IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 24 April 2024, accepted 12 June 2024, date of publication 20 June 2024, date of current version 27 June 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3417301

RESEARCH ARTICLE

Group Membership Verification via Nonlinear Sparsifying Transform Learning

BEHROOZ RAZEGHI[1,7], (Senior Member, IEEE), MARZIEH GHEISARI[2], AMIR ATASHIN[3], DIMCHE KOSTADINOV[4], SÉBASTIEN MARCEL[1,5], (Senior Member, IEEE), DENIZ GÜNDÜZ[6], (Fellow, IEEE), AND SLAVA VOLOSHYNOVSKIY[7], (Senior Member, IEEE)

[1]Idiap Research Institute, 1920 Martigny, Switzerland
[2]École Normale Supérieure, 75005 Paris, France
[3]Eindhoven University of Technology, 5612 AZ Eindhoven, The Netherlands
[4]Sony Advanced Visual Sensing AG, 8952 Schlieren, Switzerland
[5]School of Criminal Sciences of the Faculty of Law, University of Lausanne, 1015 Lausanne, Switzerland
[6]Imperial College London, SW7 2AZ London, U.K.
[7]Department of Computer Science, University of Geneva, 1205 Geneva, Switzerland

Corresponding authors: Behrooz Razeghi (behrooz.razeghi@idiap.ch) and Slava Voloshynovskiy (svolos@unige.ch)

This work was supported in part by the University of Geneva and in part by the Swiss Center for Biometrics Research and Testing.

ABSTRACT In today's digitally interconnected landscape, confirming the genuine associations between entities—whether they are items, devices, or individuals—and specific groups is critical. This paper introduces a new group membership verification method while ensuring minimal information loss, coupled with privacy-preservation and discrimination priors. Instead of verifying based on a similarity score in the original data space, we use a min-max functional measure in a transformed space. This method comprises two stages: (i) generating candidate nonlinear transform representations, and (ii) evaluating the min-max measure over these representations for both group assignment and transform selection. We simultaneously determine group membership and pick the appropriate representation from the candidate set based on the evaluation score. To solve within this framework, we employ an iterative alternating algorithm that both learns the parameters of candidate transforms and assigns group membership. Our method's efficacy is assessed on public datasets across various verification and identification scenarios and further tested on real-world image databases, CFP and LFW.

INDEX TERMS Group testing, transform learning, discrimination, identification, verification.

I. INTRODUCTION

Group testing, initially conceived during World War II to efficiently identify rare defects in large populations, has evolved into a fundamental principle across a spectrum of fields, from medical diagnostics to contemporary computer science applications. This approach, which pools and assesses samples concurrently, maximizes resource efficiency and identifies specific attributes in subsets of populations. As today is systems—be it cloud infrastructures or the Internet of Things (IoT)—deepen their interconnectedness, there's a burgeoning demand for robust tools that can authenticate entity affiliations within designated groups. Such verifications not only aid in identification but also ensure authentication in our interconnected digital ecosystems. Yet, despite the advances, contemporary systems present intricate challenges, underscoring a notable gap: the integration of traditional group testing techniques with entity affiliation verifications. Addressing this gap is paramount. It promises not just enhanced membership verification, especially using facial images, but also fortifies the twin pillars of privacy and security in digital realms.

Verification Principles and Challenges: Group membership verification, at its core, seeks to confirm the legitimacy of a specific item, device, or individual within a designated group. This procedure unfolds in two pivotal

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

stages: firstly, the *verification* of the entity's affiliation to the group, and secondly, the *identification* of the entity itself. A paramount challenge in this domain is striking a delicate balance: ensuring accurate distinction between members and non-members while simultaneously preserving individual identities to uphold privacy.

Verification Mechanism: At the core of the verification process lies the collection of templates. These encompass representations of items (via passive physical unclonable functions (PUFs)), devices (through active PUFs), and unique individual traits (captured by biometrics). These templates are securely stored within a server-side data structure. Upon a verification request, the client transmits a distinct signature to the server, which subsequently determines access authorization. It is essential that this mechanism preserves utmost anonymity to shield individual identities. The primary goal of this protocol is to enable the server to verify group membership with accuracy, without having detailed insight into the system's internal mechanisms.

Operational Model: In our operational framework, a central server manages the group membership protocol, processing requests from various clients. When a client acquires a new template, it communicates with the server. It is essential to recognize the potential risks posed by such servers. While some might operate with adversarial intentions, our model assumes the server follows an ''honest but curious'' behavior. This implies that although the server executes its functions reliably and adheres to the prescribed protocol, there might be an inherent inclination to interpret cached templates or analyze the nature of incoming requests. Such servers could attempt to glean additional information from the data they handle, aiming either to reconstruct the original data or infer relationships between different queries. The system's design meticulously ensures that the server cannot infer private templates, guaranteeing accurate validation of a user's group affiliation and determining their group identity when necessary.

A. RELATED WORK

The paradigm of anonymous authentication for group members has been a staple in cryptography for years [1]. However, the applications we consider, particularly for biometrics, diverge significantly from established models of authentication, identification, and secret binding. Traditional approaches ensure security either at the server or client end, but they invariably result in the revelation of the user's identity.

Aggregating signals into a unified representation is widely adopted in computer vision. Techniques such as Bag of Words (BoW) [2] and VLAD [3] consolidate local descriptors from an image into a comprehensive descriptor. Contemporary adaptations like BoW encoding convolutional features of CNNs have been introduced [4]. Arandjelovic et al. further refined VLAD with a learnable pooling layer, termed NetVLAD [5], with subsequent iterations exploring soft assignment to multiple clusters [6]. Distinctly, Zhong et al. [7] developed a descriptor for the faces of celebrities present in the same photo. Although their system excels with two faces per image, performance drops with an increasing number of faces. Our approach, while inspired by these techniques, diverges primarily because our queries comprise a singular face, and our group representations usually encompass more than two faces captured under diverse conditions.

The amalgamation of templates into a collective representation is prominent in biometrics. For example, [8] combined multiple captures of a single person's face to offset challenges from poses, expressions, and image quality. Contrastingly, our focus is on aggregating unique faces from different individuals in a group. Notably, while conventional methods prioritize retrieving visually analogous elements, they do not inherently provide security or privacy.

The methodologies introduced in [9] and [10] pivot on the notion of transforming randomly-selected templates into discrete embeddings. These are then coalesced to form a singular group representation. Their cost-effectiveness, coupled with the inherent challenges of identity reconstruction, positions these strategies as particularly compelling. Further intricacies, including the impact of the sparsity level of high-dimensional features characterizing group members on aspects like security, compactness, and verification efficacy, are expounded upon in [11]. In a progressive stride, [12] departs from the conventional approach. Instead of statically computing group representations based on predefined entities, this study concurrently learns group representation and their corresponding assignments. By adopting variance as a metric for dissimilarity, the approach endeavors to minimize inter-group differences while amplifying intra-group distinctions. Empirical assessments indicate that this dual learning mechanism fosters enhanced performance, all while maintaining robust security measures.

B. CONTRIBUTION

We propose a novel group membership assignment based on *joint modeling and learning of nonlinear transforms with priors* and *nonlinear transform representation and group representative assignment*. The model parameters are learned by minimizing an empirical expectation of the model log likelihood, which, under the discrimination prior, corresponds to maximizing the discrimination power measure. The proposed model allows a rejection option over continuous, discontinuous, and overlapping regions in the transform domain. Our learning strategy is based on an iterative, alternating algorithm with three steps. The honest but curious server cannot reconstruct the signatures, i.e., the data structure is protected in terms of security requirements. Moreover, the privacy of the users is guaranteed by anonymous verification.

C. OUTLINE

Sec. III elucidates our framework. A comprehensive performance analysis is presented in Sec. VII. Concluding remarks are in Sec. VIII.

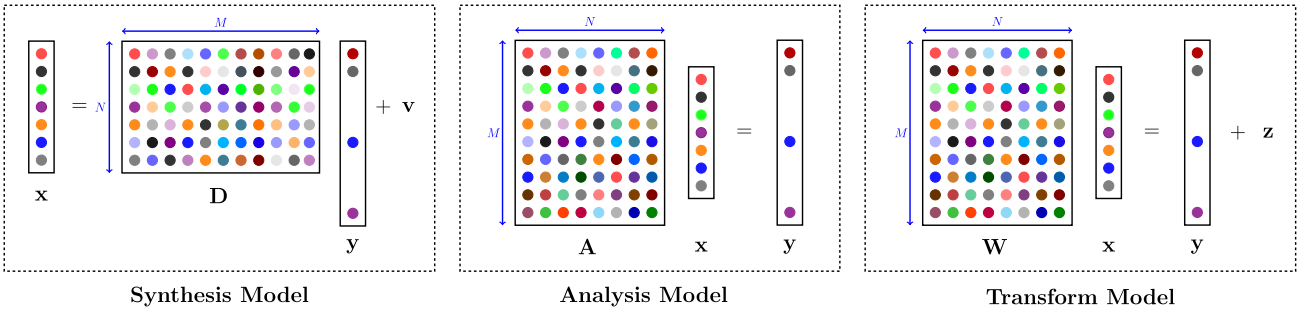


FIGURE 1. Visualization of three sparse data representation models.

D. NOTATION

Vectors and matrices are denoted by boldface lower-case ($\mathbf{x}$) and upper-case ($\mathbf{X}$) letters, respectively. We consider the same notation for a random vector $\mathbf{x}$ and its realization. The difference should be clear from the context. We use the notation $[N]$ for the set $\{1, \cdots, N\}$. The superscript $(\cdot)^T$ stands for the transpose.

II. PRELIMINARIES ON SPARSE DATA REPRESENTATION

The foundational principle of *sparse representation* is to express data using fewer components than traditionally required, without compromising the integrity or essential characteristics of the data. This principle has found widespread application across various domains, including feature extraction, clustering, classification, and signal reconstruction, underscoring its versatility and impact. Three primary models of sparse data representation are the synthesis model, the analysis model, and the transform model. In the following section, we briefly review the essence of these models. Figure 1 visualizes the core concepts of these models.

A. SYNTHESIS MODEL

The synthesis model assumes that a high-dimensional data sample $\mathbf{x}_i \in \mathbb{R}^N$ can be approximated as a linear combination of a select few basis columns (atoms) from a dictionary $\mathbf{D} = [\mathbf{d}_1, \cdots, \mathbf{d}_M] \in \mathbb{R}^{N \times M}$, yielding a sparse representation $\mathbf{y}_i \in \mathbb{R}^M$, where $\|\mathbf{y}_i\|_0 \ll M$. The deterministic formulation of the sparse synthesis model is expressed as $\mathbf{x}_i = \mathbf{D}\mathbf{y}_i + \mathbf{v}_i$, where $\mathbf{v}_i \in \mathbb{R}^N$ denotes the approximation error in the original data domain—essentially, the residual between the actual data and its sparse approximation. The synthesis model is also known as the regression model with sparsity regularized penalty.

B. ANALYSIS MODEL

The analysis model employs a dictionary $\mathbf{A} \in \mathbb{R}^{M \times N}$ with $M > N$ to analyze the data sample $\mathbf{x}_i \in \mathbb{R}^N$. Given this data sample $\mathbf{x}_i \in \mathbb{R}^N$ and dictionary $\mathbf{A} \in \mathbb{R}^{M \times N}$, the model assumes the sparse representation $\mathbf{y}_i = \mathbf{A}\mathbf{x}_i$, i.e., $\|\mathbf{y}_i\|_0 \ll M$.

C. TRANSFORM MODEL

The transform model assumes that a high-dimensional data sample $\mathbf{x}_i$ can be effectively sparsified through a linear transformation $\mathbf{W} \in \mathbb{R}^{M \times N}$, resulting in $\mathbf{W}\mathbf{x}_i = \mathbf{y}_i + \mathbf{z}_i$. In this deterministic formulation, $\mathbf{y}_i \in \mathbb{R}^M$ denotes the sparse representation, which satisfies $\|\mathbf{y}_i\|_0 \ll M$, and $\mathbf{z}_i$ represents the approximation error within the transform domain. This model fundamentally employs linear transformation but significantly enhances its applicability through the integration of nonlinearities. Nonlinearities can be introduced via a generalized element-wise nonlinearity operator $\psi_{\boldsymbol{\theta}}(\mathbf{W}\mathbf{x}_i)$, where $\boldsymbol{\theta}$ denotes the parameters modulating the nonlinearity. Examples of nonlinear transforms include the hard thresholding function, the soft thresholding function, and ReLU activation function.

III. PROPOSED FRAMEWORK OVERVIEW

A. ASSIGNMENT AND LEARNING APPROACH

In our research, we introduce a methodology for learning a privacy-preserving assignment. This stands in contrast to the works cited in [13], [14], [15], [16], [17], [18], and [19], where various generic privacy-preserving identification or search mechanisms have been put forward. Unlike these approaches, our protocol assigns group membership not by assessing similarity in the original data space, but through evaluating a min-max functional measure within a transformed space. This assignment procedure unfolds in two distinct steps: **(i)** generation of candidate nonlinear transform representations, and **(ii)** evaluation of the min-max measure across these representations leading to the assignment of group membership.

In our pursuit to develop an assignment mechanism that is sparse, discriminative, information-preserving, and respects privacy, we employ specific priors when modeling the candidate nonlinear transforms. In the learning process, we simultaneously focus on two core objectives: **(i)** estimating the parameters inherent to the candidate transforms, and **(ii)** assigning group membership. To achieve both the discriminative and privacy-preserving goals, we advocate for support intersection measures. According to these measures, every representation of a data instance that is assigned to a specific group should maximize its support intersection with other representations within that same group. On the contrary, a representation associated with a particular group should minimize its support intersection with representations

from different groups. Here, "support intersection" between vector representations denotes the count of non-zero elements that both vectors contain at the same index positions.

B. SETUP

Consider a dataset of K data vectors, $\mathbf{x}_i \in \mathbb{R}^N$, represented as $\mathbf{X} \in \mathbb{R}^{N\times K}$. For simplicity, we postulate that each entity, whether an item, device, or individual, belongs to a singular group. This means that each data vector, $\mathbf{x}_i$, is affiliated with one distinct group $c \in \{1, \cdots, C\}$. We further assume that entities within the same group exhibit similar features, positioning them within proximate sub-spaces.

Our primary aim is to construct an information-preserving group representation, $\boldsymbol{\theta}_c = [\boldsymbol{\nu}_c, \boldsymbol{\tau}_c] \in \mathcal{T}^{L\times 2}, c \in [C]$, for a set of members. This representation should adhere to security requirements; specifically, it should shield the inherent structure of the data from servers that, while trustworthy, may be intrusively curious (honest, but curious servers). Here, the vectors $\boldsymbol{\nu}_c$ and $\boldsymbol{\tau}_c$ are parameters representing *similarity* and *dissimilarity* within group $c \in [C]$. Once these security requirements are met, we commit these group representatives to a public server. Our coding alphabet, denoted as $\mathcal{T}$, can adopt various forms such as binary, ternary, continuous, and so forth. In this research, we consider a non-quantized alphabet.

Our secondary goal revolves around deriving an information-preserving transform representation, $\mathbf{y}_i$. This representation, corresponding to the original space group member $\mathbf{x}_i$, should remain distinguishable from non-members. Simultaneously, it is imperative that it maintains the privacy of the underlying entity, thus not disclosing the member's identity.

C. FRAMEWORK OVERVIEW

Our framework comprises the subsequent steps, as illustrated in Figure 2:

1) PREPARATION AT OWNER SIDE

The owner jointly estimates the sparse representations and assigns the group representatives from the data they possess. This estimation is achieved using a *trained linear* mapping, followed by a *generalized element-wise nonlinearity*. These group representatives are then transmitted to the server, which serves as a storage facility.

2) QUERYING AT CLIENT SIDE

The client produces a sparse representation from their query data. This is performed utilizing the same *trained linear* mapping and the *generalized element-wise nonlinearity*. Subsequently, the client forwards this sparse representation to the server.

3) SEARCHING AT SERVER SIDE

The server uses the received data to conduct similarity and dissimilarity searches among the group representatives. This process aims to precisely identify the group that most closely matches the client's query.

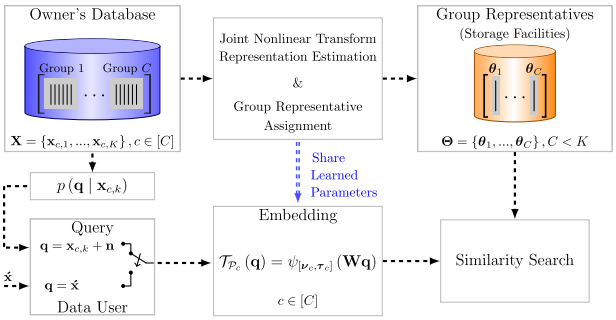


FIGURE 2. General block diagram of the proposed framework.

IV. ASSIGNMENT PRINCIPLE AND PROBABILISTIC MODEL

In this section, we elucidate the core principle guiding our group assignment methodology. Our focus is on the evaluation of candidate representations through a specific measure, and the probabilistic model integral to our approach of privacy-preserving group membership assignment.

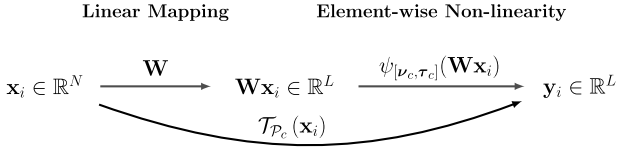
A. GROUP MEMBERSHIP ASSIGNMENT PRINCIPLE

The procedure for our group membership assignment is bifurcated into:

(i) Candidate Representation Generation, and
(ii) Group and Nonlinear Transform Representation Assignment.

1) CANDIDATE REPRESENTATION GENERATION

To derive a candidate representation, denoted as $\mathbf{q}_{i,c}$, we employ a nonlinear transform (NT) symbolized as $\mathcal{T}_{\mathcal{P}_c}$ equipped with parameters $\mathcal{P}_c = \{\mathbf{W} \in \mathbb{R}^{L\times N}, \boldsymbol{\nu}_c \in \mathbb{R}^L, \boldsymbol{\tau}_c \in \mathbb{R}^L\}$. This NT is applied to the input data $\mathbf{x}_i$. We provide a schematic representation of this NT-based generation process below. Here, the function $\psi_{[\boldsymbol{\nu}_c,\boldsymbol{\tau}_c]}(\cdot)$ denotes an element-wise nonlinearity parameterized by $[\boldsymbol{\nu}_c, \boldsymbol{\tau}_c] =: \boldsymbol{\theta}_c$, which we will unpack in subsequent sections.



For the generation of an ensemble of candidate representations, symbolized as $\{\mathbf{q}_{i,1}, \cdots, \mathbf{q}_{i,C}\}$, we leverage a set of C NTs represented as $\mathcal{T}_T = \{\mathcal{T}_{\mathcal{P}_1}, \cdots, \mathcal{T}_{\mathcal{P}_C}\}$, paired with their respective parameter sets $\mathcal{P}_T = \{\mathcal{P}_1, \cdots, \mathcal{P}_C\}$. Each NT, denoted as $\mathcal{T}_{\mathcal{P}_c}$, is uniquely characterized by its parameter set $\mathcal{P}_c = \{\mathbf{W} \in \mathbb{R}^{L\times N}, \boldsymbol{\nu}_c \in \mathbb{R}^L, \boldsymbol{\tau}_c \in \mathbb{R}^L\}, c \in [C]$, though all share a common linear map $\mathbf{W}$. Yet, every NT possesses unique pairs $[\boldsymbol{\nu}_c, \boldsymbol{\tau}_c]$ tied to a specific group $c \in [C]$. For brevity, we also label the set of these parameter pairs as $\boldsymbol{\theta} = \{\boldsymbol{\theta}_1, \cdots, \boldsymbol{\theta}_C\} = \{[\boldsymbol{\nu}_1, \boldsymbol{\tau}_1], \cdots, [\boldsymbol{\nu}_C, \boldsymbol{\tau}_C]\}$.

2) GROUP AND NT REPRESENTATION ASSIGNMENT

Given the candidate representations, we proceed with the actual assignment. Unlike traditional similarity (or dissimilarity) measures, we employ a min-max functional measure. Each candidate representation undergoes evaluation through this measure. Based on the evaluation score, we determine the group membership and its associated NT representation.

B. PROBABILISTIC MODEL

To introduce the probabilistic model of our framework, let's start with the decomposition of the joint probability distribution $p(\mathbf{x}_i, \mathbf{y}_i, \boldsymbol{\theta}, \mathbf{W})$. Using the chain rule, we have $p(\mathbf{x}_i, \mathbf{y}_i, \boldsymbol{\theta}, \mathbf{W}) = p(\mathbf{x}_i, \mathbf{y}_i, \boldsymbol{\theta} \mid \mathbf{W})\, p(\mathbf{W})$. Assuming $p(\mathbf{x}_i \mid \mathbf{W}) = p(\mathbf{x}_i)$ and applying the chain rule again, we have:

$$p(\mathbf{x}_i, \mathbf{y}_i, \boldsymbol{\theta} \mid \mathbf{W}) = p(\mathbf{y}_i, \boldsymbol{\theta} \mid \mathbf{x}_i, \mathbf{W})\, p(\mathbf{x}_i) \tag{1a}$$

$$= p(\mathbf{x}_i \mid \mathbf{y}_i, \boldsymbol{\theta}, \mathbf{W})\, p(\mathbf{y}_i, \boldsymbol{\theta} \mid \mathbf{W}) \tag{1b}$$

Therefore, using Bayes' rule, we have:

$$p(\mathbf{y}_i, \boldsymbol{\theta} \mid \mathbf{x}_i, \mathbf{W}) \propto p(\mathbf{x}_i \mid \mathbf{y}_i, \boldsymbol{\theta}, \mathbf{W})\, p(\mathbf{y}_i, \boldsymbol{\theta} \mid \mathbf{W}) \tag{2}$$

Let's assume $p(\mathbf{y}_i, \boldsymbol{\theta} \mid \mathbf{W}) = p(\mathbf{y}_i, \boldsymbol{\theta})$, i.e., neglects the dependence on $\mathbf{W}$. This independence assumption allows us more flexibility in the class of assumptions related to sparsity and discrimination for the parametric prior $p(\mathbf{y}_i, \boldsymbol{\theta})$. Also, for simplification, let $p(\mathbf{W} \mid \mathbf{x}_i) = p(\mathbf{W})$. Given K data samples $\mathbf{X} = [\mathbf{x}_1, \cdots \mathbf{x}_K]$ and using (2), we can consider the following learning model:

$$p(\mathbf{Y}, \mathbf{W} \mid \mathbf{X}) = p(\mathbf{Y} \mid \mathbf{W}, \mathbf{X})\, p(\mathbf{W} \mid \mathbf{X}) \tag{3a}$$

$$= \prod_{i=1}^{K} \int_{\boldsymbol{\theta}} p(\mathbf{y}_i, \boldsymbol{\theta} \mid \mathbf{x}_i, \mathbf{W})\, p(\mathbf{W})\, \mathrm{d}\boldsymbol{\theta} \tag{3b}$$

$$= \prod_{i=1}^{K} \int_{\boldsymbol{\theta}} p(\mathbf{x}_i \mid \mathbf{y}_i, \boldsymbol{\theta}, \mathbf{W})\, p(\mathbf{y}_i, \boldsymbol{\theta})\, p(\mathbf{W})\, \mathrm{d}\boldsymbol{\theta} \tag{3c}$$

Using this probabilistic formulation, we characterize our learning model through the following components:

- A sparsifying error coupled with a $\boldsymbol{\theta}$ adjustment error, represented by the prior $p(\mathbf{x}_i \mid \mathbf{y}_i, \boldsymbol{\theta}, \mathbf{W})$.
- Discrimination and sparsity prior $p(\mathbf{y}_i, \boldsymbol{\theta})$.
- Linear map $\mathbf{W}$ prior $p(\mathbf{W})$.

By adopting this structured probabilistic perspective, we aim for a holistic understanding of the dynamics between the observed inputs, their associated transformations, model parameters, and their linear mapping.

V. ADDRESSING PROBABILISTIC MODELING PRIORS

A. SPARSIFYING ERROR AND MODEL PARAMETERS ADJUSTMENT ERROR PRIOR

In our pursuit to capture the intricate relationship between observed data and its inherent structure, we formulate a probabilistic construct as presented below:

$$p(\mathbf{x}_i \mid \mathbf{y}_i, \boldsymbol{\theta}, \mathbf{W}) \propto \exp\left[-\frac{1}{\beta_{\text{spa}}} d_{\text{spa}}(\mathbf{W}\mathbf{x}_i, \mathbf{y}_i) - \frac{1}{\beta_{\text{adj}}} d_{\text{adj}}(\mathbf{W}\mathbf{x}_i, \boldsymbol{\theta})\right], \tag{4}$$

where $d_{\text{spa}}(\cdot,\cdot) : \mathbb{R}^L \times \mathbb{R}^L \to \mathbb{R}$ represents the sparsifying error measure, and $d_{\text{adj}}(\cdot,\cdot) : \mathbb{R}^L \times \mathbb{R}^L \to \mathbb{R}$ represents the model parameters adjustment error. The parameters β_{spa} and β_{adj} are scaling factors.

Our primary objective for introducing this construct is twofold. Initially, we encapsulate the sparsifying error vector $\mathbf{W}\mathbf{x}_i - \mathbf{y}_i$. Subsequently, and crucially, we seek to refine the alignment of our model's estimations with observed data by addressing discrepancies. This is achieved by implementing an adjustment to the discrimination parameter error vector, denoted as $\mathbf{W}\mathbf{x}_i - \boldsymbol{\nu}_c - \boldsymbol{\tau}_c$. We define our measures as follows:

$$d_{\text{spa}}(\mathbf{W}\mathbf{x}_i, \mathbf{y}_i) = \frac{1}{2}\|\mathbf{W}\mathbf{x}_i - \mathbf{y}_i\|_2^2, \tag{5}$$

$$d_{\text{adj}}(\mathbf{W}\mathbf{x}_i, \boldsymbol{\theta}) = \frac{1}{2}\|\mathbf{W}\mathbf{x}_i - \boldsymbol{\nu}_c - \boldsymbol{\tau}_c\|_2^2. \tag{6}$$

where the index c corresponds to the assigned class of the data sample $\mathbf{x}_i$. These measures directly influence our probabilistic model's likelihood expressions, described as:

$$p(\mathbf{x}_i \mid \mathbf{y}_i, \mathbf{W}) \propto \exp\left[-\frac{1}{\beta_{\text{spa}}} d_{\text{spa}}(\mathbf{W}\mathbf{x}_i, \mathbf{y}_i)\right], \tag{7}$$

$$p(\mathbf{x}_i \mid \boldsymbol{\theta}, \mathbf{W}) \propto \exp\left[-\frac{1}{\beta_{\text{adj}}} d_{\text{adj}}(\mathbf{W}\mathbf{x}_i, \boldsymbol{\theta})\right]. \tag{8}$$

B. SPARSITY AND DISCRIMINATION PRIOR

We model our sparsity-inducing and discrimination prior as follows:

$$p(\mathbf{y}_i, \boldsymbol{\theta}) \propto \exp\left[-\frac{1}{\beta_{\text{disc}}} d_{\text{disc}}(\mathbf{y}_i, \boldsymbol{\theta}) - \frac{1}{\beta_{\text{p}}} d_{\text{p}}(\boldsymbol{\theta}) - \frac{1}{\beta_1} d_1(\mathbf{y}_i)\right], \tag{9}$$

where d_{disc} represents the discrimination prior measure, d_{p} denotes the similarity/dissimilarity prior measure, and d_1 indicates the sparsity measure. The parameters β_{disc}, β_{p}, and β_1 are scaling factors. By considering the ℓ_1-norm as the sparsity measure for vector $\mathbf{y}_i$, i.e., $d_1(\mathbf{y}_i) = \|\mathbf{y}_i\|_1$, we introduce a prior that induces sparsity in $\mathbf{y}_i$. The sparsity-inducing prior on $\mathbf{y}_i$ is given by:

$$p(\mathbf{y}_i) \propto \exp\left(-\|\mathbf{y}_i\|_1 / \beta_1\right). \tag{10}$$

The discrimination prior is modeled similarly as:

$$p(\boldsymbol{\theta} \mid \mathbf{y}_i) \propto \exp\left[-\frac{1}{\beta_{\text{disc}}} d_{\text{disc}}(\mathbf{y}_i, \boldsymbol{\theta}) - \frac{1}{\beta_{\text{p}}} d_{\text{p}}(\boldsymbol{\theta})\right] \tag{11}$$

In order to define the measures d_{disc} and d_{p}, we first describe our min-max discrimination measure based on the following assumptions:

1) The discrimination measure is characterized by relationships based on the support intersection between $\mathbf{y}_i$, $\boldsymbol{\nu}_c$ and $\boldsymbol{\tau}_c$.

2) The discrimination measure has a min-max structure, with its expression factored with respect to $\boldsymbol{\nu}_c$ and $\boldsymbol{\tau}_c$.

We then elucidate our foundational measures related to *similarity*, *dissimilarity*, and *strength*, focusing on the support intersection between the representations and the $\boldsymbol{\theta}_c$ parameter pair. Following this, we detail our min-max discrimination functional measure.

1) QUANTIFYING REPRESENTATION SIMILARITY AND DISSIMILARITY

The measure Sim quantifies the *similarity* between two representations $\mathbf{y}_1$ and $\mathbf{y}_2$. It is defined as:

$$\mathsf{Sim}(\mathbf{y}_1, \mathbf{y}_2) = \|\mathbf{y}_1^+ \odot \mathbf{y}_2^+\|_1 + \|\mathbf{y}_1^- \odot \mathbf{y}_2^-\|_1, \tag{12}$$

where $\mathbf{y}_1 = \mathbf{y}_1^+ - \mathbf{y}_1^-, \mathbf{y}_2 = \mathbf{y}_2^+ - \mathbf{y}_2^-$, $\mathbf{y}_1^+ = \max(\mathbf{y}_1, 0)$ and $\mathbf{y}_1^- = \max(-\mathbf{y}_1, 0)$.

The term $\|\mathbf{y}_1^+ \odot \mathbf{y}_2^+\|_1$ measures the similarity in the positive components. Essentially, for two vectors to have high similarity in their positive components, they should both have positive values at the same indices, and these values should be relatively large. This is captured by the element-wise multiplication ($\odot$) and the L_1 norm. Similarly, the term $\|\mathbf{y}_1^- \odot \mathbf{y}_2^-\|_1$ evaluates the agreement in their negative components. The overall similarity score, therefore, combines the extent to which the positive components agree and the negative components agree. The larger the score, the more similar the two representations are in both their positive and negative components.

The measure Dis, related to *dissimilarity* (oppositeness) between two representations $\mathbf{y}_1$ and $\mathbf{y}_2$ is defined as:

$$\mathsf{Dis}(\mathbf{y}_1, \mathbf{y}_2) = \|\mathbf{y}_1^+ \odot \mathbf{y}_2^-\|_1 + \|\mathbf{y}_1^- \odot \mathbf{y}_2^+\|_1. \tag{13}$$

The term $\|\mathbf{y}_1^+ \odot \mathbf{y}_2^-\|_1$ computes the contrast between the positive components of $\mathbf{y}_1$ and the negative components of $\mathbf{y}_2$. A high value in this term indicates that there are indices where $\mathbf{y}_1$ has a positive value while $\mathbf{y}_2$ has a negative value, or vice versa. The element-wise multiplication emphasizes this opposition. The term $\|\mathbf{y}_1^- \odot \mathbf{y}_2^+\|_1$ mirrors this idea, identifying negative values in $\mathbf{y}_1$ where there are positive values in $\mathbf{y}_2$.

Furthermore, the measure Stg, which correlates to the *strength* on the support intersection, is articulated as:

$$\mathsf{Stg}(\mathbf{y}_1, \mathbf{y}_2) = \|\mathbf{y}_1 \odot \mathbf{y}_2\|_2^2. \tag{14}$$

Figure 3 depcits a visual representation of the similarity and dissimilarity contributions pertaining to two typical sparse representations, $\mathbf{y}_1$ and $\mathbf{y}_2$.

2) MIN-MAX DISCRIMINATION PRIOR MEASURE

We introduce a min-max functional, denoted as $d_{\mathsf{disc}}(\mathbf{y}_i, \boldsymbol{\theta})$:

$$d_{\mathsf{disc}}(\mathbf{y}_i, \boldsymbol{\theta}) = \min_d \max_c \left[\mathsf{Sim}(\mathbf{y}_i, \boldsymbol{\tau}_d) + \mathsf{Dis}(\mathbf{y}_i, \boldsymbol{\nu}_c)\right] + \min_d \mathsf{Stg}(\mathbf{y}_i, \boldsymbol{\tau}_d). \tag{15}$$

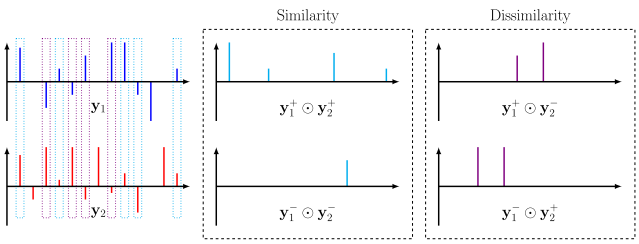


FIGURE 3. Visualization of similarity and dissimilarity measures.

This formulated metric ensures that the transform representation, $\mathbf{y}_i$, aligns with the following criteria:

1) The similarity relative to $\boldsymbol{\tau}_d$ is minimized, as measured by Sim.
2) The intersection strength regarding support with respect to $\boldsymbol{\tau}_d$ is minimized, as evaluated by Stg.
3) The similarity relative to $\boldsymbol{\nu}_c$ is maximized, as measured by Sim.

To understand the dynamics of this prior measure, consider the formulation as akin to a dance of forces within a physical field. In this space, the point—our transform representation, denoted as $\mathbf{y}_i$—is swayed by two principal forces: one symbolized by $\mathsf{Sim}(\mathbf{y}_i, \boldsymbol{\tau}_d)$ and the other by $\mathsf{Dis}(\mathbf{y}_i, \boldsymbol{\nu}_c)$. The force governed by $\min_d \mathsf{Sim}(\mathbf{y}_i, \boldsymbol{\tau}_d)$ highlights the minimum similarity metric as defined by Sim. This force operates with an intent to maximize the "distance" between $\mathbf{y}_i$ and $\boldsymbol{\tau}_d$, effectively pushing $\mathbf{y}_i$ away from regions that resemble $\boldsymbol{\tau}_d$. In contrast, the force dictated by $\max_c \mathsf{Dis}(\mathbf{y}_i, \boldsymbol{\nu}_c)$ is somewhat paradoxical. While the term 'dissimilarity' might suggest repulsion, within our defined metric space, it functions differently. This force underscores $\mathbf{y}_i$'s pursuit to align as closely as possible with the representation $\boldsymbol{\nu}_c$ by amplifying its 'oppositeness' or dissimilarity. Consequently, it attracts $\mathbf{y}_i$ towards $\boldsymbol{\nu}_c$, seeking regions where this measure of 'oppositeness' is maximized.

At the heart of this formulation is the score $d_{\mathsf{disc}}(\mathbf{y}_i, \boldsymbol{\theta})$. This score is not just a mere numerical value but a manifestation of equilibrium. It embodies the delicate balance, almost a dance, between the aforementioned forces. As they act upon $\mathbf{y}_i$, they ensure its positioning is not arbitrary but follows the intricate choreography laid out by the relations of Sim and Dis.

3) MODEL PARAMETERS θ PRIOR MEASURE

We introduce the measure $d_{\mathsf{p}}(\boldsymbol{\theta})$ to quantify the combined influence of different parameter sets within the indexed range $\{1, \cdots, C\}$. To achieve this, we harness pairwise interactions between distinct parameter sets, computing their aggregate effects as follows:

$$d_{\mathsf{p}}(\boldsymbol{\theta}) = \sum_{c \in \{1,\cdots,C\}} \sum_{d \in \{1,\cdots,C\} \setminus c} d_{\mathsf{disc}}(\boldsymbol{\nu}_c, \boldsymbol{\theta}_d) + d_{\mathsf{disc}}(\boldsymbol{\tau}_c, \boldsymbol{\theta}_d). \tag{16}$$

Here, the notation $\{1, \cdots, C\} \setminus c$ refers to the set $\{1, \cdots, c-1, c+1, \cdots, C\}$, effectively excluding the element c from the set $\{1, \cdots, C\}$. Note that we have

$p(\boldsymbol{\theta}) \propto \exp\left(-d_{\mathsf{p}}(\boldsymbol{\theta})/\beta_{\mathsf{p}}\right)$. Indeed, $d_{\mathsf{p}}(\boldsymbol{\theta})$ quantifies both the similarity and dissimilarity between the parameters $\boldsymbol{\theta}$ by employing the discrimination measure d_{disc}.

The dual summation ensures that for every parameter c, we account for its relationship with every other distinct parameter, d. In doing so, d_{p} encapsulates a holistic view of parameter inter-dependencies and interactions in the model.

C. LINEAR MAP PRIOR

The purpose of the 'linear map prior' is to penalize information loss while simultaneously discouraging the adoption of undesirable matrices. To achieve this dual objective, we regularize both the condition number and the expected coherence of the matrix $\mathbf{W}$. Specifically, the prior is defined as:

$$p(\mathbf{W}) \propto \exp\left(-\Omega(\mathbf{W})\right), \tag{17}$$

where

$$\Omega(\mathbf{W}) = \frac{1}{\beta_{\mathsf{W}_1}}\|\mathbf{W}\|_F^2 + \frac{1}{\beta_{\mathsf{W}_2}}\|\mathbf{W}\mathbf{W}^T - \mathbf{I}\|_F^2 - \frac{1}{\beta_{\mathsf{W}_3}}\log|\det \mathbf{W}^T\mathbf{W}|. \tag{18}$$

In this formulation:

- The term $\|\mathbf{W}\|_F^2$ serves as a regularizer, penalizing the magnitude of matrix $\mathbf{W}$ to ensure stability.
- $\|\mathbf{W}\mathbf{W}^T - \mathbf{I}\|_F^2$ aims to minimize the deviation of $\mathbf{W}$ from orthogonality.
- Lastly, $\log|\det \mathbf{W}^T\mathbf{W}|$ assesses the volume scaling factor of $\mathbf{W}$, thereby promoting full rank matrices.

For a more detailed exploration and justification of the constraint $\Omega(\mathbf{W})$ on the linear map $\mathbf{W}$, readers are directed to the works of [14] and [15], along with associated references contained therein.

VI. LEARNING MODEL TO ASSIGN GROUP MEMBERSHIP

A. PROBLEM FORMULATION

Consider a given training dataset, denoted by $\mathbf{X}$. The task of directly maximizing the probability function $p(\mathbf{x}_i \mid \mathbf{y}_i, \boldsymbol{\theta}, \mathbf{W})\, p(\boldsymbol{\theta}|\mathbf{y}_i)\, p(\mathbf{y}_i)\, p(\mathbf{W})$ over the parameter space $\{\mathbf{Y}, \boldsymbol{\theta}, \mathbf{W}\}$ is computationally burdensome, primarily due to the high dimensionality and nonlinearities involved. To circumvent this challenge, we pivot our approach towards minimizing the negative logarithm of our probabilistic model (3c), over the variables $\{\mathbf{Y}, \boldsymbol{\theta}, \mathbf{W}\}$. That is we aim to minimize the negative log likelihood

$$-\log \prod_{i=1}^{K} \int_{\boldsymbol{\theta}} p(\mathbf{x}_i \mid \mathbf{y}_i, \boldsymbol{\theta}, \mathbf{W})\, p(\mathbf{y}_i, \boldsymbol{\theta})\, p(\mathbf{W})\, \mathrm{d}\boldsymbol{\theta} \tag{19}$$

over the variables $\{\mathbf{Y}, \boldsymbol{\theta}, \mathbf{W}\}$. However, minimizing the exact negative logarithm is difficult since it requires integrating to compute the marginal and the partitioning function of the prior $p(\mathbf{y}, \boldsymbol{\theta})$. Instead, we consider minimizing the negative logarithm of its maximum point-wise estimate:

$$\int_{\boldsymbol{\theta}_{\mathsf{est}}} p(\mathbf{x}_i \mid \mathbf{y}_i, \boldsymbol{\theta}_{\mathsf{est}}, \mathbf{W})\, p(\mathbf{y}_i, \boldsymbol{\theta}_{\mathsf{est}})\, \mathrm{d}\boldsymbol{\theta}_{\mathsf{est}} \leq \kappa\, p(\mathbf{x}_i \mid \mathbf{y}_i, \boldsymbol{\theta}, \mathbf{W})\, p(\mathbf{y}_i, \boldsymbol{\theta}), \tag{20}$$

where κ is a constant. Here, we assume that $\boldsymbol{\theta}$ are the parameters at which $p(\mathbf{x}_i \mid \mathbf{y}_i, \boldsymbol{\theta}_{\mathsf{est}}, \mathbf{W})\, p(\mathbf{y}_i, \boldsymbol{\theta}_{\mathsf{est}})$ attains its maximum value.

Considering Eq. (3), (19), (20), we arrive at the following problem formulation:

$$\{\widehat{\mathbf{Y}}, \widehat{\boldsymbol{\theta}}, \widehat{\mathbf{W}}\} = \arg\min_{\mathbf{Y},\boldsymbol{\theta},\mathbf{W}} \sum_{i=1}^{K} \Big[-\log p(\mathbf{x}_i \mid \mathbf{y}_i, \boldsymbol{\theta}, \mathbf{W}) - \log p(\mathbf{y}_i, \boldsymbol{\theta}) \Big] - \log p(\mathbf{W}). \tag{21}$$

By substituting the measure obtained from section V, we can define our optimization problem as follows:

$$\{\widehat{\mathbf{Y}}, \widehat{\boldsymbol{\theta}}, \widehat{\mathbf{W}}\} = \arg\min_{\mathbf{Y},\boldsymbol{\theta},\mathbf{W}} \sum_{i=1}^{K} \Big[d_{\mathsf{spa}}(\mathbf{u}_i, \mathbf{y}_i) + \lambda_{\mathsf{adj}}\, d_{\mathsf{adj}}(\mathbf{u}_i, \boldsymbol{\theta}) + \lambda_{\mathsf{disc}}\, d_{\mathsf{disc}}(\mathbf{y}_i, \boldsymbol{\theta}) + \lambda_{\mathsf{p}}\, d_{\mathsf{p}}(\boldsymbol{\theta}) + \lambda_1 \|\mathbf{y}_i\|_1 \Big] + \lambda_{\mathsf{W}}\, \Omega(\mathbf{W}), \tag{22}$$

where $\mathbf{u}_i = \mathbf{W}\mathbf{x}_i$ represents the transformed input data. The collection $\{\lambda_{\mathsf{adj}}, \lambda_{\mathsf{disc}}, \lambda_{\mathsf{p}}, \lambda_1, \lambda_{\mathsf{W}}\}$ denotes the set of Lagrangian multipliers, each serving as an inverse coefficient to their associated scaling parameters. We set $\lambda_{\mathsf{spa}} = 1$.

An essential point of clarification here is that our derived solution to Eq. (22) does not equate to the maximum a posteriori (MAP) solution. While the MAP provides a point estimate by maximizing the posterior distribution, it is computationally intensive due to the complexity of calculating higher-dimensional integrals in the parameter space.

Instead, our framework articulated in Eq. (22) encapsulates an integrated marginal minimization (IMM) strategy. This method entails a sequence of iterative steps, where, in each iteration, we maximize the terms of our model with respect to the variables $\mathbf{Y}$, $\boldsymbol{\theta}$, and $\mathbf{W}$. This integrated marginal minimization offers several advantages:

- Computational Efficiency: Unlike the MAP approach, which often requires complex integrations over the parameter space, our method iteratively maximizes simpler terms, allowing for more efficient computational procedures.
- Flexibility: Given the iterative nature of our method, it is more adaptable to various datasets and can better accommodate changes in data distribution, especially in scenarios with limited or evolving data.
- Stability: The integrated approach reduces the risk of settling into local optima that do not represent the broader dataset well. By considering marginal effects iteratively, we can capture more global patterns in the data, enhancing the model's generalization capabilities.

In essence, while the MAP solution offers a theoretical ideal, practical constraints necessitate the use of strategies

like integrated marginal minimization, which strike a balance between theoretical rigor and computational tractability. Our methodology facilitates the identification of a joint local maximum in the space $\{\mathbf{Y}, \boldsymbol{\theta}, \mathbf{W}\}$ for the likelihood and prior probabilities.

B. LEARNING ALGORITHM

We propose an alternating block coordinate descent algorithm that progresses across three stages:

(i) Simultaneous estimation of the representation $\mathbf{y}_i$ and assignment of group membership c,
(ii) Update of the group parameters, represented as $\boldsymbol{\theta} = \{\boldsymbol{\theta}_1, \cdots, \boldsymbol{\theta}_C\} = \{[\boldsymbol{\nu}_1, \boldsymbol{\tau}_1], \cdots, [\boldsymbol{\nu}_C, \boldsymbol{\tau}_C]\}$,
(iii) Update of the linear map $\mathbf{W}$.

1) NT REPRESENTATION ESTIMATION AND ASSIGNMENT

Given the dataset $\mathbf{X}$, the latest estimate of the group membership parameters $\boldsymbol{\theta}$, and the current approximation of the linear map $\mathbf{W}$, the expression in (22) simplifies to the ensuing representation estimation problem:

$$[\widehat{\mathbf{y}}_1, \cdots, \widehat{\mathbf{y}}_K] = \arg \min_{[\mathbf{y}_1, \cdots, \mathbf{y}_K]} \sum_{i=1}^{K} \Big[\frac{1}{2} \|\mathbf{u}_i - \mathbf{y}_i\|_2^2 + \lambda_1 \|\mathbf{y}_i\|_1 + \lambda_{\text{disc}}\, d_{\text{disc}}\,(\mathbf{y}_i, \boldsymbol{\theta}) \Big]. \quad (23)$$

To resolve this problem, we implement a two-fold approach. Initially, we fix the group index and determine an estimated candidate for the nonlinear transform representation. In the second stage, we evaluate the candidate representations against the predefined group parameters to assign each data point to its most fitting group.

Candidate NT Representation Estimation: During the first phase, observe that for each pair $[\boldsymbol{\tau}_c, \boldsymbol{\nu}_c]$, (23) reduces to the following constraint projection problem:

$$(\mathsf{P_S}): \widehat{\mathbf{q}}_{i,c} = \arg \min_{\mathbf{y}_i} \frac{1}{2} \|\mathbf{u}_i - \mathbf{q}_{i,c}\|_2^2 + \lambda_1 \mathbf{1}^T |\mathbf{q}_{i,c}| + \lambda_{\text{disc}}\, d_{\text{disc}}\big(\mathbf{q}_{i,c}, [\boldsymbol{\nu}_c, \boldsymbol{\tau}_c]\big). \quad (24)$$

We provide a detailed exposition that $(\mathsf{P_S})$ per $\mathbf{y}$ reduces to $\min_{\mathbf{y}} \frac{1}{2}\|\mathbf{u} - \mathbf{y}\|_2^2 + \mathbf{g}^T |\mathbf{y}| + \mathbf{s}^T (\mathbf{y} \odot \mathbf{y}) + \lambda_1 \mathbf{1}^T |\mathbf{y}|$ and per $[\boldsymbol{\tau}_c, \boldsymbol{\nu}_c]$ has a closed-form solution as:

$$\widehat{\mathbf{q}}_{i,c} = \psi\,(\mathbf{u}_i) := \text{sign}(\mathbf{u}_i) \odot \max(|\mathbf{u}_i| - \mathbf{g}_i - \lambda_1 \mathbf{1}, \mathbf{0}) \oslash \mathbf{k}_c, \quad (25)$$

where $\mathbf{k}_c = (1 + 2\,\mathbf{s}_c)$ and $\mathbf{y} = \mathbf{y}_i$, $\mathbf{s}_c^T = \lambda_{\text{disc}} (\boldsymbol{\tau}_c \odot \boldsymbol{\tau}_c)^T$, $\mathbf{g}_i^T = \lambda_{\text{disc}} \left(\mathbf{h}_1^T - \mathbf{h}_2^T\right)$, $\mathbf{h}_1^T |\mathbf{y}| = \left(\mathbf{y}^+\right)^T \boldsymbol{\tau}_c^+ + \left(\mathbf{y}^-\right)^T \boldsymbol{\tau}_c^-$, $\mathbf{h}_2^T |\mathbf{y}| = \left(\mathbf{y}^+\right)^T \boldsymbol{\nu}_c^- + \left(\mathbf{y}^-\right)^T \boldsymbol{\nu}_c^+$, $\mathbf{h}_1 = \max(\boldsymbol{\tau}_c, \mathbf{0}) \odot \text{sign}(\max(\mathbf{y}, \mathbf{0})) + \max(-\boldsymbol{\tau}_c, \mathbf{0}) \odot \text{sign}(\max(-\mathbf{y}, \mathbf{0}))$, $\mathbf{h}_2 = \max(\boldsymbol{\nu}_c, \mathbf{0}) \odot \text{sign}(\max(\mathbf{y}, \mathbf{0})) + \max(-\boldsymbol{\nu}_c, \mathbf{0}) \odot \text{sign}(\max(-\mathbf{y}, \mathbf{0}))$.

Proof: Given the available database $\mathbf{X}$ and the current estimate of the linear map $\mathbf{W}$, the representation estimation problem is formulated in (23). Let $\mathbf{y} = \mathbf{y}_i$ and $\mathbf{x} = \mathbf{x}_i$, the above problem per single $\mathbf{y}_i$ reduces to:

$$\min_{\mathbf{y}} \frac{1}{2} \|\mathbf{W}\mathbf{x} - \mathbf{y}\|_2^2 + \lambda_1 \mathbf{1}^T |\mathbf{y}| + \lambda_{\text{disc}} \Big((\mathbf{y}^+)^T \boldsymbol{\tau}_c^+ + (\mathbf{y}^-)^T \boldsymbol{\tau}_c^- + (\boldsymbol{\tau}_c \odot \boldsymbol{\tau}_c)^T (\mathbf{y} \odot \mathbf{y}) - \left[(\mathbf{y}^+)^T \boldsymbol{\nu}_c^+ + (\mathbf{y}^-)^T \boldsymbol{\nu}_c^- \right] \Big). \quad (26)$$

By denoting:

$$\mathbf{u} = \mathbf{W}\mathbf{x} \quad (27)$$

$$\mathbf{h}_1^T |\mathbf{y}| = \left(\mathbf{y}^+\right)^T \boldsymbol{\tau}^+ + \left(\mathbf{y}^-\right)^T \boldsymbol{\tau}^-, \quad (28)$$

$$\mathbf{h}_2^T |\mathbf{y}| = \left(\mathbf{y}^+\right)^T \boldsymbol{\nu}^+ + \left(\mathbf{y}^-\right)^T \boldsymbol{\nu}^-, \quad (29)$$

$$\mathbf{s}_c^T = \lambda_{\text{disc}} (\boldsymbol{\tau}_c \odot \boldsymbol{\tau}_c)^T, \quad (30)$$

$$\mathbf{g}_i^T = \lambda_{\text{disc}} \left(\mathbf{h}_1^T - \mathbf{h}_2^T\right), \quad (31)$$

where $\mathbf{h}_1 = \max(\boldsymbol{\tau}_c, \mathbf{0}) \odot \text{sign}(\max(\mathbf{y}, \mathbf{0})) + \max(-\boldsymbol{\tau}_c, \mathbf{0}) \odot \text{sign}(\max(-\mathbf{y}, \mathbf{0}))$, $\mathbf{h}_2 = \max(\boldsymbol{\nu}_c, \mathbf{0}) \odot \text{sign}(\max(\mathbf{y}, \mathbf{0})) + \max(-\boldsymbol{\nu}_c, \mathbf{0}) \odot \text{sign}(\max(-\mathbf{y}, \mathbf{0}))$, the problem is simplified as:

$$\min_{\mathbf{y}} \frac{1}{2} \|\mathbf{y} - \mathbf{u}\|_2^2 + \mathbf{g}_i^T |\mathbf{y}| + \mathbf{s}_c^T (\mathbf{y} \odot \mathbf{y}) + \lambda_1 \mathbf{1}^T |\mathbf{y}|. \quad (32)$$

Taking the first order derivative w.r.t $\mathbf{y}$ and using the sign magnitude decomposition of $\mathbf{y} = \text{sign}(\mathbf{y}) \odot |\mathbf{y}|$ and $\mathbf{u} = \text{sign}(\mathbf{u}) \odot |\mathbf{u}|$ gives:

$$\begin{aligned} &\text{sign}(\mathbf{y}) \odot |\mathbf{y}| \odot (\mathbf{1} + 2\mathbf{s}_c) - \text{sign}(\mathbf{u}) \odot |\mathbf{u}| + \lambda_1 \text{sign}(\mathbf{y}) \\ &\quad + \lambda_{\text{disc}} \big(\text{sign}\left(\mathbf{y}^+\right) \odot \boldsymbol{\tau}_c^+ + \text{sign}\left(\mathbf{y}^-\right) \odot \boldsymbol{\tau}_c^- \\ &\quad - \text{sign}(\mathbf{y}^+) \odot \boldsymbol{\nu}_c^+ - \text{sign}(\mathbf{y}^-) \odot \boldsymbol{\nu}_c^- \big) = \mathbf{0}. \end{aligned} \quad (33)$$

Let $\text{sign}(\mathbf{y}) = \text{sign}(\mathbf{u})$, then by Hadamard multiplying from the left side by $\text{sign}(\mathbf{u})$ and noting that $\text{sign}(\mathbf{u}) \odot \text{sign}\left(\mathbf{u}^+\right) = \text{sign}\left(\mathbf{u}^+\right)$, $\text{sign}(\mathbf{u}) \odot \text{sign}\left(\mathbf{u}^-\right) = \text{sign}\left(\mathbf{u}^-\right)$ and taking into account the positive values for magnitude we have:

$$\begin{aligned} &|\mathbf{y}| \odot (\mathbf{1} + 2\mathbf{s}_c) \\ &= \max \Big(|\mathbf{u}| - \lambda_{\text{disc}} \big(\text{sign}\left(\mathbf{u}^+\right) \odot \boldsymbol{\tau}_c^+ - \text{sign}\left(\mathbf{u}^-\right) \\ &\quad \odot \boldsymbol{\tau}_c^- - \text{sign}\left(\mathbf{u}^+\right) \odot \boldsymbol{\nu}_c^+ + \text{sign}\left(\mathbf{u}^-\right) \odot \boldsymbol{\nu}_c^- \big) - \lambda_1 \mathbf{1}, \mathbf{0} \Big), \end{aligned} \quad (34)$$

Note that $\mathbf{h}_1 = \text{sign}\left(\mathbf{u}^+\right) \odot \boldsymbol{\tau}^+ - \text{sign}\left(\mathbf{u}^-\right) \odot \boldsymbol{\tau}^-$ and $\mathbf{h}_2 = \text{sign}\left(\mathbf{u}^+\right) \odot \boldsymbol{\nu}^+ - \text{sign}\left(\mathbf{u}^-\right) \odot \boldsymbol{\nu}^-$. Therefore, the closed-form solution of problem (26) is given as:

$$\mathbf{y} = \text{sign}\,(\mathbf{u}) \odot \max\,(|\mathbf{u}| - \mathbf{g}_i - \lambda_1 \mathbf{1}, \mathbf{0}) \oslash (\mathbf{1} + 2\mathbf{s}_c)\,, \quad (35)$$

which completes the proof. □

Assignment: During the second step, given all the candidate representations $\mathbf{u}_{i,c}, c \in [1, \cdots, C]$, we evaluate the scores using the composite function:

$$\mathsf{S}(\mathbf{q}_{i,c}, [\boldsymbol{\nu}_c, \boldsymbol{\tau}_c]) = \mathsf{Sim}\left(\mathbf{q}_{i,c}, \boldsymbol{\tau}_c\right) + \mathsf{Stg}\left(\mathbf{y}_i, \boldsymbol{\tau}_c\right) + \mathsf{Sim}\left(\mathbf{y}_i, \boldsymbol{\nu}_c\right).$$

Based on these scores, we assign the data point $\mathbf{x}_i$ to a group characterized by the index $\widehat{c} \in [1, \cdots, C]$. Concurrently, we choose the representation $\mathbf{q}_{i,\widehat{c}}$, that results in a minimal evaluation score. Specifically, this assignment is governed by:

$$\widehat{c} = \arg \min_{c} \mathsf{S}(\mathbf{q}_{i,c}, \boldsymbol{\nu}_c, \boldsymbol{\tau}_c), \quad (36)$$

$$\widehat{\mathbf{y}}_i = \mathbf{q}_{i,\widehat{c}}\,. \quad (37)$$

2) GROUP PARAMETERS θ UPDATE

Given the current estimate of the linear map $\mathbf{W}$ and representations $\mathbf{y}_i$, we can reformulate the problem (22) as follows:

$$\widehat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \sum_{i=1}^{K} \big[\lambda_{\mathsf{disc}}\, d_{\mathsf{disc}}(\mathbf{y}_i, \boldsymbol{\theta}) + \lambda_{\mathsf{adj}}\, d_{\mathsf{adj}}(\mathbf{u}_i, \boldsymbol{\theta})\big) + \lambda_{\mathsf{p}}\, d_{\mathsf{p}}(\boldsymbol{\theta})\,\big]. \tag{38}$$

With the reformulated problem in place, our solution strategy employs a distinct two-phase procedure. In the first phase, we focus on the update mechanism for single parameters $\boldsymbol{\nu}_c$. Subsequently, the second phase addresses the intricacies of the $\boldsymbol{\tau}_c$ parameters update. The division into these phases is derived from the inherent structure of the problem, ensuring that each component is accurately captured and updated. The specifics of each phase are presented in the following sections.

Single $\boldsymbol{\nu}_c$ Parameter Update: Given $\{\boldsymbol{\theta}_1, \cdots, \boldsymbol{\theta}_{c-1}, \boldsymbol{\theta}_{c+1}, \cdots, \boldsymbol{\theta}_C\}$, problem (38) per $\boldsymbol{\nu}_c$ reduces to

$$(\mathsf{P}_{\mathsf{T}_1}) : \widehat{\boldsymbol{\nu}}_c = \arg\min_{\boldsymbol{\nu}_c} \sum_{i\in\mathcal{I}_c} \frac{1}{2}\|\mathbf{u}_i - \boldsymbol{\nu}_c - \boldsymbol{\tau}_c\|_2^2 + \lambda_{\mathsf{disc}} \sum_{d\in\{1,\cdots,C\}\setminus c} d_{\mathsf{disc}}\left(\boldsymbol{\nu}_c, \boldsymbol{\theta}_d\right), \tag{39}$$

where $\mathcal{I}_c$ is a set, which contains all indexes i for data $\mathbf{x}_i$ that were assigned to group indexed by c. The solution for $(\mathsf{P}_{\mathsf{T}_1})$ aligns structurally with the solution from (25). Notably, they differ in their respective thresholding and normalization vectors.

Proof: Given $\boldsymbol{\theta}_s = \{\boldsymbol{\theta}_{1\setminus c}, \boldsymbol{\theta}_2\}$, problem (38) per $\boldsymbol{\nu}_{c1}$ reduces to:

$$\begin{aligned}\min_{\boldsymbol{\nu}_{c1}} \sum_m \frac{1}{2}\|\mathbf{W}\mathbf{x}_{c1,m} - \boldsymbol{\nu}_{c1} - \boldsymbol{\tau}_{c1}\|_2^2 \\ + \lambda_{\mathsf{disc}} \sum_m \big(\mathsf{Sim}\left(\mathbf{y}_{c1,m}, \boldsymbol{\tau}_{c1}\right) \\ - \mathsf{Sim}\left(\mathbf{y}_{c1,m}, \boldsymbol{\nu}_{c1}\right) + \mathsf{Stg}\left(\mathbf{y}_{c1,m}, \boldsymbol{\tau}_{c1}\right)\big) \\ + \lambda_{\mathsf{p}} \sum_{c\neq c1} \big(\mathsf{Sim}\left(\boldsymbol{\nu}_{c1}, \boldsymbol{\tau}_c\right) - \mathsf{Sim}\left(\boldsymbol{\nu}_{c1}, \boldsymbol{\nu}_c\right) + \mathsf{Stg}\left(\boldsymbol{\nu}_{c1}, \boldsymbol{\tau}_c\right)\big).\end{aligned} \tag{40}$$

Let $\mathbf{v} = \boldsymbol{\nu}_{c1}$ and $\mathbf{u} = \sum_m \mathbf{W}\mathbf{x}_{c1,m} - \boldsymbol{\tau}_{c1}$. The first order derivative with respect to $\mathbf{v}$ is:

$$\begin{aligned}(M\mathbf{v} - \mathbf{u}) - \lambda_{\mathsf{disc}} \sum_m \left(\mathrm{sign}\left(\mathbf{v}^+\right) \odot \mathbf{y}_{c1,m}^+ - \mathrm{sign}\left(\mathbf{v}^-\right) \odot \mathbf{y}_{c1,m}^-\right) \\ + \lambda_{\mathsf{p}} \sum_{c\neq c1} \Big(\mathrm{sign}\left(\mathbf{v}^+\right) \odot \boldsymbol{\tau}_c^+ - \mathrm{sign}\left(\mathbf{v}^-\right) \odot \boldsymbol{\tau}_c^- \\ - \mathrm{sign}\left(\mathbf{v}^+\right) \odot \boldsymbol{\nu}_c^+ + \mathrm{sign}\left(\mathbf{v}^-\right) \odot \boldsymbol{\nu}_c^- \\ + 2\,\mathbf{v} \odot (\boldsymbol{\tau}_c \odot \boldsymbol{\tau}_c)\Big).\end{aligned} \tag{41}$$

Denote:

$$\begin{aligned}\mathbf{h}_y^T|\mathbf{v}| &= \left(\mathbf{v}^+\right)^T \mathbf{y}_{c1,m}^+ - \left(\mathbf{u}^-\right)^T \mathbf{y}_{c1,m}^-, \\ \mathbf{h}_1^T|\mathbf{v}| &= \left(\mathbf{v}^+\right)^T \boldsymbol{\tau}_c^+ - \left(\mathbf{u}^-\right)^T \boldsymbol{\tau}_c^-, \\ \mathbf{h}_2^T|\mathbf{v}| &= \left(\mathbf{v}^+\right)^T \boldsymbol{\nu}_c^+ - \left(\mathbf{u}^-\right)^T \boldsymbol{\nu}_c^-, \\ \mathbf{s}_c^T &= \lambda_{\mathsf{p}}(\boldsymbol{\tau}_c \odot \boldsymbol{\tau}_c)^T, \\ \mathbf{g}_c^T &= \lambda_{\mathsf{p}}\left(\mathbf{h}_1^T - \mathbf{h}_2^T\right), \\ \mathbf{p}_c^T &= \lambda_{\mathsf{disc}}\mathbf{h}_y^T,\end{aligned} \tag{42}$$

where

$$\begin{aligned}\mathbf{h}_y = \sum_m \max\left(\mathbf{y}_{c1,m}, \mathbf{0}\right) \odot \mathrm{sign}\left(\max\left(\mathbf{u}, \mathbf{0}\right)\right) \\ - \sum_m \max\left(-\mathbf{y}_{c1,m}, \mathbf{0}\right) \odot \mathrm{sign}\left(\max\left(-\mathbf{u}, \mathbf{0}\right)\right),\end{aligned} \tag{43}$$

$$\begin{aligned}\mathbf{h}_1 = \sum_{c\neq c1} \max\left(\boldsymbol{\tau}_c, \mathbf{0}\right) \odot \mathrm{sign}\left(\max\left(\mathbf{u}, \mathbf{0}\right)\right) \\ - \sum_{c\neq c1} \max\left(-\boldsymbol{\tau}_c, \mathbf{0}\right) \odot \mathrm{sign}\left(\max\left(-\mathbf{u}, \mathbf{0}\right)\right),\end{aligned} \tag{44}$$

$$\begin{aligned}\mathbf{h}_2 = \sum_{c\neq c1} \max\left(\boldsymbol{\nu}_c, \mathbf{0}\right) \odot \mathrm{sign}\left(\max\left(\mathbf{u}, \mathbf{0}\right)\right) \\ - \sum_{c\neq c1} \max\left(-\boldsymbol{\nu}_c, \mathbf{0}\right) \odot \mathrm{sign}\left(\max\left(-\mathbf{u}, \mathbf{0}\right)\right),\end{aligned} \tag{45}$$

Take the magnitude decomposition $\mathbf{v} = \mathrm{sign}\,(\mathbf{v}) \odot |\mathbf{v}|$ and $\mathbf{u} = \mathrm{sign}\,(\mathbf{u}) \odot |\mathbf{u}|$ and let $\mathrm{sign}\,(\mathbf{v}) = \mathrm{sign}\,(\mathbf{u})$. By Hadamard multiplying from the left side by $\mathrm{sign}\,(\mathbf{u})$ and noting that the magnitude can be only positive, the closed-form solution is simplified as:

$$\mathbf{v} = \mathrm{sign}\,(\mathbf{u}) \odot \max\left(|\mathbf{u}| + \mathbf{p}_c - \mathbf{g}_c, \mathbf{0}\right) \oslash (M + 2\mathbf{s}_c), \tag{46}$$

which completes the proof. □

Single $\boldsymbol{\tau}_c$ Parameter Update: Given $\{\boldsymbol{\theta}_1, \cdots, \boldsymbol{\theta}_{c-1}, \boldsymbol{\theta}_{c+1}, \cdots, \boldsymbol{\theta}_C\}$, problem (38) per $\boldsymbol{\tau}_c$ reduces to:

$$(\mathsf{P}_{\mathsf{T}_2}) : \widehat{\boldsymbol{\tau}}_c = \arg\min_{\boldsymbol{\nu}_c} \sum_{i\in\mathcal{I}_c} \frac{1}{2}\|\mathbf{u}_i - \boldsymbol{\nu}_c - \boldsymbol{\tau}_c\|_2^2 + \lambda_{\mathsf{disc}} \sum_{d\in\{1,\cdots,C\}\setminus c} d_{\mathsf{disc}}\left(\boldsymbol{\tau}_c, \boldsymbol{\theta}_d\right). \tag{47}$$

Similar to the previous parameter update, the solution for this problem adheres structurally to (25). Nevertheless, the distinction between the two lies in their thresholding and normalization vectors.

3) LINEAR MAP $\mathbf{W}$ UPDATE

Consider a given data set $\mathbf{X}$, its corresponding representations $\mathbf{Y}$, and the group membership parameters $\boldsymbol{\theta}$. With these, the problem in (22) can be restructured specifically for the linear map $\mathbf{W}$ update as:

$$\widehat{\mathbf{W}} = \arg\min_{\mathbf{W}} \frac{1}{2}\|\mathbf{W}\mathbf{X} - \mathbf{R}\|_F^2 + \frac{\lambda_{\mathsf{W}_1}}{2}\|\mathbf{W}\|_F^2 + \frac{\lambda_{\mathsf{W}_2}}{2}\|\mathbf{W}\mathbf{W}^T - \mathbf{I}\|_F^2 - \frac{\lambda_{\mathsf{W}_3}}{2}\log|\det \mathbf{W}^T\mathbf{W}|, \tag{48}$$

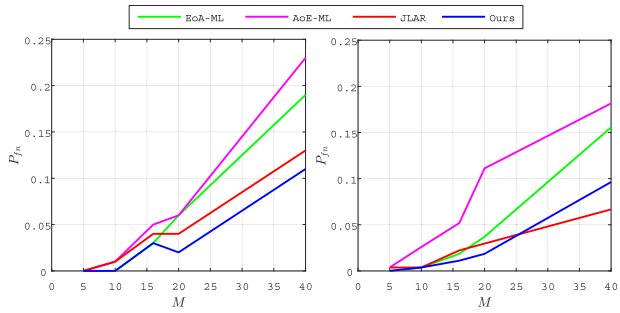


FIGURE 4. Performances on group verification for varying group size M. P_{fn} at $P_{\text{fp}} = 0.05$ for CFP (left) and LFW (right).

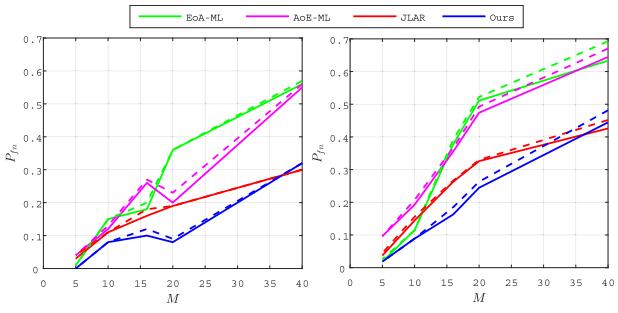


FIGURE 5. Performances on group identification for varying group size M on CFP (left) and LFW (right). P_{fn} at $P_{\text{fp}} = 0.05$ for the first step (solid line) and P_ϵ for the second step of group identification (dashed).

where we denote $\mathbf{R} = [\mathbf{r}_1, \cdots, \mathbf{r}_L]$ with $\mathbf{r}_i = \mathbf{y}_i - \lambda_{\text{adj}}(\boldsymbol{\nu}_c + \boldsymbol{\tau}_c)$. Furthermore, λ_{W_1}, λ_{W_2}, and λ_{W_3} are inversely related to the scaling parameters β_{W_1}, β_{W_2}, and β_{W_3}, respectively. The solution to this problem aligns with the method proposed in [15].

VII. NUMERICAL EVALUATION

In this section, we assess the efficacy of our proposed scheme through its application in face recognition across two scenarios. We then compare its performance against established methods: EoA-ML, AoE-ML [10], and JLAR [12]. While EoA-ML and AoE-ML enroll K individuals across C random groups without joint optimization, both our proposed method and JLAR are distinct in that they jointly learn group assignments and representations.

A. FACE DATASETS

We extract face descriptors using a network that has been pre-trained on the VGG-Face architecture. These descriptors are then processed through PCA to reduce their dimensionality. Subsequently, the reduced descriptors are L_2-normalized, resulting in standardized feature vectors of size $N = 1,024$.

1) CFP

The Celebrities in Frontal-Profile (CFP) database encompasses 500 individuals, each represented by 10 frontal and 4 profile images captured in unconstrained environments. From this collection, we exclusively utilize $K = 400$ frontal images. For the purpose of defining impostors, a subset of $K_q = 100$ individuals is randomly chosen.

2) LFW

The Labeled Faces in the Wild (LFW) dataset includes 13,233 facial images collected from the internet. In our study, we use their pre-aligned versions. The enrollment set is derived from $K = 1680$ individuals, each represented by at least two images within the LFW database. For each individual, one random template is enrolled as $\mathbf{x}_i$. A separate group of $K_q = 263$ individuals is randomly selected from the dataset to represent impostors.

B. SCENARIO #1: GROUP VERIFICATION

Consider a scenario where a user claims their membership to group c. We define two hypotheses: $\mathcal{H}_1$ affirms the user's claim, while $\mathcal{H}_0$ refutes it, categorizing the user as an impostor. The user's signature, denoted by $\mathbf{q}$, is transformed into $\mathbf{q}_c$ using the function $\mathcal{T}_{\mathcal{P}_c}(\mathbf{q})$, where $\mathcal{P}_c$ are nonlinear transform parameters. Once embedded, both the signature $\mathbf{q}_c$, and its associated group c are transmitted to the system. The system then contrasts $\mathbf{q}_c$ with the group's representative parameters $\boldsymbol{\theta}_c$ to decide on acceptance ($t = 1$) or rejection ($t = 0$). This constitutes a binary hypothesis test with two potential error outcomes: The false positive rate, denoted as $P_{\text{fp}} := \mathbb{P}(t = 1|\mathcal{H}_0)$, and the false negative rate, represented as $P_{\text{fn}} := \mathbb{P}(t = 0|\mathcal{H}_1)$. The figure of merit is P_{fn} when $P_{\text{fp}} = 0.05$.

Figure. 4 provides a comparative analysis, illustrating the enhanced efficiency of our scheme in group verification relative to other baseline methods. Complementing this, TABLE 1 offers a detailed quantitative evaluation for four specific group sizes $M = 5, 10, 20, 40$, with all values rounded to three decimal places. Particularly, our method demonstrates marked benefits on the CFP dataset. Furthermore, the observed improvements with both JLAR and our scheme become considerably more noticeable for larger group sizes, suggesting that grouping similar vectors leads to minimal information loss.

C. SCENARIO #2: GROUP IDENTIFICATION

This scenario addresses open set identification, wherein a querying user may either be enrolled or categorized as an impostor. The system's identification procedure bifurcates into two phases. In the first phase, the system determines whether the user is enrolled. While this mirrors the verification described previously, the pivotal distinction lies in the *group's indeterminacy*. The system calculates the min-max score for each group, denoted as $\delta_c = \mathsf{S}(\mathbf{q}_c, \boldsymbol{\theta}_c)$ for all $c \in [C]$. An acceptance decision ($t = 1$) is rendered if the lowest score among these C scores is below a specified threshold τ. The figure of merit is P_{fn} when $P_{\text{fp}} = 0.05$. Upon achieving $t = 1$ in the first phase, the system advances to the second. Here, the system deduces the likely group

TABLE 1. Performances on group verification for varying group size M. P_{fn} at $P_{\text{fp}} = 0.05$ for CFP (up) and LFW (down).

	$M=5$				$M=10$				$M=20$				$M=40$			
	EoA-ML	AoE-ML	JLAR	NSTL (ours)	EoA-ML	AoE-ML	JLAR	NSTL (ours)	EoA-ML	AoE-ML	JLAR	NSTL (ours)	EoA-ML	AoE-ML	JLAR	NSTL (ours)
P_{fn} (CFP)	0.000	0.000	0.000	0.000	0.012	0.012	0.000	0.000	0.060	0.060	0.039	0.022	0.189	0.229	0.128	0.109
P_{fn} (LFW)	0.005	0.005	0.000	0.000	0.005	0.027	0.005	0.005	0.037	0.116	0.031	0.021	0.154	0.181	0.068	0.098

TABLE 2. Performances on group identification for varying group size M on CFP (up) and LFW (down). P_{fn} at $P_{\text{fp}} = 0.05$ for the first step of group identification.

	$M=5$				$M=10$				$M=20$				$M=40$			
	EoA-ML	AoE-ML	JLAR	NSTL (ours)	EoA-ML	AoE-ML	JLAR	NSTL (ours)	EoA-ML	AoE-ML	JLAR	NSTL (ours)	EoA-ML	AoE-ML	JLAR	NSTL (ours)
P_{fn} (CFP)	0.011	0.049	0.042	0.000	0.151	0.113	0.110	0.086	0.368	0.201	0.193	0.088	0.554	0.543	0.298	0.312
P_{fn} (LFW)	0.032	0.098	0.043	0.028	0.110	0.192	0.145	0.090	0.510	0.484	0.317	0.246	0.631	0.639	0.420	0.437

TABLE 3. The Detection and Identification Rate (DIR) vs. P_{fp} on CFP for group size $M = 16$ (up) and $M = 10$ (down).

	$P_{\text{fp}}=0.05$				$P_{\text{fp}}=0.2$				$P_{\text{fp}}=0.4$			
	EoA-ML	AoE-ML	JLAR	NSTL (ours)	EoA-ML	AoE-ML	JLAR	NSTL (ours)	EoA-ML	AoE-ML	JLAR	NSTL (ours)
DIR ($M=16$)	0.75	0.71	0.75	0.86	0.89	0.81	0.87	0.92	0.90	0.92	0.93	0.93
DIR ($M=10$)	0.85	0.84	0.86	0.90	0.95	0.94	0.93	0.97	0.96	0.95	0.94	0.99

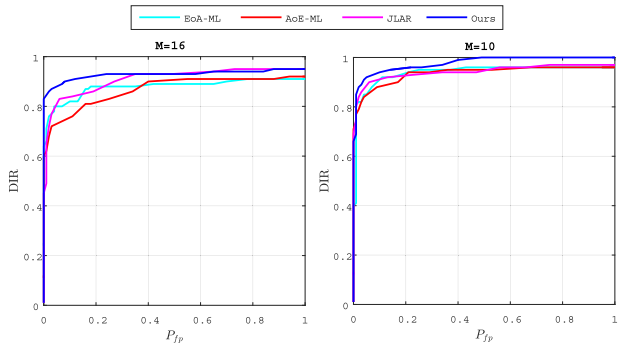


FIGURE 6. The Detection and Identification Rate (DIR) vs. P_{fp} on CFP.

membership, represented by $\widehat{c} = \arg\min_{c\in[C]} \delta_c$. Two pivotal figures of merit (performance metrics) of this phase: the error probability, $P_\epsilon := \mathbb{P}(\widehat{c} \neq c)$, and the Detection and Identification Rate (DIR), defined as $\mathsf{DIR} := (1 - P_\epsilon)(1 - P_{\text{fn}})$.

The performance metrics for the group identification scenario are delineated in Fig 5. These metrics highlight the enhancements that our proposed scheme offers in this scenario. Table 2 provides a detailed quantitative comparison for four group sizes $M = 5, 10, 20, 40$, with all values rounded to three decimal places, corresponding to the results depicted in Figure 5. Additionally, Figure 6 shows the impact of group size on the Detection and Identification Rate (DIR) for the CFP dataset. Evidently, packing more templates into one group will cause performance deterioration. Table 3 provides a quantitative comparison for three false positive ratea $P_{\text{fp}} = 0.05, 0.2, 0.4$, also rounded to two decimal places, corresponding to the results depicted in Figure 6.

FIGURE 7. Examples of group identification on LFW. black frames (enrolled samples), green frames (successful queries) and red frames (failed queries).

Figure 7 presents some enrolled and querying faces of LFW dataset. All the failed identification examples show a change of lighting, pose or expression or occlusion.

VIII. CONCLUSION

We introduce a new group membership methodology which achieves two primary objectives: (i) it jointly learns nonlinear transform representations, incorporating prior information, and (ii) it determines group representatives utilizing a maximum likelihood approach grounded in functional measures. We have further proposed an efficient algorithm tailored for the optimal estimation of model parameters. Evaluations

of our proposed framework were conducted on image databases, underscoring its applicability and proficiency in face verification and identification tasks.

ACKNOWLEDGMENT

Implementation codes available at (https://gitlab.idiap.ch/biometric/code.group_membership_verification).

REFERENCES

[1] S. Schechter, T. Parnell, and A. Hartemink, "Anonymous authentication of membership in dynamic groups," in *Proc. Int. Conf. Financial Cryptography*, 1999, pp. 184–195.
[2] Z. Sivic, "Video Google: A text retrieval approach to object matching in videos," in *Proc. 9th IEEE Int. Conf. Comput. Vis.*, Oct. 2003, pp. 1470–1477.
[3] H. Jégou, F. Perronnin, M. Douze, J. Sánchez, P. Pérez, and C. Schmid, "Aggregating local image descriptors into compact codes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 9, pp. 1704–1716, Sep. 2012.
[4] E. Mohedano, K. McGuinness, N. E. O'Connor, A. Salvador, F. Marques, and X. Giro-I-Nieto, "Bags of local convolutional features for scalable instance search," in *Proc. ACM Int. Conf. Multimedia Retr.*, Jun. 2016, pp. 327–331.
[5] R. Arandjelovic, P. Gronat, A. Torii, T. Pajdla, and J. Sivic, "NetVLAD: CNN architecture for weakly supervised place recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 5297–5307.
[6] F. Radenovic, G. Tolias, and O. Chum, "Fine-tuning CNN image retrieval with no human annotation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 7, pp. 1655–1668, Jul. 2019.
[7] Y. Zhong, R. Arandjelovic, and A. Zisserman, "Compact deep aggregation for set retrieval," in *Proc. Eur. Conf. Comput. Vis. (ECCV) Workshops*, Sep. 2018, pp. 1–27.
[8] Y. Zhong, R. Arandjelović, and A. Zisserman, "Ghostvlad for set-based face recognition," in *Proc. Asian Conf. Comput. Vis.*, Perth, WA, Australia. Cham, Switzerland: Springer, Dec. 2018, pp. 35–50.
[9] M. Gheisari, T. Furon, L. Amsaleg, B. Razeghi, and S. Voloshynovskiy, "Aggregation and embedding for group membership verification," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Brighton, U.K., May 2019, pp. 2592–2596.
[10] M. Gheisari, T. Furon, and L. Amsaleg, "Privacy preserving group membership verification and identification," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2019, pp. 74–82.
[11] M. Gheisari, T. Furon, and L. Amsaleg, "Group membership verification with privacy: Sparse or dense?" in *Proc. IEEE Int. Workshop Inf. Forensics Secur. (WIFS)*, Dec. 2019, pp. 1–7.
[12] M. Gheisari, T. Furon, and L. Amsaleg, "Joint learning of assignment and representation for biometric group membership," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2020, pp. 2922–2926.
[13] S. Ravishankar and Y. Bresler, "Learning sparsifying transforms," *IEEE Trans. Signal Process.*, vol. 61, no. 5, pp. 1072–1086, Mar. 2013.
[14] S. Ravishankar and Y. Bresler, "Sparsifying transform learning with efficient optimal updates and convergence guarantees," *IEEE Trans. Signal Process.*, vol. 63, no. 9, pp. 2389–2404, May 2015.
[15] D. Kostadinov, S. Voloshynovskiy, and S. Ferdowsi, "Learning overcomplete and sparsifying transform with approximate and exact closed form solutions," in *Proc. 7th Eur. Workshop Vis. Inf. Process. (EUVIP)*, Nov. 2018, pp. 1–6.
[16] B. Razeghi, S. Voloshynovskiy, D. Kostadinov, and O. Taran, "Privacy preserving identification using sparse approximation with ambiguization," in *Proc. IEEE Workshop Inf. Forensics Secur. (WIFS)*, Rennes, France, Dec. 2017, pp. 1–6.
[17] B. Razeghi, S. Voloshynovskiy, S. Ferdowsi, and D. Kostadinov, "Privacy-preserving identification via layered sparse code design: Distributed servers and multiple access authorization," in *Proc. 26th Eur. Signal Process. Conf. (EUSIPCO)*, Rome, Rome, Italy, Sep. 2018, pp. 2578–2582.
[18] B. Razeghi and S. Voloshynovskiy, "Privacy-preserving outsourced media search using secure sparse ternary codes," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2018, pp. 1992–1996.
[19] D. Kostadinov, B. Razeghi, S. Rezaeifar, and S. Voloshynovskiy, "Supervised joint nonlinear transform learning with discriminative-ambiguous prior for generic privacy-preserved features," in *Proc. 53rd Annu. Conf. Inf. Sci. Syst. (CISS)*, Mar. 2019, pp. 1–6.

BEHROOZ RAZEGHI (Senior Member, IEEE) received the M.Sc. degree in electrical engineering (communications) from the Ferdowsi University of Mashhad, in 2014, the M.Sc. degree in mathematics (numerical analysis) from Iran University of Science and Technology, in 2017, and the Ph.D. degree (Hons.) in computer science from the University of Geneva, in 2022. He is currently a Postdoctoral Researcher with the Biometric Security & Privacy Group, Idiap Research Institute. He was a member of the Stochastic Information Processing Group, University of Geneva. He has also spent time as a Visiting Research Fellow with Harvard University and as a Visiting Research Scholar with the Imperial College London. His research interests include the application of mathematical, statistical, information theories to machine learning, data privacy, and signal processing, with a particular emphasis on development of trustworthy AI technologies in biometric systems and healthcare problems.

MARZIEH GHEISARI received the Ph.D. degree from the Inria Research Center, Rennes, in 2021. She is currently a Postdoctoral Researcher with the Computational Bioimaging and Bioinformatics Group, École Normale Supérieure/PSL University. She was part of the LinkMedia Research Team, Inria Research Center, focused on multimedia information extraction and representation. Her doctoral research was dedicated to designing privacy-preserving schemes for biometric verification. Currently, her postdoctoral work is centered on advancing techniques in image and video synthesis and computational bioimaging.

AMIR ATASHIN received the bachelor's degree in software engineering and the master's degree in computer science from the Ferdowsi University of Mashhad, Iran. He is currently pursuing the Engineering Doctorate degree in data science with Eindhoven University of Technology. His research interests include AI and ML, areas which have guided his academic and research endeavors. His commitment to exploring AI and ML is aimed at contributing to practical solutions and advancements in the field.

DIMCHE KOSTADINOV received the B.Sc. degree in electrical engineering, major: computer science, information technology and automation from the Faculty of Electrical Engineering and Information Technologies, Ss. Cyril and Methodius University, Skopje, North Macedonia, the M.Sc. degree in electrical engineering and information technologies, major: digital signal processing, and the Ph.D. degree (Hons.) in computer science from the University of Geneva, in 2019. He is currently a Senior ML Researcher with Sony Advanced Visual Sensing AG, Zurich, Switzerland. He was a member of the Stochastic Information Processing Group, University of Geneva.

SÉBASTIEN MARCEL (Senior Member, IEEE) received the Ph.D. degree in signal processing from CNET, the research center of France Telecom (now Orange Laboratories), Université de Rennes I, France, in 2000. He heads the Biometrics Security and Privacy Group, Idiap Research Institute, Switzerland, and conducts research on face recognition, speaker recognition, vein recognition, attack detection (presentation attacks, morphing attacks, and deepfakes), and template protection. He is currently a Professor with the University of Lausanne (School of Criminal Justice) and a Lecturer with the École Polytechnique Fédérale de Lausanne (EPFL). He is also the Director of the Swiss Center for Biometrics Research and Testing, which conducts certifications of biometric products.

DENIZ GÜNDÜZ (Fellow, IEEE) received the B.S. degree in electrical and electronics engineering from METU, Ankara, Turkey, in 2002, and the M.S. and Ph.D. degrees in electrical engineering from New York University Polytechnic School of Engineering, Brooklyn, NY, USA, in 2004 and 2007, respectively. He is currently a Professor in information processing with the Electrical and Electronic Engineering Department, Imperial College London. Previously, he was a Research Associate with CTTC, Barcelona, Spain, a Consulting Assistant Professor with the Department of Electrical Engineering, Stanford University, and a Postdoctoral Research Associate with the Department of Electrical Engineering, Princeton University. He is also a part-time Faculty Member with the University of Modena and Reggio Emilia, Italy, and has held visiting positions with the University of Padova (2018–2020) and Princeton University (2009–2012). He is a Distinguished Lecturer of the IEEE Information Theory Society (2020-2022). He serves as an Area Editor for IEEE TRANSACTIONS ON INFORMATION THEORY, IEEE TRANSACTIONS ON COMMUNICATIONS, and the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS (JSAC)—Special Series on Machine Learning in Communications and Networks. He is also an Editor of IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS. He served as the Co-Chair for the IEEE Information Theory Society Student Committee, from 2012 to 2015. He served as the General Co-Chair for the 2016 IEEE Information Theory Workshop and was the Co-Chair for the 2012 European School of Information Theory (ESIT). He was a recipient of the IEEE Communications Society—Communication Theory Technical Committee (CTTC) Early Achievement Award, in 2017, Starting Grant of the European Research Council (ERC), the 2014 IEEE Communications Society Best Young Researcher Award for the Europe, Middle East, and Africa Region, and several best paper awards, including the 2007 IEEE International Symposium on Information Theory (ISIT) Best Student Paper Award. His research interests include communications and information theory, machine learning, and privacy.

SLAVA VOLOSHYNOVSKIY (Senior Member, IEEE) received the degree in radio engineering from Lviv Polytechnic Institute, Lviv, Ukraine, in 1993, and the Ph.D. degree in electrical engineering from State University Lvivska Politechnika, Lviv, in 1996. From 1998 to 1999, he was a Visiting Scholar with the University of Illinois at Urbana–Champaign. Since 1999, he has been with the University of Geneva, Switzerland, where he is currently a Professor with the Department of Computer Science and the Head of the Stochastic Information Processing Group. He has coauthored over 350 journals and conference papers in his research areas and holds 15 patents. His current research interests include information-theoretic aspects of self-supervised and generative machine learning, digital data hiding, content fingerprinting, and physical object security. He was a recipient of the Swiss National Science Foundation Professorship Grant, in 2003. He served as an Associate Editor for IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY (2013–2015). He was an elected member of the IEEE Information Forensics and Security Technical Committee (2011–2013), where he was the Area Chair of Information-Theoretic Security and has been an Associate Member, since 2015. He was the General Chair of the ACM Multimedia Security Conference, in 2006, and the Technical Co-Chair of the Workshop on Information Forensics and Security, in 2015. He has served as a consultant to the private industry in the above areas.