Face Sketch Matching via Coupled Deep Transform Learning

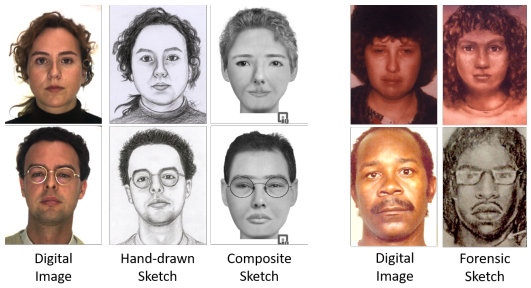
Shruti Nagpal[1], Maneet Singh[1], Richa Singh[1,2], Mayank Vatsa[1,2], Afzel Noore[2], and Angshul Majumdar[1]
[1]IIIT-Delhi, India, [2]West Virginia University
{shrutin, maneets, rsingh, mayank, angshul}@iiitd.ac.in, afzel.noore@mail.wvu.edu

Abstract

Face sketch to digital image matching is an important challenge of face recognition that involves matching across different domains. Current research efforts have primarily focused on extracting domain invariant representations or learning a mapping from one domain to the other. In this research, we propose a novel transform learning based approach termed as DeepTransformer, which learns a transformation and mapping function between the features of two domains. The proposed formulation is independent of the input information and can be applied with any existing learned or hand-crafted feature. Since the mapping function is directional in nature, we propose two variants of DeepTransformer: (i) semi-coupled and (ii) symmetrically-coupled deep transform learning. This research also uses a novel IIIT-D Composite Sketch with Age (CSA) variations database which contains sketch images of 150 subjects along with age-separated digital photos. The performance of the proposed models is evaluated on a novel application of sketch-to-sketch matching, along with sketch-to-digital photo matching. Experimental results demonstrate the robustness of the proposed models in comparison to existing state-of-the-art sketch matching algorithms and a commercial face recognition system.

1. Introduction

Face recognition systems have been evolving over the past few decades, particularly with the availability of large scale databases and access to sophisticated hardware. Large scale face recognition challenges such as MegaFace [13] and Janus [16] further provide opportunities for bridging the gap between unconstrained and constrained face recognition. However, the availability of new devices and applications continuously open new challenges. One such challenging application is matching sketches with digital face photos. In criminal investigations, eyewitnesses provide a first hand account of the event, along with a description of the appearance of the suspect based on their memory. A sketch artist interviews the eyewitness of a particular case and a sketch image of the suspect is created. Such a sketch drawn by an artist is termed as a *hand-drawn sketch*. To eliminate the inter-artist variations and automate the process of sketch generation, law enforcement agencies have started using software generated *composite sketches*. In such cases, the eyewitness is interviewed by an officer and a sketch is created using the drag-and-drop features available in sketch generation tools such as FACES [2], evoFIT [1] and IdentiKit [3]. As shown in Figure 1(a), the information content in the two domains/modalities (sketch and digital image) vary significantly. The digital image is an information-rich representation whereas, the sketch image contains only the edge information and lacks texture details. Figure 1(b) shows real world examples of forensic hand-drawn sketch and corresponding photo pairs. Along with domain differences, variations caused by eyewitness description makes this problem further challenging.



(a) Viewed sketches (b) Forensic sketches

Figure 1: Illustrating the variations in the information content of digital images and different types of sketches.

Traditionally, a sketch image is matched with digital mugshot images for identifying the suspect. The literature is spread across hand-drawn and composite sketch to digital photo matching [30], with algorithms being evaluated [17, 23] on viewed sketches [18, 45, 49]. Viewed sketches are drawn while looking at the digital photos. Such sketches do not reflect real scenario and fail to capture the challenging nature of the problem. Choi *et al.* [8] have established the limitations of viewed sketches and emphasized the need for new databases and algorithms imitating real scenarios.

Sketch	Authors (Year)	Feature Extraction	Classification
Hand-drawn	Bhatt *et al.* [5] (2012)	Proposed MCWLD	Memetically optimized chi-squared distance
	Khan *et al.* [14] (2012)	Facial Self Similarity descriptor	Nearest neighbor classifier
	Mignon *et al.*[26] (2012)	Proposed Cross modal metric learning (CMML)	
	Klare *et al.* [15] (2013)	MLBP, SIFT + Heterogenous Prototype	Cosine similarity
	Cai *et al.* [47] (2013)	Coupled least squares regression method with a local consistency constraint	
	Tsai *et al.* [42] (2014)	Domain adaptation based proposed DiCA	Subject-specific SVM
	Lin *et al.* [23] (2016)	Affine transformations	CNNs over Mahalanobis and Cosine scores
Composite	Chugh *et al.* [9] (2013)	Histogram of image moments and HoG	Chi-squared distance
	Han *et al.* [11] (2013)	MLBP of ASM features	Similarity on normalized histogram intersection
	Mittal *et al.* [28] (2015)	Deep Boltzmann Machines	Neural Networks
	Mittal *et al.* [27] (2017)	HoG + DAISY	Chi-squared distance + Attribute feedback
Both	Klum *et al.* [18] (2014)	SketchID- automated system based on holistic [15] and component [11] based algorithms	
	Ouyang *et al.*[32] (2016)	Learned a mapping to reverse the forgetting process of the eyewitness	

Table 1: A brief literature review of sketch-to-photo matching problem.

Table 1 summarizes the literature of facial sketch recognition which shows that both handcrafted and learned representation have been explored. Sketch recognition has traditionally been viewed as a domain adaptation task due to the cross-domain data. Such techniques can be applied for viewed sketch recognition, where the variations across different types of images is primarily governed by the changes in the domain. However, in case of forensic sketch matching for face images, there are several factors apart from the difference in domain which make the problem further challenging, such as memory gap [32] and the bias observed due to the eye-witness [28]. In this work, we propose a novel transform learning based formulation, *DeepTransformer*, which learns meaningful coupled representations for sketch and digital images. Further, two important and challenging application scenarios are used for performance evaluation: (i) age separated digital to sketch matching (both composite and hand-drawn) and (ii) sketch to sketch matching. The effectiveness of the proposed formulation is evaluated on hand-drawn and forensic sketch databases, including a novel sketch database. The key contributions are:

- This is the first work incorporating the concept of Deep Learning in Transform Learning framework. Specifically, novel deep coupled transform learning formulations, *Semi-Coupled* and *Symmetrically-Coupled Deep Transform Learning*, have been presented which imbibe qualities of deep learning with domain adaption.
- This is the first work which presents sketch to sketch matching as an important, yet unattended application for law enforcement. As shown in Figure 1, composite and hand-drawn sketches have significant difference in their information content. Such matching can be useful for crime linking, where different methods may have been used to generate the sketches.
- IIIT-D CSA dataset [1] contains age-separated images of an individual against a sketch image, for 150 subjects. The dataset also contains 3529 digital images.

[1]Dataset will be available at www.iab-rubric.org/resources/csa.html

2. Preliminaries

Dictionary Learning has been used in literature to learn filters and feature representations [22, 31]. For a given input $\mathbf{X}$, a dictionary $\mathbf{D}$ is learned along with the coefficients $\mathbf{Z}$:

$$\min_{\mathbf{D},\mathbf{Z}} \|\mathbf{X}-\mathbf{D}\mathbf{Z}\|_F^2, \; such\ that \; \|\mathbf{Z}\|_0 \leq \tau \tag{1}$$

where, the l_o-norm imposes a constraint of sparsity on the learned coefficients. It can be observed that dictionary learning is a synthesis formulation; i.e., the learned coefficients and dictionary are able to $synthesize$ the given input $\mathbf{X}$. Ravishankar and Bresler [36] proposed it's analysis equivalent, termed as transform learning. It analyzes the data by learning a transform or basis to produce coefficients. Mathematically, for input data $\mathbf{X}$, it can be expressed as:

$$\min_{\mathbf{T},\mathbf{Z}} \|\mathbf{T}\mathbf{X}-\mathbf{Z}\|_F^2, \; such\ that \; \|\mathbf{Z}\|_0 \leq \tau \tag{2}$$

where, $\mathbf{T}$ and $\mathbf{Z}$ are the transform and coefficients, respectively. Relating transform learning to the dictionary learning formulation in Equation 1, it can be seen that dictionary learning is an inverse problem while transform learning is a forward problem. In order to avoid the degenerate solutions of Equation 2, the following formulation is proposed [36]:

$$\min_{\mathbf{T},\mathbf{Z}} \|\mathbf{T}\mathbf{X}-\mathbf{Z}\|_F^2 + \lambda\left(\epsilon\|\mathbf{T}\|_F^2 - log\det\mathbf{T}\right) s.t. \; \|\mathbf{Z}\|_0 \leq \tau \tag{3}$$

The factor '$log\det\mathbf{T}$' refers to the log-determinant regularizer [20], which imposes a full rank on the learned transform to prevent degenerate solutions. The additional penalty term $\|\mathbf{T}\|_F^2$ is to balance scale. In literature, an alternating minimization approach has been presented [37, 38] to solve the above transform learning problem, i.e.:

$$\mathbf{Z} \leftarrow \min_{\mathbf{Z}} \|\mathbf{T}\mathbf{X}-\mathbf{Z}\|_F^2, \; such\ that \; \|\mathbf{Z}\|_0 \leq \tau \tag{4}$$

$$\mathbf{T} \leftarrow \min_{\mathbf{T}} \|\mathbf{T}\mathbf{X}-\mathbf{Z}\|_F^2 + \lambda\left(\epsilon\|\mathbf{T}\|_F^2 - log\det\mathbf{T}\right) \tag{5}$$

The coefficients in Equation 4 are updated using Orthogonal Matching Pursuit (OMP) [34], and transform matrix $\mathbf{T}$ is updated using a closed form solution presented in [40]. The

proof for convergence of the update algorithm can be found in [38]. There is a computational advantage of transform learning over dictionary learning. The latter is a synthesis formulation, and during the test stage, for a given x_{test} it needs to solve a problem of the form:

$$\min_{z_{test}} \|x_{test} - \mathbf{D}z_{test}\|_F^2, \; such\ that \; \|z_{test}\|_0 \leq \tau \quad (6)$$

This is an iterative optimization problem, and thus time consuming, whereas, transform learning is an analysis framework, and at testing time, only the given equation is solved:

$$\min_{z_{test}} \|\mathbf{T}x_{test} - z_{test}\|_F^2, \; such\ that \; \|z_{test}\|_0 \leq \tau \quad (7)$$

This can be solved using one step of hard thresholding [6], making test feature generation very fast and real time.

3. DeepTransformer: Proposed Coupled Deep Transform Learning

Transform Learning has been used for several applications such as blind compressive sensing, online learning, along with image and video de-noising [35, 39, 40]. This research addresses the challenging task of sketch recognition by proposing two novel formulations: semi-coupled, and symmetrically-coupled transform learning. This is the first work which incorporates a mapping function in the transform learning framework in order to reduce between-domain variations. Further, both the models have been extended to propose Semi-Coupled DeepTransformer and Symmetrically-Coupled DeepTransformer.

3.1. Semi-Coupled Deep Transform Learning

As a result of varying information content of images belonging to different domains, there is a need to reduce the domain gap while performing recognition. This is often achieved by mapping the information content of one domain's data onto the other. In real world scenarios of photo to sketch matching, generally a probe sketch image is matched with a gallery of mugshot digital images. This presents the requirement of transforming data from one domain (sketch) onto the other (digital image). For such instances, where the data from only one domain is required to be mapped to the other, Semi-Coupled Transform Learning is proposed. Let $\mathbf{X_1}$ be the data of first domain and $\mathbf{X_2}$ be the data of second domain. The proposed model learns two transform matrices, $\mathbf{T_1}$ and $\mathbf{T_2}$ (one for each domain) and their corresponding features $\mathbf{Z_1}$ and $\mathbf{Z_2}$, such that the features from the first domain can be linearly mapped ($\mathbf{M}$) onto the other. Mathematically this is expressed as:

$$\begin{aligned}\min_{\mathbf{T_1},\mathbf{T_2},\mathbf{Z_1},\mathbf{Z_2},\mathbf{M}} & \|\mathbf{T_1X_1} - \mathbf{Z_1}\|_F^2 + \|\mathbf{T_2X_2} - \mathbf{Z_2}\|_F^2 \\ + \lambda\big(\epsilon\|\mathbf{T_1}\|_F^2 + \epsilon\|\mathbf{T_2}\|_F^2 - & \log\det\mathbf{T_1} - \log\det\mathbf{T_2}\big) \\ & + \mu\|\mathbf{Z_2} - \mathbf{MZ_1}\|_F^2\end{aligned} \quad (8)$$

Equation 8 is solved using alternating minimization approach. Specifically, this equation can be decomposed into five sub-problems, one for each variable, and then each is solved individually, as explained below.

Sub-Problem 1:

$$\min_{\mathbf{T_1}} \|\mathbf{T_1X_1} - \mathbf{Z_1}\|_F^2 + \lambda(\epsilon\|\mathbf{T_1}\|_F^2 - \log\det\mathbf{T_1}) \quad (9)$$

Sub-Problem 2:

$$\min_{\mathbf{T_2}} \|\mathbf{T_2X_2} - \mathbf{Z_2}\|_F^2 + \lambda(\epsilon\|\mathbf{T_2}\|_F^2 - \log\det\mathbf{T_2}) \quad (10)$$

The solution for Equations 9, 10 is similar to the one for Equation 5.

Sub-Problem 3:

$$\begin{aligned}&\min_{\mathbf{Z_1}} \|\mathbf{T_1X_1} - \mathbf{Z_1}\|_F^2 + \mu\|\mathbf{Z_2} - \mathbf{MZ_1}\|_F^2 \\ &\equiv \min_{\mathbf{Z_1}} \left\| \begin{pmatrix}\mathbf{T_1X1} \\ \sqrt{\mu}\mathbf{Z_2}\end{pmatrix} - \begin{pmatrix}\mathbf{I} \\ \sqrt{\mu}\mathbf{M}\end{pmatrix}\mathbf{Z_1} \right\|_F^2\end{aligned} \quad (11)$$

Sub-Problem 4:

$$\begin{aligned}&\min_{\mathbf{Z_2}} \|\mathbf{T_2X_2} - \mathbf{Z_2}\|_F^2 + \mu\|\mathbf{Z_2} - \mathbf{MZ_1}\|_F^2 \\ &\equiv \min_{\mathbf{Z_2}} \left\| \begin{pmatrix}\mathbf{T_2X2} \\ \sqrt{\mu}\mathbf{MZ_1}\end{pmatrix} - \begin{pmatrix}\mathbf{I} \\ \sqrt{\mu}\,\mathbf{I}\end{pmatrix}\mathbf{Z_2} \right\|_F^2\end{aligned} \quad (12)$$

The above two equations are least square problems with a closed form solution, and thus can be minimized for feature representations $\mathbf{Z_1}$ and $\mathbf{Z_2}$.

Sub-Problem 5:

$$\min_{\mathbf{M}} \|\mathbf{Z_2} - \mathbf{MZ_1}\|_F^2 \quad (13)$$

Finally, a mapping $\mathbf{M}$ is learned between the representations $\mathbf{Z_1}$ and $\mathbf{Z_2}$ by solving the above least square equation.

Inspired by the success of deep learning [12, 19, 41] to model high level abstractions and learn large variations in data, this research introduces *deep* transform learning. For a k-layered architecture, Semi-Coupled DeepTransformer can be expressed as:

$$\begin{aligned}\min_{\theta} \Big[\sum_{j=1}^{k} \Big(& \|\mathbf{T_1^j I_1^j} - \mathbf{Z_1^j}\|_\mathbf{F}^\mathbf{2} + \|\mathbf{T_2^j I_2^j} - \mathbf{Z_2^j}\|_\mathbf{F}^\mathbf{2} + \\ & + \lambda\big(\epsilon\|\mathbf{T_1^j}\|_F^2 + \epsilon\|\mathbf{T_2^j}\|_F^2 - \log\det\mathbf{T_1^j} - \log\det\mathbf{T_2^j}\big)\Big) + \\ & \|\mathbf{Z_2^k} - \mathbf{MZ_1^k}\|_\mathbf{F}^\mathbf{2}\Big]\end{aligned} \quad (14)$$

where, $\theta = \{\forall_{\mathbf{j=1}}^{\mathbf{k}}(\mathbf{T_1^j}, \mathbf{T_2^j}, \mathbf{Z_1^j}, \mathbf{Z_2^j}), \mathbf{M}\}$. ($\mathbf{T_1^j}$, $\mathbf{I_1^j}$, and $\mathbf{Z_1^j}$) and ($\mathbf{T_2^j}$, $\mathbf{I_2^j}$, and $\mathbf{Z_2^j}$) refer to the transform matrix, input, and learned representations of the j^{th} layer for the two domains respectively. $\mathbf{M}$ refers to the learned linear mapping between the final representations of the k^{th} layer ($\mathbf{Z_1^k}$, $\mathbf{Z_2^k}$). The input to the model, $\mathbf{I_1^1}$ and $\mathbf{I_2^1}$ are $\mathbf{X_1}$ and $\mathbf{X_2}$, i.e. training data of the first and second domains, respectively. For subsequent layers, $\mathbf{I_1^j}$ and $\mathbf{I_2^j}$ correspond to the feature representations learned at the previous layers, i.e. $\mathbf{Z_1^{j-1}}$ and

$\mathbf{Z_2^{j-1}}$ respectively. As we go deeper and increase the value of k, Equation 14 can be solved similar to Equation 8. The problem can be divided into $(4k)$+1 sub-problems via alternating minimization approach: separate sub-problems for solving the transform matrices $(2k)$, and the learned representations $(2k)$, and one for the final mapping $\mathbf{M}$. However, solving $(4k)$+1 sub-problems can be computationally expensive as the number of parameters is large. As a cost effective alternative, the proposed model can be learned with greedy layer-wise optimization. Here we explain the layer-wise optimization for a 2-layered semi-coupled deep transform learning model (similar greedy layer-wise optimization can be followed for $k > 2$).

Layer One: Learn the first layer transform matrices $(\mathbf{T_1^1}, \mathbf{T_2^1})$ for both domains, along with the representations of the input data $(\mathbf{Z_1^1}, \mathbf{Z_2^1})$:

$$\min_{\mathbf{T_1^1}, \mathbf{Z_1^1}} \left\|\mathbf{T_1^1 X_1} - \mathbf{Z_1^1}\right\|_F^2 + \lambda\left(\epsilon \left\|\mathbf{T_1^1}\right\|_F^2 - log\det \mathbf{T_1^1}\right) \tag{15a}$$

$$\min_{\mathbf{T_2^1}, \mathbf{Z_2^1}} \left\|\mathbf{T_2^1 X_2} - \mathbf{Z_2^1}\right\|_F^2 + \lambda\left(\epsilon \left\|\mathbf{T_2^1}\right\|_F^2 - log\det \mathbf{T_2^1}\right) \tag{15b}$$

Layer Two: Using the representations learned in the first layer as input, semi-coupled transform learning is applied at the second layer to obtain the transform matrices for the second layer, for both domains $(\mathbf{T_1^2}, \mathbf{T_2^2})$:

$$\begin{aligned}\min_{\mathbf{T_1^2}, \mathbf{T_2^2}, \mathbf{Z_1^2}, \mathbf{Z_2^2}, \mathbf{M}} &\left\|\mathbf{T_1^2 Z_1^1} - \mathbf{Z_1^2}\right\|_F^2 + \left\|\mathbf{T_2^2 Z_2^1} - \mathbf{Z_2^2}\right\|_F^2 \\ &+ \lambda\left(\epsilon \left\|\mathbf{T_1^2}\right\|_F^2 + \epsilon \left\|\mathbf{T_2^2}\right\|_F^2 - log\det \mathbf{T_1^2} - log\det \mathbf{T_2^2}\right) \\ &+ \mu \left\|\mathbf{Z_2^2} - \mathbf{M Z_1^2}\right\|_F^2\end{aligned} \tag{16}$$

3.2. Symmetrically-Coupled Deep Transform Learning

In real world scenarios, a given sketch image may be matched with a dataset of different type of sketches for crime-linking. In such cases, learning a single mapping function using semi-coupled transform learning may not be useful. For such cases, symmetrically-coupled transform learning is proposed, where two linear maps are learned; one from the first domain to the second one, and the other from the second domain to the first one. This leads to the following formulation:

$$\begin{aligned}\min_{\mathbf{T_1}, \mathbf{T_2}, \mathbf{Z_1}, \mathbf{Z_2}, \mathbf{M_1}, \mathbf{M_2}} &\left\|\mathbf{T_1 X_1} - \mathbf{Z_1}\right\|_F^2 + \left\|\mathbf{T_2 X_2} - \mathbf{Z_2}\right\|_F^2 \\ &+ \lambda\left(\epsilon \left\|\mathbf{T_1}\right\|_F^2 + \epsilon \left\|\mathbf{T_2}\right\|_F^2 - log\det \mathbf{T_1} - log\det \mathbf{T_2}\right) \\ &+ \mu\left(\left\|\mathbf{Z_2} - \mathbf{M_1 Z_1}\right\|_F^2 + \left\|\mathbf{Z_1} - \mathbf{M_2 Z_2}\right\|_F^2\right)\end{aligned} \tag{17}$$

where, $\mathbf{M_2}$ and $\mathbf{M_1}$ correspond to the mapping matrices to transform feature representations of domain two into those of domain one, and vice versa, respectively. As before, with alternating minimization, Equation 17 can be optimized with the help of the following sub-problems:

Sub-Problem 1:

$$\min_{\mathbf{T_1}} \left\|\mathbf{T_1 X_1} - \mathbf{Z_1}\right\|_F^2 + \lambda(\epsilon \left\|\mathbf{T_1}\right\|_F^2 - log\det \mathbf{T_1}) \tag{18}$$

Sub-Problem 2:

$$\min_{\mathbf{T_2}} \left\|\mathbf{T_2 X_2} - \mathbf{Z_2}\right\|_F^2 + \lambda(\epsilon \left\|\mathbf{T_2}\right\|_F^2 - log\det \mathbf{T_2}) \tag{19}$$

Updates for the transform matrices $(\mathbf{T_1}, \mathbf{T_2})$ remain the same as shown in Equations 9 and 10.

Sub-Problem 3:

$$\begin{aligned}\min_{\mathbf{Z_1}} \left\|\mathbf{T_1 X_1} - \mathbf{Z_1}\right\|_F^2 + \mu\big(&\left\|\mathbf{Z_2} - \mathbf{M_1 Z_1}\right\|_F^2 \\ &+ \left\|\mathbf{Z_1} - \mathbf{M_2 Z_2}\right\|_F^2\big)\end{aligned} \tag{20}$$

Sub-Problem 4:

$$\begin{aligned}\min_{\mathbf{Z_2}} \left\|\mathbf{T_2 X_2} - \mathbf{Z_2}\right\|_F^2 + \mu\big(&\left\|\mathbf{Z_2} - \mathbf{M_1 Z_1}\right\|_F^2 \\ &+ \left\|\mathbf{Z_1} - \mathbf{M_2 Z_2}\right\|_F^2\big)\end{aligned} \tag{21}$$

The above two equations for learning the representations $(\mathbf{Z_1}, \mathbf{Z_2})$ of the two domains are least square minimizations, and thus have closed form solutions.

Sub-Problem 5:

$$\min_{\mathbf{M_1}} \left\|\mathbf{Z_2} - \mathbf{M_1 Z_1}\right\|_F^2 \tag{22}$$

Sub-Problem 6:

$$\min_{\mathbf{M_2}} \left\|\mathbf{Z_1} - \mathbf{M_2 Z_2}\right\|_F^2 \tag{23}$$

Similar to Equation 13, mappings $(\mathbf{M_1}, \mathbf{M_2})$ can be learned by solving the above using least square minimization. As discussed, the DeepTransfomer is solved using alternating minimization approach. Each of the subproblems of Equations 8 and 17 are solved with guaranteed convergence [36]. Specifically, learning Z has analytical solution and transform updates are done by conjugate gradients which can only decrease. Overall, the model has monotonically decreasing cost function and therefore, will converge.

We extend Equation 17 and propose symmetrically-coupled *deep* transform learning where, $\theta = \{\forall_{\mathbf{j=1}}^{\mathbf{k}}(\mathbf{T_1^j}, \mathbf{T_2^j}, \mathbf{Z_1^j}, \mathbf{Z_2^j}), \mathbf{M_1}, \mathbf{M_2}\}$, and $\mathbf{M_2}$ and $\mathbf{M_1}$ correspond to the mapping matrices to transform feature representations of domain two into those of domain one, and vice versa, respectively. It is mathematically expressed as:

$$\begin{aligned}\min_{\theta} \Big[\sum_{j=1}^{k} \Big(&\|\mathbf{T_1^j I_1^j} - \mathbf{Z_1^j}\|_\mathbf{F}^2 + \|\mathbf{T_2^j I_2^j} - \mathbf{Z_2^j}\|_\mathbf{F}^2 + \\ &+ \lambda\big(\epsilon\|\mathbf{T_1^j}\|_F^2 + \epsilon\|\mathbf{T_2^j}\|_F^2 - log\det \mathbf{T_1^j} - log\det \mathbf{T_2^j}\big)\Big) + \\ &\|\mathbf{Z_2^k} - \mathbf{M_1 Z_1^k}\|_\mathbf{F}^2 + \|\mathbf{Z_1^k} - \mathbf{M_2 Z_2^k}\|_\mathbf{F}^2\Big]\end{aligned} \tag{24}$$

This formulation can be solved using alternating minimization approach with $(4k$+2) sub-problems where the last two sub-problems are related to learning mappings $\mathbf{M_1}$ and

$\mathbf{M_2}$. However, like Semi-Coupled DeepTransformer, we optimize symmetrically coupled deep transform algorithm in a greedy layer wise manner. The optimization for a 2-layer Symmetrically-Coupled DeepTransformer is as follows:

Layer One:

$$\min_{\mathbf{T_1^1}, \mathbf{Z_1^1}} \left\|\mathbf{T_1^1 X_1} - \mathbf{Z_1^1}\right\|_F^2 + \lambda\left(\epsilon \left\|\mathbf{T_1^1}\right\|_F^2 - log\det \mathbf{T_1^1}\right) \tag{25a}$$

$$\min_{\mathbf{T_2^1}, \mathbf{Z_2^1}} \left\|\mathbf{T_2^1 X_2} - \mathbf{Z_2^1}\right\|_F^2 + \lambda\left(\epsilon \left\|\mathbf{T_2^1}\right\|_F^2 - log\det \mathbf{T_2^1}\right) \tag{25b}$$

Layer Two:

$$\begin{aligned}\min_{\mathbf{T_1^2}, \mathbf{T_2^2}, \mathbf{Z_1^2}, \mathbf{Z_2^2}, \mathbf{M}} & \left\|\mathbf{T_1^2 Z_1^1} - \mathbf{Z_1^2}\right\|_F^2 + \left\|\mathbf{T_2^2 Z_2^1} - \mathbf{Z_2^2}\right\|_F^2 \\ + \lambda\Big(\epsilon \left\|\mathbf{T_1^2}\right\|_F^2 + \epsilon \left\|\mathbf{T_2^2}\right\|_F^2 - & log\det \mathbf{T_1^2} - log\det \mathbf{T_2^2}\Big) \\ + \mu\big(\left\|\mathbf{Z_2^2} - \mathbf{M_1 Z_1^2}\right\|_F^2 + & \left\|\mathbf{Z_1^2} - \mathbf{M_2 Z_2^2}\right\|_F^2\big)\end{aligned} \tag{26}$$

The first layer learns the low level representation of each domain independently, while the second layer learns the high level representations and mapping between the representations of the two domains/modalities. The proposed model thus encodes domain specific features, followed by features incorporating the between-domain variations.

3.3. DeepTransformer for Sketch Recognition

The proposed two layer DeepTransformer is used for performing face sketch matching. For semi-coupled DeepTransformer, the following steps are performed:

Training: Given a set of sketch and digital training pairs, $\mathbf{X_s}, \mathbf{X_d}$, transform matrices $(\mathbf{T_s^1}, \mathbf{T_d^1}, \mathbf{T_s^2}, \mathbf{T_d^2})$ and coefficient vectors $(\mathbf{Z_s^1}, \mathbf{Z_d^1}, \mathbf{Z_s^2}, \mathbf{Z_d^2})$ are learned using Equation 14, along with a mapping, $\mathbf{M}$, between $\mathbf{Z_s^2}, \mathbf{Z_d^2}$. A two hidden layer neural network classifier is trained to make identification decisions.

Testing: For a given probe sketch image, x_{sTest}, the first and second layer feature representations are extracted using the learned transform spaces:

$$z_{sTest}^1 = \mathbf{T_s^1} x_{sTest};\ z_{sTest}^2 = \mathbf{T_s^2} z_{sTest}^1 \tag{27}$$

The mapping $\mathbf{M}$ is used to transform the feature vector onto the digital image space, i.e, $z_{dTest}^2 = \mathbf{M_1} z_{sTest}^2$. The feature representation of the sketch (probe) in the digital image feature space, z_{dTest} is now used for performing recognition using the trained neural network. For sketch to sketch matching (i.e. cases where mappings to and from different modalities are required), similar steps can be followed for utilizing symmetrically-coupled deep transform learning.

4. Databases and Experimental Protocol

Face sketch databases [18, 45] generally comprise of viewed sketches, either hand-drawn or composite. Viewed sketches are created by looking at a digital image and sketching it simultaneously. This fails to capture the uncertainty in the recall process that humans encounter or the variations in characteristics, like a hairstyle modification, that are generally present between a sketch and digital image acquired at different times. In this research, we utilize a novel IIIT- D *Composite Sketch with Age variations (CSA)* dataset [10], which is the **first** publicly available dataset containing multiple age-separated digital images for a given sketch image. Inspired by Bhatt *et al.* [5], the human forgetting process is incorporated by creating semi-forensic composite sketches. The user is shown the digital image of a subject for a few minutes, and is asked to create the composite sketch after a period of 30 minutes based on his/her memory. The database consists of 3529 sketches and digital face images pertaining to 150 individuals. Out of the 150 subjects, 52 are selected from the FG-NET Aging Database [21], 82 are selected from IIIT-D Aging Database [48], and the remaining subjects are collected from the Internet. The composite sketch images are created using FACES [2], a popular software to generate photo-like composite sketches.

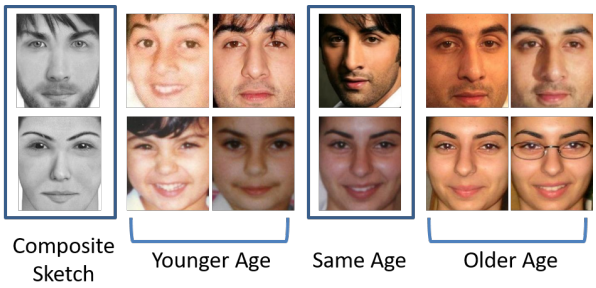


Figure 2: Sample images from the IIIT-D CSA database showing a sketch and age-separated face images of two subjects.

In IIIT-D CSA dataset, the digital images span over an age range of 1 to 65 years. For each subject, an image is chosen from the middle of his/her age range and a corresponding sketch image is generated. Following this, each subject's digital images are divided into three categories:

(i) Younger age group: This category models the scenario when the digital images (gallery) are younger than the probe sketch image. This set contains 1713 digital images.

(ii) Same age group: This category represents the scenario when the age of an individual is similar in both digital image (gallery) and sketch image (probe). A total of 150 digital images exist in this set.

(iii) Older age group: This category imitates the scenario when the digital images (gallery) of the individuals are at an age older than the sketch. It consists of 1516 digital images.

Overall, IIIT-D CSA consists of 150 composite sketch images, one for each subject, and 3379 digital images belonging to different age categories. Figure 2 shows sample images from the dataset.

Protocol	Gallery Type	Probe (Sketch)	Databases for Feature Learning	Test Database	No. of Training Pairs	Size of Gallery	Size of Probe
Matching sketch to age-separated digital images: Semi-Coupled DeepTransformer							
P1	Younger age images	Composite	CUFS, CUFSF, e-PRIP, PRIP-VSGC, IIIT-D Viewed and Semi-viewed	CSA	2129	875	90
P2	Same age images	Composite		CSA	2129	90	90
P3	Older age images	Composite		CSA	2129	1044	90
P4	Large-scale image dataset	Composite		CSA	2129	7165	90
P5		Forensic		IIIT-D Forensic	2129	7265	190
Sketch to sketch matching: Symmetrically-Coupled DeepTransformer							
P6	Composite	Hand-drawn	CUFS, e-PRIP	CUFS, e-PRIP	50	73	73
P7	Hand-drawn	Composite	CUFS, e-PRIP	CUFS, e-PRIP	50	73	73

Table 2: Details of experimental protocols. For P6-P7, unseen training and testing partitions are used from the CUFS and e-PRIP datasets (both contain sketches pertaining to AR dataset).

Apart from IIIT-D CSA, we have also used viewed hand-drawn sketch and digital image pairs from CUHK Face Sketch Dataset (**CUFS**) [45] (311 pairs of students and AR dataset [25]), CUHK Face Sketch FERET Dataset (**CUFSF**) [49] (1194 pairs), and **IIIT-D Sketch dataset** [5]. IIIT-D dataset contains viewed (238 pairs), semi-viewed (140 pairs), and forensic sketches (190 pairs). Composite sketches from PRIP Viewed Software-Generated Composite database (**PRIP-VSGC**) [18] and extended-PRIP Database (**e-PRIP**) [29] (Indian user set) are also used.

Experimental Protocol: To evaluate the efficacy of the proposed formulations two challenging problems are considered: sketch matching against age-separated digital images (semi-coupled DeepTransformer) and sketch to sketch matching (symmetrically coupled DeepTransformer). Since this is the first research that focuses on sketch to sketch matching, as well as sketch to age-separated digital matching, we have created seven different experimental protocols to understand the performance with individual cases. These protocols are classified according to the two case studies and the details are summarized in Table 2.

1. Matching Sketch to Age-Separated Digital Images: CSA test set and IIITD Forensic hand-drawn database have been used to evaluate the performance of the proposed model. Inspired from real life scenarios, the test set is divided into a gallery and probe set. The gallery contains the digital images while the probe contains the sketch image. The first three protocols evaluate the effect of age difference on the recognition performance, and the next two protocols (P4 and P5) analyze the difference in performance on matching forensic and composite sketches with large scale digital image gallery. Since sketch to digital image matching experiment involves one way mapping, the results are demonstrated with Semi-Coupled DeepTransformer.

2. Sketch to Sketch Matching: In real world crime scene linking application, one might want to match a hand-drawn sketch against a database of composite sketches, or the other way around. Therefore, for this experiment, the proposed Symmetrically-Coupled DeepTransformer is used. CUFS dataset contains hand-drawn sketch images for the AR dataset (123 subjects), while e-PRIP contains composite sketches generated by a sketch artist for the same. The following two experiments with protocols P6 and P7 are performed: **(i)** composite to hand-drawn sketch, and **(ii)** hand-drawn to composite sketch matching.

5. Results and Observations

Effectiveness of DeepTransformer is evaluated with multiple input features, namely Dense Scale Invariant Feature Transform (DSIFT) [7], Dictionary Learning (DL) [22], Class Sparsity based Supervised Encoder (L-CSSE) [24], Light CNN [46], and VGG-Face [33]. To analyze the effect of depth in this formulation, the results are computed with single layer (low level features) and with two layers (high level features) of DeepTransformer. Two kinds of comparative experiments are performed. The first one compares the performance of one layer and two layers deep transform learning algorithms with two classifiers, i.e., Euclidean distance and neural network. The second comparison is performed with existing algorithms like Semi-Coupled Dictionary Learning algorithm (SCDL) [44] and Multi-Modal Sharable and Specific feature learning algorithm (MMSS) [43]. Both the techniques have been used in literature for performing cross-domain recognition, wherein the former is a coupled dictionary learning based approach (synthesis technique), and the latter incorporated transform learning with convolutional neural networks for addressing cross-domain recognition. Comparison has also been drawn with state-of-the-art sketch recognition algorithms, namely MCWLD [5] and GSMFL [23], and a commercial-off-the-shelf system (COTS), Verilook [4]. In all the experiments, for training the networks, data augmentation is performed on the gallery images to increase per-class samples by varying the illumination and flipping the images along the y-axis. The key observations from experimental results are:

Performance with Different Features: Tables 4 and 3 present the rank-10 identification accuracies of DeepTrans-

former with different features, for both applications of sketch matching: sketch to sketch matching and sketch to photo matching. Table 3 presents the accuracies for sketch to digital image matching, where the proposed Semi-Coupled DeepTransformer has been used. The results show that DeepTransformer enhances the performance of existing feature extraction techniques by at least 10% as compared to Euclidean distance matching, and at most 22% when neural network (NNET) is used for classification. Upon comparing accuracies across features, it is observed that DeepTransformer achieves the best results with L-CSSE features for all protocols. Similar results can be seen from Table 4 where Symmetrically-Coupled DeepTransformer has been used for sketch to sketch matching. Experimentally, it can be observed that providing class-specific features to DeepTransformer results in greater improvement. L-CSSE is a supervised deep learning model built over an autoencoder. The model incorporates supervision by adding a $l_{2,1}$ norm regularizer during the feature learning to facilitate class-specific feature learning. The model utilizes both global and local facial regions to compute feature vector and has been shown to achieve improved results for existing face recognition problems. Further, we also observe that L-CSSE encodes the high frequency features in both local and global regions which are pertinent to digital face to sketch matching. Moreover, improved performance is observed for hand-crafted, as well as representation learning based features, thus promoting the use of DeepTransformer for different types of feature extraction techniques and input data.

Comparison with Existing Approaches: Table 5 shows that the proposed, DeepTransformer with L-CSSE features outperforms existing algorithms for both the applications of sketch recognition. In case of sketch to digital image matching, with younger age protocol (P1), Semi-Coupled DeepTransformer attains a rank-10 accuracy of **42.6%**, which is at least 15% better than existing algorithms, and around 24% better than COTS. Similar trends are observed for P2 and P3 protocols, where the proposed DeepTransformer outperforms existing techniques and the commercial-off-the-shelf system by a margin of at least 13% and 11% respectively. Additionally, the matching accuracy achieved by the proposed Symmetrically-Coupled DeepTransformer exceeds existing techniques for the task of sketch to sketch matching as well (P6, P7). An improvement of at least 14% and at most 20% is seen with the proposed DeepTransformer (L-CSSE as feature) for the given protocols. This accentuates the use of DeepTransformer for addressing the problem of real world sketch matching.

Effect of Layer-by-Layer Training and Number of Layers: We compare the performance of the proposed DeepTransformer with and without layer-by-layer training (i.e. Equations 14 and 24 for direct solving for $k = 2$ and layer-by-layer training as per Equations 15-16 and 25-26). On a 108-core server with 256GB RAM, for protocols P1 to P3, training Semi-Coupled DeepTransformer with layer-by-layer training requires 142 seconds which is 12 seconds faster than without layer-by-layer training. For both the cases, for $k = 2$, the rank-10 accuracies are same which shows that layer-by-layer training is cost effective. We also analyze the effect of number of layers and, as shown in Tables 4 and 3, 1-8% improvement in rank-10 accuracy is observed for different protocols upon going deeper.

Features	**Euclidean Distance**	**NNET**	**DeepTransformer**	
			1-Layer	**2-Layer**
Gallery with Younger Age Digital Images (P1)				
DSIFT	8.9	15.6	26.7	27.8
DL	2.2	14.4	17.8	17.8
VGG	1.1	11.1	12.2	12.2
Light CNN	8.9	12.2	30.0	27.8
L-CSSE	14.4	19.7	34.1	**42.6**
Gallery with Same Age Digital Images (P2)				
DSIFT	7.8	26.7	25.6	27.8
DL	1.1	13.3	15.6	17.8
VGG	2.2	12.2	14.4	14.4
Light CNN	11.1	25.6	32.2	34.4
L-CSSE	16.3	30.2	37.7	**44.2**
Gallery with Older Age Digital Images (P3)				
DSIFT	5.6	21.1	23.3	24.4
DL	2.2	13.3	17.8	18.9
VGG	2.2	11.1	12.2	12.2
Light CNN	7.8	20.0	24.4	28.9
L-CSSE	9.9	20.0	28.9	**36.0**

Table 3: Rank-10 accuracies (%) for protocols P1 to P3 using proposed Semi-Coupled DeepTransformer.

Performance on Large-Scale Dataset: The performance of the proposed DeepTransformer has also been evaluated for a large-scale real world dataset using protocols P4 and P5. Figure 3 presents the Cumulative Match Characteristic curves (CMCs) for IIIT-D CSA composite and IIIT-D Forensic hand-drawn sketch database respectively. The proposed Semi-Coupled DeepTransformer achieves a rank-50 accuracy of 33.7%, which is an improvement of at least 5% from other algorithms on IIIT-D CSA dataset. Similar results can be observed on the forensic sketches as well.

The experimental results showcase the efficacy of the proposed DeepTransformer, in terms of the improvement in identification accuracies with different features, and in comparison with other existing models. The results suggest that the DeepTransformer is robust to the type of feature and has the ability to learn over varying input spaces. Moreover, efficient training of the symmetrically coupled DeepTransformer with as few as 50 digital-sketch pairs (P6 and P7) motivate the use of the proposed architecture for small sample size problems as well. The evaluation on different real world protocols further strengthens the usage of the pro-

	Gallery: Composite, Probe: Hand-drawn (P6)				Gallery: Hand-drawn, Probe: Composite (P7)			
Features	**Euclidean Distance**	**NNET**	**DeepTransformer**		**Euclidean Distance**	**NNET**	**DeepTransformer**	
			1-Layer	**2-Layer**			**1-Layer**	**2-Layer**
DSIFT	4.1	16.4	24.7	28.8	2.7	13.7	23.3	30.1
DL	4.1	15.1	17.8	19.2	6.9	16.4	19.2	20.6
VGG	6.9	12.3	24.7	27.4	6.9	15.1	19.2	20.6
Light CNN	8.2	15.1	26.0	30.1	8.2	16.4	20.6	28.8
L-CSSE	10.9	17.8	28.4	**31.5**	10.9	20.9	31.5	**33.6**

Table 4: Rank-10 accuracies (%) for sketch to sketch matching (P6, P7) using Symmetrically-Coupled DeepTransformer.

Algorithm	**P1**	**P2**	**P3**	**P6**	**P7**
MCWLD [5]	26.8	30.7	24.4	16.5	19.2
GSMFL [23]	25.2	29.3	23.3	16.5	19.2
SCDL [44]	23.3	25.6	18.9	15.1	13.7
MMSS [43]	22.2	27.8	21.1	13.3	15.1
Verilook (COTS) [4]	17.8	16.6	12.2	10.9	13.7
DeepTransformer (with L-CSSE)	**42.6**	**44.2**	**36.0**	**31.5**	**33.6**

Table 5: Rank-10 accuracies (%) comparing proposed DeepTransformer with existing algorithms and COTS.

posed model for addressing cross domain matching tasks.

6. Conclusion

This research focuses on the challenging problem of face sketch recognition and proposes a novel transform learning based formulation, called as *DeepTransformer*. Two models: Semi-Coupled and Symmetrically-Coupled DeepTransformer have been presented, both of which aim to reduce the variations between two domains. The highlight of the proposed formulation is that it provides the flexibility of using an existing feature extractor and classifier in the framework. The proposed DeepTransfomer is evaluated with real world scenarios of age-separated digital image to sketch matching and sketch to sketch matching. Results are also shown on the IIIT-D Composite Sketch with Age variations database of 150 subjects. Comparison with existing state-of-the-art algorithms and commercial-off-the-shelf system further instantiates the efficacy of both the semi-coupled and symmetrically coupled variants of the purposed DeepTransformer.

7. Acknowledgment

This research is partially supported by MEITY (Government of India), India. M. Vatsa, R. Singh, and A. Majumdar are partially supported through Infosys Center for Artificial Intelligence. S. Nagpal is partially supported through TCS PhD Fellowship. The authors acknowledge T. Chugh for his help in database creation. R. Singh also thank NVIDIA Corp. for Tesla K40 GPU for research.

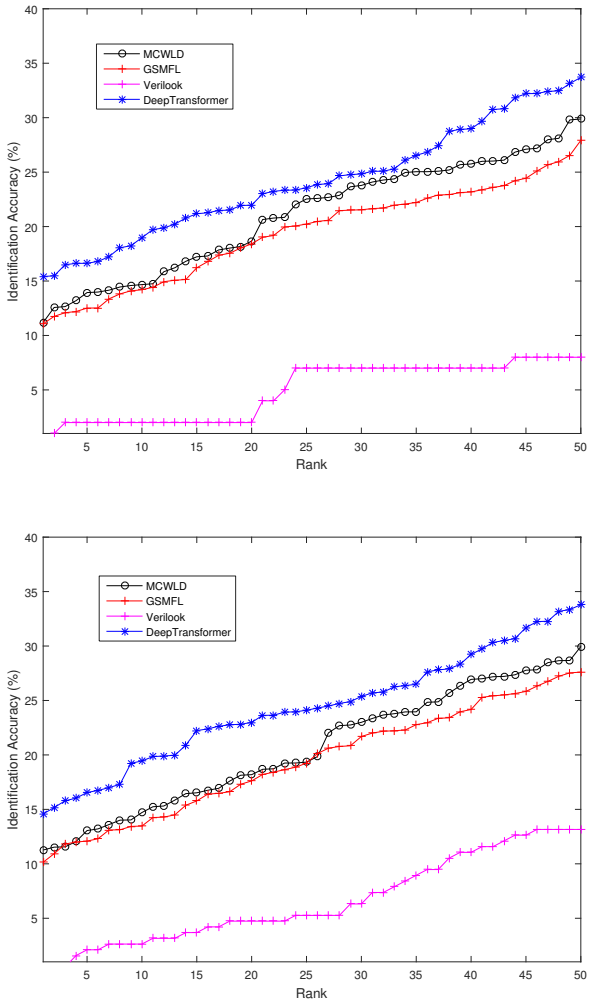


Figure 3: CMC curves for P4 and P5 experiments: (a) CSA, and (b) IIIT-D Forensic datasets.

References

[1] evofit. http://www.evofit.co.uk/. 1

[2] Faces. http://www.facesid.com/products.html. 1, 5

[3] Identi-kit. http://identikit.net/. 1

[4] Verilook. http://www.neurotechnology.com/verilook.html. 6, 8

[5] H. S. Bhatt, S. Bharadwaj, R. Singh, and M. Vatsa. Memetically optimized MCWLD for matching sketches with digital face images. *IEEE Transactions on Information Forensics and Security*, 7(5):1522–1535, 2012. 2, 5, 6, 8

[6] T. Blumensath and M. E. Davies. Iterative thresholding for sparse approximations. *Journal of Fourier Analysis and Applications*, 14(5-6):629–654, 2008. 3

[7] A. Bosch, A. Zisserman, and X. Muñoz. Image classification using random forests and ferns. In *IEEE International Conference on Computer Vision*, pages 1–8, 2007. 6

[8] J. Choi, A. Sharma, D. W. Jacobs, and L. S. Davis. Data insufficiency in sketch versus photo face recognition. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, pages 1–8, 2012. 1

[9] T. Chugh, H. S. Bhatt, R. Singh, and M. Vatsa. Matching age separated composite sketches and digital face images. In *IEEE International Conference on Biometrics: Theory, Applications and Systems*, pages 1–6, 2013. 2

[10] T. Chugh, M. Singh, S. Nagpal, R. Singh, and M. Vatsa. Transfer learning based evolutionary algorithm for composite face sketch recognition. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2017. 5

[11] H. Han, B. F. Klare, K. Bonnen, and A. K. Jain. Matching composite sketches to face photos: A Component-Based Approach. *IEEE Transactions on Information Forensics and Security*, 8(1):191–204, 2013. 2

[12] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 770–778, 2016. 3

[13] I. Kemelmacher-Shlizerman, S. M. Seitz, D. Miller, and E. Brossard. The MegaFace benchmark: 1 million faces for recognition at scale. *CoRR*, abs/1512.00596, 2015. 1

[14] Z. Khan, Y. Hu, and A. Mian. Facial self similarity for sketch to photo matching. In *International Conference on Digital Image Computing Techniques and Applications*, pages 1–7, 2012. 2

[15] B. Klare and A. K. Jain. Heterogeneous face recognition using kernel prototype similarities. *IEEE Transactions on Pattern Analysis Machine Intelligence*, 35(6):1410–1422, 2013. 2

[16] B. F. Klare, B. Klein, E. Taborsky, A. Blanton, J. Cheney, K. Allen, P. Grother, A. Mah, M. J. Burge, and A. K. Jain. Pushing the frontiers of unconstrained face detection and recognition: IARPA Janus Benchmark A. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1931–1939, 2015. 1

[17] B. F. Klare, Z. Li, and A. Jain. Matching forensic sketches to mug shot photos. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(3):639–646, 2011. 1

[18] S. J. Klum, H. Han, B. F. Klare, and A. K. Jain. The facesketchid system: Matching facial composites to mugshots. *IEEE Transactions on Information Forensics and Security*, 9(12):2248–2263, 2014. 1, 2, 5, 6

[19] A. Krizhevsky, I. Sutskever, and G. E. Hinton. ImageNet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*, pages 1097–1105. 2012. 3

[20] B. Kulis, M. Sustik, and I. Dhillon. Learning low-rank kernel matrices. In *International Conference on Machine Learning*, pages 505–512, 2006. 2

[21] A. Lanitis. Comparative evaluation of automatic age-progression methodologies. *EURASIP Journal on Advances in Signal Processing*, pages 101:1–101:10, 2008. 5

[22] D. D. Lee and H. S. Seung. Learning the parts of objects by nonnegative matrix factorization. *Nature*, 401:788–791, 1999. 2, 6

[23] L. Lin, G. Wang, W. Zuo, X. Feng, and L. Zhang. Cross-domain visual matching via generalized similarity measure and feature learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(6):1089–1102, 2017. 1, 2, 6, 8

[24] A. Majumdar, R. Singh, and M. Vatsa. Face verification via class sparsity based supervised encoding. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(6):1273–1280, 2017. 6

[25] A. Martínez and R. Benavente. The AR face database. Technical Report 24, Computer Vision Center, 1998. 6

[26] A. Mignon and F. Jurie. CMML: A New Metric Learning Approach for Cross Modal Matching. In *Asian Conference on Computer Vision*, page 14, 2012. 2

[27] P. Mittal, A. Jain, G. Goswami, M. Vatsa, and R. Singh. Composite sketch recognition using saliency and attribute feedback. *Information Fusion*, 33:86–99, 2017. 2

[28] P. Mittal, M. Vatsa, and R. Singh. Composite sketch recognition via deep network - a transfer learning approach. In *International Conference on Biometrics*, pages 251–256, 2015. 2

[29] P. Mittal, M. Vatsa, and R. Singh. Composite sketch recognition via deep network - a transfer learning approach. In *International Conference on Biometrics*, pages 251–256, 2015. 6

[30] S. Nagpal, M. Vatsa, and R. Singh. Sketch recognition: What lies ahead? *Image and Vision Computing*, 55, Part 1:9 – 13, 2016. 1

[31] B. A. Olshausen and D. J. Field. Sparse coding with an overcomplete basis set: A strategy employed by v1? *Vision Research*, 37(23):3311 – 3325, 1997. 2

[32] S. Ouyang, T. M. Hospedales, Y. Z. Song, and X. Li. Forgetmenot: Memory-aware forensic facial sketch matching. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 5571–5579, 2016. 2

[33] O. M. Parkhi, A. Vedaldi, and A. Zisserman. Deep face recognition. In *British Machine Vision Conference*, 2015. 6

[34] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad. Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition. In *Conference on Signals, Systems and Computers*, pages 1–3, 1993. 2

[35] L. Pfister and Y. Bresler. Automatic parameter tuning for image denoising with learned sparsifying transforms. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 6040–6044, 2017. 3

[36] S. Ravishankar and Y. Bresler. Learning sparsifying transforms. *IEEE Transactions on Signal Processing*, 61(5):1072–1086, 2013. 2, 4

[37] S. Ravishankar and Y. Bresler. Efficient blind compressed sensing using sparsifying transforms with convergence guarantees and application to magnetic resonance imaging. *SIAM Journal on Imaging Sciences*, 8(4):2519–2557, 2015. 2

[38] S. Ravishankar and Y. Bresler. Online sparsifying transform learning 2014; Part II: Convergence Analysis. *IEEE Journal of Selected Topics in Signal Processing*, 9(4):637–646, 2015. 2, 3

[39] S. Ravishankar and Y. Bresler. Data-driven learning of a union of sparsifying transforms model for blind compressed sensing. *IEEE Transactions on Computational Imaging*, 2(3):294–309, 2016. 3

[40] S. Ravishankar, B. Wen, and Y. Bresler. Online sparsifying transform learning 2014; part I: Algorithms. *IEEE Journal of Selected Topics in Signal Processing*, 9(4):625–636, 2015. 2, 3

[41] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *CoRR*, abs/1409.1556, 2014. 3

[42] Y. H. Tsai, H. M. Hsu, C. A. Hou, and Y. C. F. Wang. Person-specific domain adaptation with applications to heterogeneous face recognition. In *IEEE International Conference on Image Processing*, pages 338–342, 2014. 2

[43] A. Wang, J. Cai, J. Lu, and T.-J. Cham. MMSS: Multi-modal sharable and specific feature learning for rgb-d object recognition. In *The IEEE International Conference on Computer Vision*, 2015. 6, 8

[44] S. Wang, L. Zhang, Y. Liang, and Q. Pan. Semi-coupled dictionary learning with applications to image super-resolution and photo-sketch synthesis. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 2216–2223, 2012. 6, 8

[45] X. Wang and X. Tang. Face photo-sketch synthesis and recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(11):1955–1967, 2009. 1, 5, 6

[46] X. Wu, R. He, and Z. Sun. A lightened CNN for deep face representation. *CoRR*, abs/1511.02683, 2015. 6

[47] C. Xinyuan, W. Chunheng, X. Baihua, C. Xue, L. Zhijian, and S. Yanqin. Coupled latent least squares regression for heterogeneous face recognition. In *IEEE International Conference on Image Processing*, pages 2772–2776, 2013. 2

[48] D. Yadav, R. Singh, M. Vatsa, and A. Noore. Recognizing age-separated face images: Humans and machines. *PLoS ONE*, 9, 2014. 5

[49] W. Zhang, X. Wang, and X. Tang. Coupled information-theoretic encoding for face photo-sketch recognition. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 513–520, 2011. 1, 6