Hu, "Mixed High-Order Attention Network for Person Re-Identification," in *ICCV*, Seoul, Korea, 2019, pp. 371-381.

[65] G. Maman, O. Yair, D. Eytan, and R. Talmon, "Domain Adaptation Using Riemannian Geometry of Spd Matrices," in *ICASSP*, Brighton, UK, 2019, pp. 4464-4468.

[66] P. Koniusz and H. Zhang, "Power Normalizations in Fine-grained Image, Few-shot Image and Graph Classification," *IEEE TPAMI*,vol. 44, no.2, pp. 591-609, 2021.

[67] Q.Luo, Y. Meng, L. Liu, X. Zhao, and Z. Zhou, "Cloud classification of ground-based infrared images combining manifold and texture features," *Atmos. Meas. Tech.*, vol. 11, pp. 5351-5361, 2018.

[68] X. Pennec, P. Fillard, and N. Ayache, "A Riemannian Framework for Tensor Computing," *Int. J. Computer Vision*, vol. 66, pp. 41–66, 2006.

[69] A. Barachant, S. Bonnet, M. Congedo, and C. Jutten, "Multiclass Brain–Computer Interface Classification by Riemannian Geometry," *IEEE Transactions on Biomedical Engineering*, vol. 59, pp. 920-928, 2012.

[70] P. Zanini, M. Congedo, C. Jutten, S. Said, and Y. Berthoumieu, "Transfer Learning: A Riemannian Geometry Framework With Applications to Brain–Computer Interfaces," *IEEE Transactions on Biomedical Engineering*, vol. 65, pp. 1107-1116, 2018.

[71] Z. Dong, S. Jia, C. Zhang, M. Pei, and Y. Wu, "Deep Manifold Learning of Symmetric Positive Definite Matrices with Application to Face Recognition," in *AAAI Conference on Artificial Intelligence, San Francisco, USA*, 2017, doi: 10.1609/aaai.v31i1.11232.

[72] Z. Huang and L. Van Gool, "A Riemannian Network for SPD Matrix Learning," in *AAAI Conference on Artificial Intelligence*, San Francisco, USA, 2017, doi: 10.1609/aaai.v31i1.10866.

[73] P. Li, Q. Wang, W. Zuo, and L. Zhang, "Log-Euclidean Kernels for Sparse Representation and Dictionary Learning," in *ICCV*, Sydney, Australia, 2013, pp. 1601-1608.

[74] R. Vemulapalli, J. K. Pillai, and R. Chellappa, "Kernel Learning for Extrinsic Classification of Manifold Features," in *CVPR*, Portland, USA 2013, pp. 1782-1789.

[75] 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 *ECCV*, Florence, Italy, 2012, pp. 216-229.

[76] V. Arsigny, P. Fillard, X. Pennec, and N. Ayache, "Log-Euclidean metrics for fast and simple calculus on diffusion tensors," *Magnetic Resonance in Medicine*, vol. 56, pp. 411-421, 2006.

[77] S. Sra, "A new metric on the manifold of kernel matrices with application to matrix geometric means," in *NIPS* Lake Tahoe, USA, 2012.

[78] B. Kulis, M. Sustik, and I. Dhillon, "Learning low-rank kernel matrices," in *ICML*, Pittsburgh, USA, 2006, pp. 505-512.

[79] A. Cichocki, S. Cruces, and S.-i. Amari, "Log-Determinant Divergences Revisited: Alpha-Beta and Gamma Log-Det Divergences," *Entropy*, vol. 17, pp. 2988-3034, 2015.

[80] D. B. Thiyam, S. Cruces, J. Olias, and A. Cichocki, "Optimization of Alpha-Beta Log-Det Divergences and their Application in the Spatial Filtering of Two Class Motor Imagery Movements," *Entropy*, vol. 19, p. 89, 2017.

[81] M. Harandi, M. Salzmann, and R. Hartley, "Dimensionality Reduction on SPD Manifolds: The Emergence of Geometry-Aware Methods," *IEEE TPAMI*, vol. 40, no.1, pp. 48-62, 2018.

[82] L. Zhou, L. Wang, J. Zhang, Y. Shi, and Y. Gao, "Revisiting Metric Learning for SPD Matrix Based Visual Representation," in *CVPR*, Honolulu, USA, 2017, pp. 7111-7119.

[83] H. Q. Minh, M. S. Biagio, and V. Murino, "Log-Hilbert-Schmidt metric between positive definite operators on Hilbert spaces," in *NIPS* Montreal, Canada, 2014.

[84] Y. Chen, E. K. Garcia, M. R. Gupta, A. Rahimi, and L. Cazzanti, "Similarity-based Classification: Concepts and Algorithms," *J. Mach. Learn. Research*, vol. 10, pp. 747–776, 2009.

[85] R. Hadsell, S. Chopra, and Y. LeCun, "Dimensionality Reduction by Learning an Invariant Mapping," in *CVPR*, New York, USA, 2006, pp. 1735-1742.

[86] S. Said, L. Bombrun, Y. Berthoumieu, and J. H. Manton, "Riemannian Gaussian Distributions on the Space of Symmetric Positive Definite Matrices," *IEEE Transactions on Information Theory*, vol. 63, pp. 2153-2170, 2017.

[87] A. Cichocki and S.-I. Amari, "Families of Alpha- Beta- and Gamma-Divergences: Flexible and Robust Measures of Similarities," *Entropy*, vol. 12, pp. 1532-1568, 2010.

[88] A. Edelman, T. A. Arias, and S. T. Smith, "The geometry of algorithms with orthogonality constraints," *SIAM journal on Matrix Analysis and Applications*, vol. 20, pp. 303-353, 1998.

[89] P.-A. Absil, R. Mahony, and R. Sepulchre, *Optimization Algorithms on Matrix Manifolds*, Princeton USA, Princeton University Press, 2008.

[90] S. Bonnabel, "Stochastic Gradient Descent on Riemannian Manifolds," *IEEE Transactions on Automatic Control*, vol. 58, pp. 2217-2229, 2013.

[91] M. K. Kalera, S. Srihari, and A. Xu, "Offline signature verification and identification using distance statistics," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 18, pp. 1339-1360, 2004.

[92] J. F. Vargas, M. A. Ferrer, C. M. Travieso, and J. B. Alonso, "Off-line signature verification based on grey level information using texture features," *Pattern Recognition*, vol. 44, pp. 375-385, 2011.

[93] A. Alaei, S. Pal, U. Pal, and M. Blumenstein, "An Efficient Signature Verification Method Based on an Interval Symbolic Representation and a Fuzzy Similarity Measure," *IEEE TIFS*, vol. 12, no. 10, pp. 2360-2372, 2017.

[94] R. Kumar, J. D. Sharma, and B. Chanda, "Writer-independent off-line signature verification using surroundedness feature," *Pattern Recognition Letters*, vol. 33, pp. 301-308, 2012.

[95] J.-X. Ren, Y.-J. Xiong, H. Zhan, and B. Huang, "2C2S: A two-channel and two-stream transformer based framework for offline signature verification," *Engineering Applications of Artificial Intelligence*, vol. 118, p. 105639, 2023.

[96] H. Li, P. Wei, Z. Ma, C. Li, and N. Zheng, "TransOSV: Offline Signature Verification with Transformers," *Pattern Recognition*, vol. 145, p. 109882, 2024.

[97] M. Okawa, "From BoVW to VLAD with KAZE features: Offline signature verification considering cognitive processes of forensic experts," *Pattern Recognition Letters*, vol. 113, pp. 75-82, 2018.

[98] M. Okawa, "Synergy of foreground–background images for feature extraction: Offline signature verification using Fisher vector with fused KAZE features" *Pattern Recognition*, vol. 79, pp. 480-489, 2018.

[99] A. K. Bhunia, A. Alaei, and P. P. Roy, "Signature verification approach using fusion of hybrid texture features," *Neural Computing and Applications*, vol. 31, pp. 8737-8748, 2019.
[100] Y. Zheng, B. K. Iwana, M. I. Malik, S. Ahmed, W. Ohyama, and S. Uchida, "Learning the Micro Deformations by Max-pooling for Offline Signature Verification," *Pattern Recognition*, p. 108008, 2021.
[101] S. Pal, A. Alaei, U. Pal, and M. Blumenstein, "Performance of an Off-Line Signature Verification Method Based on Texture Features on a Large Indic-Script Signature Dataset," in DAS, Santorini, Greece, 2016, pp. 72-77.
[102] D. Tsourounis, I. Theodorakopoulos, E. N. Zois, and G. Economou, "From text to signatures: Knowledge transfer for efficient deep feature learning in offline signature verification," *Expert Systems with Applications*, vol. 189, p. 116136, 2022.
[103] J. Jiang, S. Lai, L. Jin, Y. Zhu, J. Zhang, and B. Chen, "Forgery-free signature verification with stroke-aware cycle-consistent generative adversarial network," *Neurocomputing*, vol. 507, pp. 345-357, 2022.
[104] A. Jain, S. K. Singh, and K. Pratap Singh, "Multi-task learning using GNet features and SVM classifier for signature identification," *IET Biometrics*, vol. 10, pp. 117-126, 2021.
[105] M. Diaz, M. A. Ferrer, G. S. Eskander, and R. Sabourin, "Generation of Duplicated Off-Line Signature Images for Verification Systems," *IEEE TPAMI*, vol. 39, pp. 951-964, 2017.
[106] P. Maergner, N. Howe, K. Riesen, R. Ingold, and A. Fischer, "Offline Signature Verification Via Structural Methods: Graph Edit Distance and Inkball Models," in ICFHR, Niagara Falls, USA, 2018, pp. 163-168.
[107] F. M. Alsuhimat and F. S. Mohamad, "A Hybrid Method of Feature Extraction for Signatures Verification Using CNN and HOG a Multi-Classification Approach," *IEEE Access*, vol. 11, pp. 21873-21882, 2023.
[108] L. G. Hafemann, R. Sabourin, and L. S. Oliveira, "Meta-Learning for Fast Classifier Adaptation to New Users of Signature Verification Systems," *IEEE TIFS*, vol. 15, pp. 1735-1745, 2020.

Elias N. Zois, received his B.Sc. in Physics (1994), his M.Sc. degree in Electronic Engineering (1996) along with his Ph.D. (2000) all from the University of Patras (UoP), Greece. Currently he is an Associate Professor at the University of West Attica. His research interests include among other, computer vision, image processing, machine learning, and biometrics.

Dimitrios Tsourounis, received his B.Sc. in Physics (2015) and M.Sc. (2017) in Electronics and Information Processing from University of Patras (UoP), Greece. He is now a Ph.D. candidate in Machine Learning at the Physics Department of (UoP). His research interests include Machine Learning, Pattern Recognition, Image Processing, Computer Vision and Biometrics.

Dimitrios Kalivas received the BS Degree in Electrical Engineering from the National Technical University of Athens in 1982, the MS Degree in Electrical Engineering from the North Carolina State University in 1983, and the Ph.D. Degree in Electrical Engineering from the University of Southern California in 1989. He is a professor in the Department of Electrical and Electronic Engineering at the University of West Attica. His current research interests are in Digital Image Processing and Analysis, Digital Video Processing and Compression, Design and Analysis of Communication Systems, Source Coding and Data Compression.