Face Recognition using Scattering Wavelet under Illicit Drug Abuse Variations

Prateekshit Pandey, Richa Singh, Mayank Vatsa
IIIT-Delhi India
{prateekshit12078, rsingh, mayank}@iiitd.ac.in

Abstract

Prolonged usage of illicit drugs alter texture and geometric variations of a face and hence, affect the performance of face recognition algorithms. This research proposes a two fold contribution for advancing the state-of-art in recognizing face images with variations caused due to substance abuse: firstly, scattering transform (ScatNet) based face recognition algorithm is proposed. The algorithm yields good results however, it is very expensive in terms of the computational time and space. Therefore, as the next contribution, an autoencoder-style mapping function (AutoScat) is proposed that learns to encode the ScatNet representation of a face image to reduce the computation time. The results are evaluated on the publicly available Illicit Drug Abuse Face database. The results show that ScatNet based face recognition algorithm outperforms two commercial matchers. Further, compared with ScatNet, AutoScat is able to achieve lower rank-1 accuracy but requires 10^{-3} times lesser computational requirements and around 400 times smaller feature space.

1. Introduction

With increasing usage of face recognition systems for critical applications such as law enforcement, it is imperative to make the systems robust to variations including pose, expression, illumination, and aging [5]. In recent past, researchers have also shown that covariates such as plastic surgery, makeup and disguise variations [3], [4], [7] have the potential of confounding the face recognition systems, posing immense threat to the security. Most recently, another such variation is identified which can affect face recognition algorithms: illicit drug abuse. Yadav et al. [9] have shown that the variations caused to the facial appearance because of prolonged substance abuse can have startling influence on automated face recognition algorithms. As shown in Figure 1, prolonged substance abuse can be detrimental to health, which evidently gets reflected in the appearance and variations in facial features of the individual. The variations ensued by substance abuse could very well be encoded by the feature extraction algorithms and subsequently perturb match score distributions.

Figure 1: Sample before-after face images with drug abuse variations (images are obtained from Internet).

As mentioned previously, Yadav et al. [9] have presented the results of different face matchers and showcased reduced performance on the Illicit Drug Abuse Face (IDAF) database. They have also presented a novel algorithm for detecting if a given image pair is affected by the drug abuse or not. However, they have not developed a face recognition algorithm which can address the variations caused due to illicit drug abuse. In this research, we have approached the problem in a two-fold approach (and thus two major contributions): firstly, we propose a Scattering Wavelet Network (ScatNet) [6] based feature extraction for recognizing faces with such variations. Scattering wavelet transform learns a deep representation of Morlet features at different scales and orientations, and has shown significant results on texture classification. However, the dimensionality of features is significantly high and it takes a lot of time to compute every node of the scattering wavelet network; therefore, it is computationally intensive. As a second contribution, we propose an auto-encoder like learning paradigm, termed as AutoScat, to encode scattering wavelet transform features into a representative lower dimensional space, which can be

computed by applying a simple linear transformation over the raw image. AutoScat aims to solve the two major drawbacks of ScatNet, namely (1) high space complexity (ScatNet works at 400 times the dimensionality of AutoScat) and (2) high computation time (ScatNet takes around 10^3 times as much computation time as AutoScat). Face identification experiments on the illicit drug abuse face database show that ScatNet outperforms two commercial matchers as well as AutoScat. On the other hand, computational and spatial advantages of AutoScat provide a balanced trade-off. The remainder of this paper is organized as follows: Section 2 covers the explanation of the Scattering wavelet transform along with the fast scattering transform and the proposed learning-based AutoScat architecture. Section 3 delineates the experimental protocol and results on the Illicit Drug Abuse Face (IDAF) database.

2. Scattering Wavelet Transformation Based Face Recognition

In this research, we propose to utilize scattering wavelet transform[1] for facial feature representation. Scattering wavelet transform has shown excellent results on generic texture-images [6] by providing feature representations robust to affine deformations. ScatNet features obviate the need for deep training networks [2], as it uses pre-specified filter parameters (Morlet in this case) instead of learning them, which is exhaustive in covering the complete frequency domain of the image. However, high time complexity and high dimensionality of feature vector limits the use of such representations in real-time applications such as face recognition. Therefore, we propose an autoencoder-style learning approach to *learn* a succinct and computationally inexpensive representation of the scattering wavelet transform. Figure 2 illustrates the steps involved in the proposed face recognition algorithm. Section 2.1 explains the formulation of scattering wavelet transform, Section 2.2 explains the fast scattering computation, and Section 2.3 explains how the autoscat features are matched.

2.1. Scattering Wavelet Transform

ScatNet is a deep wavelet based architecture for texture feature representation that applies Morlet wavelet at multiple scales and orientations. Let $I \in \mathbb{R}^2$ represent a 2-dimensional image and $\phi_J(f) = 2^{-2J}\phi(2^{-J}f)$ be a low pass (averaging) filter in the frequency domain, governed by scaling factor J. A locally affine transformation of an image is achieved by applying convolution of an averaging filter over the image, given as:

$$S_0 I(f) = I \star \phi_J(f) \tag{1}$$

[1]ScatNet is an open source MATLAB toolbox, available at http://www.di.ens.fr/data/software/scatnet/.

This representation is locally translation invariant up to 2^J pixels and rejects most of the high frequency components of the image. In order to capture high frequency components of the image, set of constant high frequency band pass filters is applied to the image. The set $\wedge$ consists of quadrature phased complex Morlet filters, or *wavelets*, $\psi_\lambda(f)$, with varying scales 2^j and rotations θ, given as

$$\psi_\lambda(f) = 2^{2j}\psi\Big(\frac{2^j}{\theta}f\Big), \lambda = j, \theta \tag{2}$$

The constant set of such band pass filters or *wavelet bank* $\wedge_0$, is given as

$$\wedge_0(J,K) = \{\psi_{\lambda_1}(f), \psi_{\lambda_2}(f), \psi_{\lambda_3}(f), ..., \psi_{\lambda_w}(f)\} \tag{3}$$

where, $\lambda_i = \{j, \theta\}$; with $j \in \{0, 1, ...J-1\}$, and $\theta = k\pi/K$, $k \in \{0, 1, ..., K-1\}$ and $w = JK$. The wavelet modulus $W(I)$ of an image I is the collective output of applying averaging filter and high frequency band pass filters (wavelets) over the image, given by:

$$W_1(I, \wedge_0) = (I \star \phi_J(f), |I \star \psi_\lambda(f)|) \tag{4}$$

where, $\psi_\lambda(f) \in \wedge_0(J,K)$. Locally translational invariant descriptors for the higher frequency components can be calculated by obtaining space average over the wavelet modulus coefficients, $|I \star \psi_\lambda(f)|$.

$$S_1 I(f, \wedge_0) = |I \star \psi_\lambda(f)| \star \phi_J \tag{5}$$

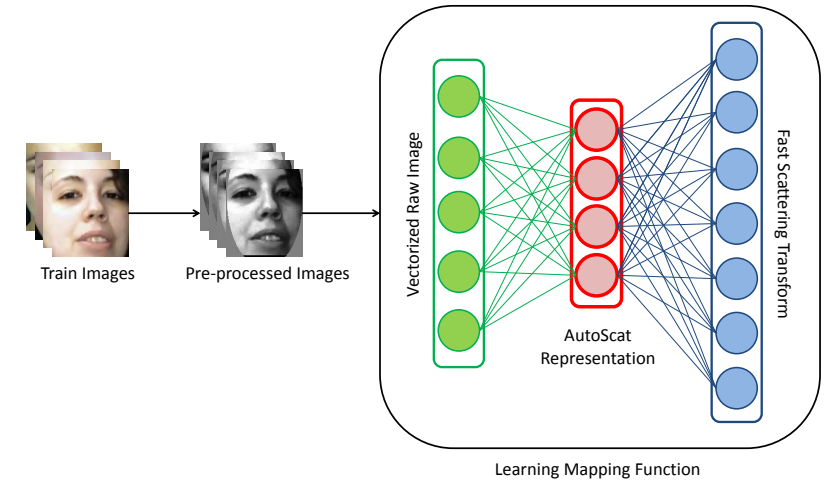
Here, $\psi_\lambda(f) \in \wedge_0(J,K)$. These are called *first-order* scattering coefficients or *first layer* of scattering coefficients. These are computed using a second wavelet modulus transform $|W_2|$ applied to $|I \star \psi_\lambda(f)|$ for each $\psi_\lambda(f) \in \wedge_0(J,K)$, which also provides complementary high-frequency wavelet coefficients:

$$|W_2(I, \wedge_1)| = \Big(|I \star \psi_{\lambda'}(f)| \star \phi_J(f), ||I \star \psi_{\lambda'}(f)| \star \psi_{\lambda''}(f)|\Big) \tag{6}$$

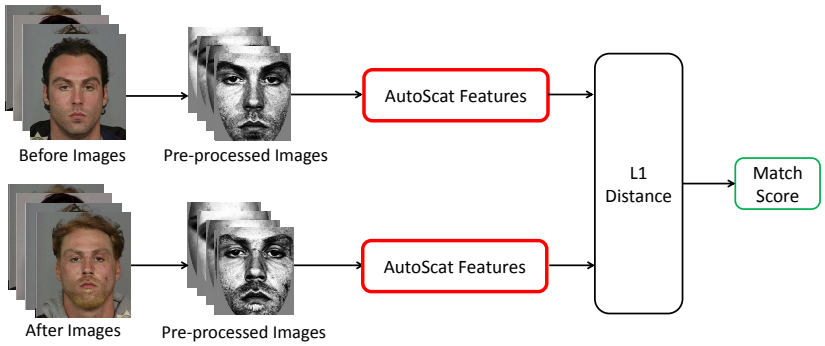
where, $\psi_{\lambda'}(f) \in \wedge_0(J,K)$ and $\psi_{\lambda''} \in \wedge_1(J,K)$. Here $\wedge_0$ and $\wedge_1$ are two different filter banks. These coefficients are averaged by the low-pass filter ϕ_J, just like it is done for computing the first-order coefficients. It defines *second-order* scattering coefficients (or *second layer* scattering coefficients), given by:

$$S_2 I(f, \wedge_0, \wedge_1) = ||I \star \psi_{\lambda'}(f)| \star \psi_{\lambda''}(f)| \star \phi_J(f) \tag{7}$$

These averages are computed by applying a third wavelet modulus transform $|W_3|$ to $||I \star \psi_{\lambda'}(f)| \star \psi_{\lambda''}(f)|$ for each $\psi_{\lambda'}(f) \in \wedge_0(J,K)$ and $\psi_{\lambda''} \in \wedge_1(J,K)$. It computes their wavelet modulus coefficients through convolutions with a



(a) Training Phase



(b) Test Phase

Figure 2: Schematic diagram of the proposed algorithm. In the training phase, fast scattering transform $\mathcal{S}$ is computed over face images. Using these features as the target layer, we train the weights W_1 for the hidden layer h of our neural network architecture. The output layer, $\hat{\mathcal{I}}$, of the architecture is computed by applying weights W_2 and sigmoid activation σ over the hidden layer. Optimal weights are computed by optimizing the error between $\hat{\mathcal{I}}$ and $\mathcal{S}$. The learned weights W_1 are used to map face images in the test set to a lower dimensional representation of their scattering wavelet transform and then matched using L1 distance metric.

new set of wavelets $\psi_{\lambda'''}$ having scaling and orientation parameters defined by λ''' (such that $\psi_{\lambda'''} \in \wedge_3(J,K)$, where $\wedge_3(J,K)$ is a new wavelet bank). Further layers can be computed by recursively applying this process, thus defining a scattering wavelet transform network, which can be extended to any order m, using a constant wavelet bank $\wedge_{m-1}$ at each step. Higher order scattering coefficients provide more robust feature representation, which are stable to affine deformations and are locally invariant.

For practical purposes, for a scattering wavelet transform network of order m, $\wedge_i = \wedge_j$ where $i \neq j$ and $i,j \in 0,1,...,m-1$, that is, the same constant wavelet bank is used at each layer of the network. Thus, the scattering wavelet coefficients at the m^{th} layer can be defined as:

$$\begin{aligned} S_m I(f,\wedge) &= |||I \star \psi_{\lambda_1}| \star \psi_{\lambda_2}|...| \star \psi_{\lambda_i}|...| \star \psi_{\lambda_m}| \star \phi_J \\ &= S[p]I(f,\wedge) \star \phi_J \end{aligned} \tag{8}$$

$\phi_{\lambda_i}(f) \in \wedge$ for $i = 1,2,...,m$, and $p = \{\phi_{\lambda_1}(f), \phi_{\lambda_2}(f), ..., \phi_{\lambda_i}(f), ..., \phi_{\lambda_m}(f)\}$ is an ordered set (or *path*) of wavelets. Here $p \in \wedge^m$ and

$$\wedge^m = \underbrace{\wedge \times \wedge \times ... \times \wedge}_{\text{m times}} \tag{9}$$

where, $\times$ represents Cartesian product.

2.2. Fast Scattering Computation

Fast scattering transform is defined over sets $p \in \mathcal{P}^m_{\downarrow}$ of frequency decreasing paths of length $m \leq \bar{m}$. For $m = 0$,

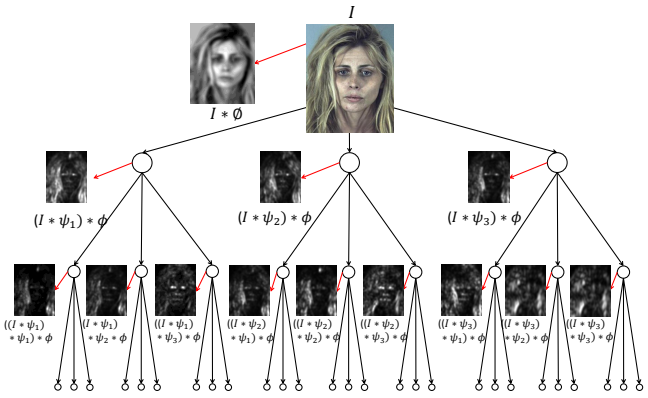


Figure 3: Visualization of the output of scattering wavelet transform over a face image.

$\mathcal{P}_{\downarrow}^{0} = \phi$ and it corresponds to the original image I itself. Let $[p + \lambda]$ be the path achieved by appending $\psi_\lambda \in \wedge$. At the $(m+1)^{th}$ layer of scattering network, the computation of $S_{m+1}I(f, \wedge) = |S[p]I(f, \wedge) \star \psi_\lambda(f)|$ can be reduced by subsampling this convolution at intervals $\alpha 2^j$, with $\alpha = 1$, if $\lambda = \{j, \theta\}$. To avoid aliasing, $\alpha = 1/2$ can be used.

for $m = 1$ *to* $\bar{m}$ **do**
  for $\forall p \in \mathcal{P}_{\downarrow}^{m-1}$ **do**
    Output
    $S_m I(\alpha 2^J f, \wedge) = S[p]I(\alpha 2^J f, \wedge) \star \phi_J(\alpha 2^J f)$
  end
  for $\forall [p + \lambda_m] \in \mathcal{P}_{\downarrow}^{m}$ *with* $\lambda_m = \{j_m, \theta_m\}$ **do**
    Compute $S[p + \lambda_m]I(\alpha 2^{j_m} f, \wedge) =$
    $|S[p]I(\alpha 2^{j_m} f, \wedge) \star \psi_{\lambda_m}(\alpha 2^{j_m} f)|$
  end
end
for $\forall p \in \mathcal{P}_{\downarrow}^{m}$ **do**
  Output $S_m I(\alpha 2^J f, \wedge) = S[p]I(f, \wedge) \star \phi_J(\alpha 2^J f)$
end

Algorithm 1: Fast Scattering Transform

The scattering network has $K^m\binom{J}{m}$ scattering outputs at layer m. Each element in $S[p]I(\alpha 2^{j_m} n, \wedge)$ is sampled at intervals $\alpha 2^{j_m}$, where $p \in \mathcal{P}_{\downarrow}^{m}$. Thus by induction on m, layer m has a total number of samples equal to $\alpha^{-2}(K/3)^m N$. The $K^m\binom{J}{m}$ scattering outputs at $S_m I(\alpha 2^J n, \wedge)$ are subsampled by 2^J, leading to much fewer coefficients.

To achieve more robust representation of the image, it is advisable to use $\bar{m} \geq 2$, $K \geq 8$ and $J \geq 4$. At such values, the computational complexity of even fast scattering algorithm becomes expensive, in terms of both time and memory. To address this challenge, in the next subsection, we propose an autoencoder-style mapping approach, AutoScat, which learns a lower dimensional and succinct representation of the scattering wavelet transform coefficients of face images.

2.3. AutoScat: Learning Succinct ScatNet Features

Fast scattering transform combines the outputs from each layer of the scattering tree to form the resultant feature representation, and thus, as the number of levels increases, the space complexity increases exponentially. Also, as elaborated by the previous section, the time complexity is very high even for the minimally optimal parameters. We proposed an algorithm which learns two linear maps:

1. from the raw image to a lower dimensional space, and

2. from the lower dimensional space to the fast scattering transform features.

It follows the encoding-decoding principle of autoencoders with the only difference that it does not encode the raw input image, rather, it encodes the *mapping* from input image to scattering wavelet transform. The learnt weights of mapping raw image to lower dimensional space are then used to compute the *AutoScat* features.

Let the input vector be $\mathcal{I} \in \mathbb{R}$ of dimensionality N, which is the vectorized form of the raw image $I \in \mathbb{R}^2$ with N pixels. The scattering wavelet transform coefficients of image I with wavelet bank $\wedge$ (characterized by 2^J scales and K orientations) at layer m is represented by $S_m(f, \wedge)$. Let the target vector be $\mathcal{S} \in R$, such that $\mathcal{S}$ is the concatenation of the vectorized formats of each element in fast scattering transform $S_m(, \wedge)$ for $m \leq \bar{m}$. The dimensionality S of the target layer is the number of coefficients in the scattering wavelet transform, given by equation (10), that is, $P = N\alpha^{-2}2^{-2J}\sum_{m=0}^{\bar{m}} K^m\binom{J}{K}$. Let $CM(.)$ be a vector function which performs column major operations over 2-dimensional vectors. Thus, $\mathcal{I} = CM(I)$ and $\mathcal{S} = concat(CM(S_m(f, \wedge)))_{m=\{0,1,..,\bar{m}\}}$.

Let the hidden layer have a dimensionality A, which is less than both the dimensionality of the input vector as well as the target vector. Let the weights defining the encoding part of the proposed model be W_1, with dimensionality $N \times A$. Similarly, let the weights defining the decoding part be W_2 with dimensionality $Q \times S$. We have used unit activation at the encoding phase, and sigmoid activation at the decoding phase. Thus, the output vector $\hat{\mathcal{I}}$ is given as

$$\hat{\mathcal{I}} = \sigma(W_2^T W_1^T \mathcal{I}) \tag{10}$$

where, σ is the sigmoid activation function. The loss function for the proposed architecture is thus given by

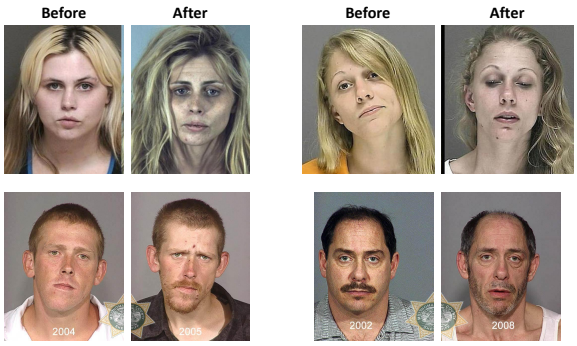


Figure 4: Sample images from the Illicit Drug Abuse Face (IDAF) database [9].

$$\begin{aligned} J(W_2, W_1) &= min||\hat{\mathcal{I}} - \mathcal{S}||_2 \\ &= min||\sigma(W_2^T W_1^T \mathcal{I}) - \\ &\quad concat(CM(S_m(f, \wedge)))_{m=\{0,1,..,\bar{m}\}}||_2 \end{aligned} \tag{11}$$

To obtain a sparse representation of the scattering wavelet transform coefficients, a sparsity constraint is added to the hidden units of the neural network, leading to the following optimization [8]:

$$J(W_2, W_1) = min\Big(\sum_{i=1}^{P} ||\hat{\mathcal{I}}_i - \mathcal{S}_i||_2 + \beta \sum_{j=1}^{A} KL(\rho||\mathcal{S}_j)\Big) \tag{12}$$

where, $\beta \sum_{j=1}^{A} KL(\rho||\mathcal{S}_j)$ is the *sparsity penalty term*. Here, β is the weight for the penalty term, ρ is the constant sparsity level, and KL(.) is the KL-Divergence metric given as:

$$KL(\rho||\hat{\rho_j}) = \rho log\frac{\rho}{\hat{\rho_j}} + (1-\rho) log\frac{1-\rho}{1-\hat{\rho_j}} \tag{13}$$

Thus, the proposed algorithm tries to learn a linear mapping of the raw image to a sparse representation of its scattering wavelet transform representation.

3. Results and Analysis

To show the efficacy of the proposed algorithm, we have used the Illicit Drug Abuse Face (IDAF) database [9], which contains before and after drug abuse face images of a 105 subjects. Figure 4 shows sample images from the database.

3.1. Experimental Protocol

The proposed algorithm is trained on a set of frontal face images under variations like expression and makeup, downloaded from the internet. The experiments are performed in identification mode and the results are reported on the unseen test set. For all the images, face images are extracted, aligned and normalized using CSU FaceEval Toolbox [1], which normalizes face images to a standard size of 150×130. Thus, for all further calculations, we have used $N = 150 \times 130 = 19500$. To compare the results of the proposed algorithm with existing state-of-the-art, the results are compared with two Commercial Off The Shelf (COTS) systems. These are the same matchers used in [9].

Table 1: Summarizing the performance in terms of identification accuracy, feature length and time.

Algorithm	Rank-10 Accuracy (%)	Feature Length	Time (ms)
ScatNet	59.05	134691	221.7
AutoScat	50.48	323	0.19

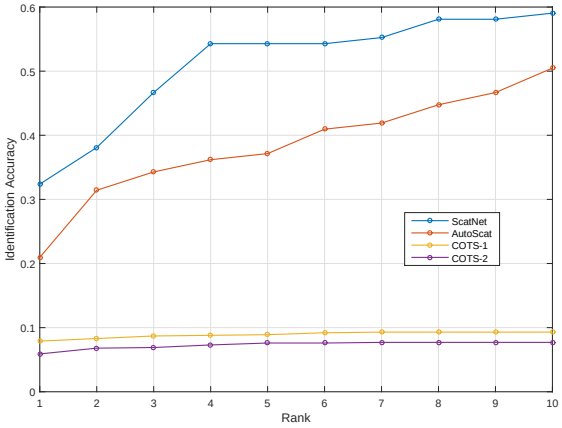


Figure 5: CMC curves for two proposed face descriptors: ScatNet and AutoScat, and two COTS.

3.2. Results and Analysis

For the experiments, we use *before* images as gallery and *after* images as probe. The results are demonstrated in terms of Cumulative Matching Characteristic (CMC) plots pertaining to the performance of ScatNet, AutoScat and two COTS systems. Figure 5 shows the CMCs on the IDAF database and Table 1 summarizes the results in terms of identification accuracy and time. The key analysis are,

- Among all the algorithms and COTS, ScatNet yields the best rank-10 accuracy. It is to be noted that since the number of samples in the IDAF database is small (105 subjects), even a small number of misclassified samples leads to large shifts in accuracy. For example, the difference in rank-10 accuracy of ScatNet and AutoScat corresponds to misclassification of just 9 subjects.
- The CMC curves corresponding to AutoScat features

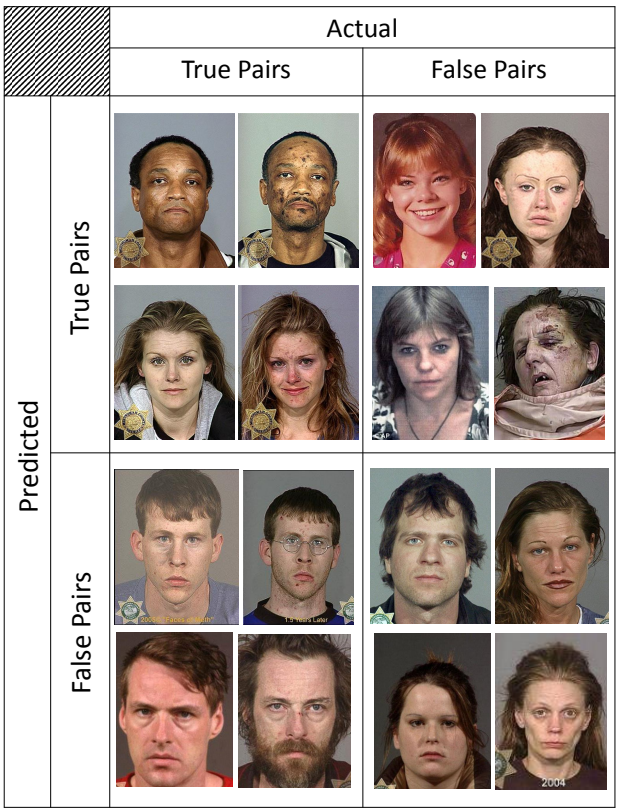


Figure 6: Sample results demonstrating the performance of AutoScat on the IDAF database at rank-10.

and ScatNet take almost the same shape demonstrating that AutoScat is able to learn a reduced dimensional representation of ScatNet. The feature space of AutoScat is much smaller than that of ScatNet; i.e. number of dimensions is reduced by a factor of 417.

- The computational time required by ScatNet is of the order of 10^3 times as compared to the time required by AutoScat. It means that a system using AutoScat will be able to process a thousand image pairs, whereas in the same time a system using ScatNet will only be able to process one pair.

- The confusion matrix of sample results shown in Figure 6 illustrates that expression and pose variations affect the performance of AutoScat; however, it is able to correctly match faces with major physiological variations.

- Compared to commercial systems, ScatNet and AutoScat perform significantly better; the differences between the rank-10 accuracies of AutoScat and COTS systems are more than 40%.

4. Conclusion

This research addresses the problem of matching face images with illicit drug abuse variations which introduces both local and global variations in face images. We propose a scattering transform based face recognition algorithm to match face images. We further propose a neural network based mapping approach, AutoScat, that learns the ScatNet representation of face images. This learnt network yields a concise representation of ScatNet which reduces the time complexity by an order of 10^{-3} and also reduces the memory requirements significantly. The experiments are performed on the illicit drug abuse face database and the results show that the proposed ScatNet and AutoScat algorithms yield improved performance compared to existing face recognition algorithms. As a future research direction, we are currently exploring methods to further improve the accuracy of AutoScat without increasing computational requirement.

References

[1] D. S. Bolme, J. R. Beveridge, M. Teixeira, and B. A. Draper. The CSU Face Identification Evaluation System: Its Purpose, Features, and Structure. In *Proceedings of 3rd International Conference on Computer Vision Systems*, pages 304–313. Springer-Verlag, 2003.

[2] J. Bruna and S. Mallat. Invariant scattering convolution networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(8):1872–1886, Aug 2013.

[3] A. Dantcheva, C. Chen, and A. Ross. Can facial cosmetics affect the matching accuracy of face recognition systems? In *Proceedings of IEEE International Conference on Biometrics: Theory, Applications and Systems*, 2009.

[4] T. Dhamecha, R. Singh, M. Vatsa, and A. Kumar. Recognizing disguised faces: Human and machine evaluation. In *PLoS ONE 9(7)*, 2014.

[5] A. K. Jain and S. Z. Li. *Handbook of Face Recognition*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2005.

[6] L. Sifre and S. Mallat. Rotation, scaling and deformation invariant scattering for texture discrimination. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 1233–1240, June 2013.

[7] R. Singh, M. Vatsa, H. Bhatt, S. Bharadwaj, A. Noore, and S. Nooreyezdan. Plastic surgery: A new dimension to face recognition. *IEEE Transactions on Information Forensics and Security*, 5(3):441–448, Sept 2010.

[8] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P.-A. Manzagol. Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. *Journal of Machine Learning Research*, 11:3371–3408, 2010.

[9] D. Yadav, N. Kohli, P. Pandey, S. R., and V. M. Effect of illicit drug abuse on face recognition. *IEEE Winter Conference on Applications of Computer Vision*, 2016.