Journal Pre-proofs

Original article

Transfer Learning for Face Recognition using Fingerprint Biometrics

Rupali Kute, Vibha Vyas, Alwin Anuse

PII:	S1018-3639(21)00110-0
DOI:	https://doi.org/10.1016/j.jksues.2021.07.011
Reference:	JKSUES 548
To appear in:	*Journal of King Saud University - Engineering Sciences*
Received Date:	1 April 2021
Accepted Date:	27 July 2021

Please cite this article as: Kute, R., Vyas, V., Anuse, A., Transfer Learning for Face Recognition using Fingerprint Biometrics, *Journal of King Saud University - Engineering Sciences* (2021), doi: https://doi.org/10.1016/j.jksues.2021.07.011

Transfer Learning for Face Recognition using Fingerprint Biometrics

Rupali Kute[1]
Assistant Professor, School of Electronics and Communication, Dr. Vishwanath Karad MIT- World Peace University, Pune, 411038, Maharashtra, INDIA
rupali.kute@mitwpu.edu.in

Vibha Vyas[2]
Associate Professor, Electronics and Telecommunication, College of Engineering Pune, Savitribai Phule Pune University, Pune, 411038, Maharashtra, INDIA
vsv.extc@coep.ac.in

Alwin Anuse[3]
Associate Professor, School of Electronics and Communication, Dr. Vishwanath Karad MIT- World Peace University, Pune, 411038, Maharashtra, INDIA
alwin.anuse@mitwpu.edu.in

Transfer Learning for Face Recognition using Fingerprint Biometrics

Abstract:
Biometrics is a set of highly automated methods used for recognition purposes including analyzing physical traits of people and statistically measuring them. In forensic applications, fingerprint and face images are mostly used for recognition. There is a need during a criminal investigation to find out a face image of a person from its fingerprint. The proposed method able to identify the face of a person using an associated fingerprint. In this paper, Bregman divergence regularization is used to learn and optimize transferring subspace. This method first gleans knowledge from samples that are meant for training and transfer it to the testing samples. This regularization helps to minimize the probability distribution differences between two different domains. It helps to find a common subspace that boosts the performance of independent and identically distributed (i.i.d.) condition of samples. Practically the samples violate this i.i.d. condition. However, it will help to identify the correct suspect.

Keywords: Dimensionality reduction algorithm, Transfer subspace learning, Bregman divergence-based regularization.

Introduction:
Biometrics and forensics are highly evolved technologies and are being studied in great depth. For high-security purposes like criminal identification, biometric systems are being used. In a forensic department, for finding probable suspects different biometric characteristics are used. Most of the time, at crime scenes, officers searches for biological traces, finger- prints, face images, and palmprint as mentioned in Monwar and Gavrilova (2009); Chin et al. (2009); Cap- pelli et al. (2003). Recently various biometric features like

iris, fingerprints, face, DNA, palm prints, hand geometry, and others are considered for different security applications referred in Meuwly and Veldhuis (2012); Jain and Klare (2011) N (1998). Lamia et al. (2020); Sharkawy and Mostfa (2021); Maya et al. (2021) have used arrtificial intelligence techniques are used to resolve the classification and prediction problem. Larrain et al. (2017) shows the face image recognition under changing pose, occlusion, expression, ambient lighting, distance, and face size using sparse fingerprint classification is possible. In the era of deep transfer learning Curricular Face experimented by Huang et al. (2020) and Fingernet experimented by Minaee et al. (2019b) have used Resnet50 and Resnet100 for feature extraction. As mentioned by Minaee et al. (2019a) deep learning models have leveraged the accuracy of face, fingerprint, ear, and palmprint recognition. The fusion of multi-biometric solves the problem of single biometric from a security applications point of view discussed in detail by Sagiroglu and Ozkaya (2009); Gad et al. (2015). Song et al. (2020) has developed the deep feature model based on Resnet50 has fused ECG, face, and fingerprint .

A huge amount of data has to be used to train the system for better results but, for classification, the major problem is insufficiently labeled data because it is expensive to acquire as well as to label. Hence the authors used data from various databases that were relevant and labeled to facilitate the learning process. The approach of transfer learning is used to handle the issue of learning with a limited number of samples. To bring about an improvement in this process of learning, a large amount of data belonging to other databases is used instead of making use of the limited supply of labeled data that is available. However, this involves some misalignment between the target and the source data. As a result of this, the instant use of the source data can hide the learning which results in a negative transfer of data. A well-known learning method considers the helper or source data for transfer from the assistant to the target domain for boosting the learning execution as mentioned by Dai et al. (2007). The data collected under various conditions and environments used by Pan et al. (2008) said that the samples are all from different domains, it causes inconsistent distribution resulting in poor performance of the system.

Su et al. (2010) Haibinyan and Poo (2011) mentioned that the probability distribution of the dataset used for training and the dataset used for testing is different. The fingerprint and face images are belonging to two different domains. They are having a different probability distribution. To associate face image with fingerprint, both the domains should have a common subspace. Hence for improvement, it is necessary that the knowledge which has been gained at the training level be transferred to the testing level. This process is undertaken by using the transfer subspace learning method as used by Kute et al. (2019, 2020).

In this paper, the authors have to find out a discriminative subspace where data from the source domain is transferred to the target domain. The distribution of the data in the target domain has to be such that there is minimum distribution difference between like data and maximum distribution difference between unlike data Obozinski et al. (2006). If samples are unlabeled in the process of transfer learning, they use various types of source data such as samples from auxiliary tasks, sharing of features from the auxiliary domain, and data from different feature spaces as mentioned by Torrey et al. (2008) and Pan et al. (2008). The auxiliary information differs based on the model, the application, and how it is used. For motivation, if the authors consider an example of classification of different images of a car, then for transfer learning at the training phase images of a bicycle, a truck or any similar vehicle can be used to obtain labeled images of a car.

All the existing reduction or subspace learning techniques such as the Principle Com- ponent Analysis (PCA), Marginal Fisher's analysis (MFA), Fisher's Linear Discriminate Analysis (FLDA), Linear Preserving Projections (LPP), Discriminative Locality Alignment (DLA) Girish et al. (2014); Zadrozny (2004); Zhang et al. (2008) are developed under the i.i.d. assumption. Practically in cross-domain, the performance of the algorithms mentioned above is poor due to the violation of the assumption. Therefore it is crucial to consider a distinction between the probability distribution of preparing and testing the samples. In this paper, there is a need to utilize the Bregman divergence-based regularization which can be related to the current dimensionality reduction algorithms e.g. PCA, FLDA Si et al. (2010); Pan and Yang (2010). By combining this regularization, the authors can acquire a subspace wherein the distinction of dissemination linking the preparation and the testing data set is restricted due to which the i.i.d. presumption is validated in the learned subspace Xia et al. (2012); Xu et al. (2016) Shao et al. (2014). As it ensures the protection of the particular data, it is possible that information from the various classes can be isolated to a great degree. In this paper as a result of the combination of regularization and subspace learning, integration of multiple biometrics like face and fingerprints is possible Si et al. (2012). In- tegration helps find the belonging class or associated face if a fingerprint is given at the testing phase. The proposed model can be used by a forensics department for investigating the correct suspect. This work performs well because face and fingerprints are coming from different domains but these domains able to share some common properties by minimizing the probability distribution difference using regularization.

The organization of the remainder of the paper is as follows: In Section 2, the authors develop a proposed framework for transfer subspace learning. The results of the experiments that the authors have conducted followed by a discussion are set out in Section 3. The Conclusion makes up Section 4 of the paper.

PROPOSED FRAMEWORK

The proposed framework is explained with the help of figure1. In the training phase, features of the face and fingerprint images are extracted using Discriminative Locality Alignment (DLA), these features are fused using Bregman-divergence regularization. The Bregman-divergence regularization reduces the probability distribution difference between face and fingerprint images and tries to bring them into the same subspace. The same label is assigned to face and fingerprint images while training. After several iterations, an optimized subspace is found out. During the testing phase, the reduced dimensional data of face and fingerprint images are encoded using optimized subspace. Face images labeled encoded data are used for KNN training and fingerprint encoded data without using a label for testing. The output of the KNN is the predicted label of the fingerprint.

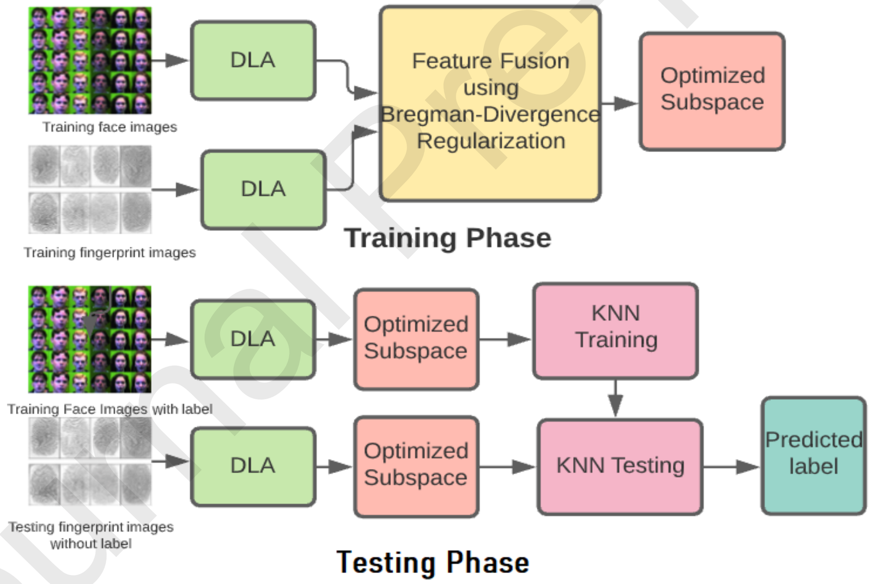

Figure1: Testing and testing phase of the proposed framework

The Principle Component Analysis (PCA), Fisher's linear discriminant analysis (LDA), and Discriminative Locality Alignment (DLA) is the maximally used dimensionality reduction algorithms. However, the problem with the PCA is, it only focuses on reducing the dimension without looking at the separation between classes. Whereas FLDA somehow manages it using within and between-class scatter matrixes, however, it also fails if the data is non-linear. It assumes that all the class samples participating equally in the dimension reduction process and it also suffers from matrix singularity problem. The proposed framework has used DLA to overcome these problems. It operates on part optimization. It imposed the discriminative information associated with every sample and its neighbors over patches and finally each part weighted by margin degree. The Bregman-divergence regularization tries to minimize the distribution difference between face and fingerprint images and bring them together in the common subspace.

For an identification or classification task, given m training samples denoted by $S=\{(x_1,l_1),(x_2,l_2),\ldots,(x_m,l_m)\}$, $x_i \in R^D$, i= 1,2,…,m , x_i denotes the feature vector of the i^{th} individual source sample and l_i is the associated label or index at the source domain. Similarly there are n testing samples denoted by $T = \{y_1,y_2,\ldots,y_n\}$, $y_i \in R^D$, i= 1,2,…,n , y_i represents the feature vector of i^{th} testing sample. The objective of the proposed framework is to find the subspace where the difference between the probability distribution of the training samples and the testing samples is minimized. Moreover, the individual subspace should be discriminative.

2.1 Subspace learning algorithm or dimensionality reduction algorithm

All samples at the source and target domains are drawn from the R^D high-dimensional space. A dimensionality reduction algorithm can be used to reduce these high-dimensional spaces to an R^d low-dimensional subspace, where images from various classes can be segregated very well. Under the assumption of the i.i.d. condition multiple algorithms have emerged. Using these algorithms the authors get highly reduced feature vectors. The objective of this part is the transformation from high to low-dimensional subspace (R^D to R^d) and that can be obtained from:

$U =$ arg min F(U) , ……………………………..

(1)

$U \in R^{D X d}$

Depending on the application the objective function F(U) is constructed from different subspace learning algorithms. In the given work the following algorithms are used:

1)PCA which is optimal for reconstruction and not for classification purposes since it considers the data is Gaussian distributed. It maximizes the mutual information between the high and low dimensional data using the decorrelation of training samples. A PCA projection matrix can maximize the trace of the covariance matrix,

$F(U) = \max \text{trace} (U^TRU)$,

= min (-trace (U^TRU)) ………………….
(2)

Subject to $U^TU=I$. In (1), $R=\sum_{i=1}^{U}(y)(y)^T$

Where $y= (x_i - m)$, R is the autocorrelation matrix of the source data, and m is the training data mean.

2) FLDA preserves the class structure which is defined by the scatter matrix (data covariance matrix). The FLDA objective function can maximize the ratio of a trace between the class scatter matrices and within-class scatter matrices for class structure preservation. The between-class scatter matrix is given by $S_B=\sum_{i=1}^{N} ni(p)(p)^T$ where $p= (m^i - m)$.

And the within-class scatter matrix is given by $S_W=\sum_{i=1}^{N}\sum_{j=1}^{ni} ni(q)(q)^T$

where $q=(x_j^i - m^i)$, where N is the class number, n_i represents how many samples there are in the i^{th} class, m^i is the mean of the samples from the i^{th} class, and m is the mean of all the samples. Finally, the FLDA projection matrix is given by:

$F(U) = trace^{-1}(U^TS_BU) * trace(U^TS_WU)$…………
.(3)

3) Zang t. et. Al.(2008) shows that DLA overcomes the issue by giving a margin degree measure to every sample. The calculation works in the accompanying three phases.

1)To begin with, to a limited extent discriminative data is forced over patches, each of which is related to one sample data and its neighbors. This patch is made by combining the given sample x_i with k1 being the nearest sample i.e. $x_{i1},x_{i2},\ldots,x_{ik1}$ from the same class and k2 being the nearest samples i.e. $x_i^1,x_i^2,\ldots,x_i^{k2}$ in different class as: $X_i=[\ x_i,\ .\ x_{i1},x_{i2},\ldots,x_{ik1}, x_i^1,x_i^2,\ldots,x_i^{k2}]$;

2) Next, in test weighting, the streamlining of each part is weighted by the margin degree, a measure of the significance of a given data given by:

$m_i= \exp(-\frac{1}{(ni+\delta)t})$;

3) Lastly, in the entire arrangement, the alignment trick is utilized to adjust all the weighted part optimization to the entire streamlining. Finally, the objective function of this can be given as:

$F(U) = \text{trace}(U^T X L X^T U)$,…………………. (4)

Where L represents the discriminative data with local geometry and $L=\sum_{i=1}^{N} SimiLiSi^T$ Where S_i is the selection matrix.

The above method performs poorly when data are drawn from different domains as this means violation of the i.i.d. condition.

2.2 Transfer subspace learning

To solve these problems the authors use the Bregman divergence based regularization Si et al. (2010) in which the probability distribution difference of the source samples and the target samples has been incorporated in the projected subspace U. The regularization parameter denoted by $D_U(P_S \| P_T)$ is integrated with equation no.(1) to obtain:

$$U = \arg \min_{U \in RD*d} F(U) + \lambda D_U(P_S \| P_T) \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad \ldots(5)$$

In (5), F(U) is an objective function that is obtained from eq.(4) and depends on the application. P_S and P_T are probability distribution of the training samples and testing samples respectively. For controlling the tradeoff, λ is used which is the regularization parameter. If prior knowledge of the samples in equation (1) is ignored, then the overfitting problem occurs. To reduce this, regularization is used which reduces the space volume also. As in Support Vector Machine, the norm is introduced in the kernel Hilbert space N (1998).Similarly, divergence-based regularization is introduced in this work to solve the overfitting and ill-posed problem.

The Bregman divergence-based regularization, used for measuring the distance between $P_S(\vec{y})$ and $P_T(\vec{y})$, is a function that is convex and is given by:

$D_U(P_S \| P_T) = \int d(\xi(P_S(\vec{y})), \xi(P_T(\vec{y}))) d\mu$ …………………… (6)

Where dμ i.e. $du(\vec{y})$ is the Lebesgue measure. The right-hand side of (6) is also called the U-divergence on the subspace R_d. By utilizing the quadratic divergence for measuring the distribution difference, the authors try to achieve a computationally tractable implementation of (6) as:

$D_U(P_S \| P_T) = \int ((P_S(\vec{y})) - (P_T(\vec{y})))^2 \, d\vec{y}$

$= \int ((P_S(\vec{y}))^2 - 2 (P_S(\vec{y})) (P_T(\vec{y})) + (P_T(\vec{y}))^2) \, d\vec{y}$ ………..(7)

To assess the circulation P_S and P_T in the anticipated subspace W, in this work, the authors start applying the kernel density estimation (KDE) strategy, which assesses the density at $y \in R_d$ as a total of kernels amongst $\vec{y}$ and each of the other samples $\vec{y}_i$ i.e.:

$P(\vec{y}) = (1/n) \sum_{i=1}^{n} G_{\Sigma}(\vec{y} - \vec{y}_i)$. Where n is the number of samples and $G_{\Sigma}(\vec{y})$ is the d-dimensional Gaussian kernel with the covariance lattice Σ.

Finally using an optimization technique like the gradient descent technique, the authors try to attain an optimal subspace U which is a linear subspace given by:

$U_{k+1} = U_k - \eta(k)\left(\frac{\partial F(U)}{\partial U} + \lambda \sum_{i=1}^{m+n} \frac{\partial DU(PS \| PT)}{\partial U}\right)$ ………………..(8)

Where η(k) is the kth learning rate factor. The derivative of DU(PS‖PT) concerning U is given in Si et al. (2010). Using a different F(U) and eq.(8) the authors have find out a solution by iteratively imposing $U^TU=I$ for association purposes.

EXPERIMENTS AND RESULT ANALYSIS

In this section, the authors undertake a comparison between traditional subspace learning algorithms like PCA, FLDA, DLA and transfer subspace learning like TPCA, TFLDA, TDLA by seeing the performance accuracy. From the given experiment the authors prove that transfer subspace learning gives better results as compared to PCA, FLDA, and DLA algorithms.

3.1 Data Preparation:

Primarily the authors used existing face data like YALE and FERET. The FERET informational set has 13,539 face pictures gathered from 1,565 people, where various sized pictures are taken. The postures, enlightenments, and facial appearances are also varied. Out of the total the authors use 700 images of 100 persons each having 7 different images. For the investigation of cross-domain face recognition, the authors need to construct two other informational collections: 1)Y2F: The YALE and FERET dataset is used for training and FERET dataset is used for testing; 2) F2Y: The YALE and FERET is the dataset used for training and YALE dataset is used for testing; After doing a comparative study on the cross-domain face data the main work is done which means applying the methods over face and fingerprint data for integration. As per a biometric review, an open face and unique fingerprint database of the same person does not exist which is available for cross biometric integration. Therefore, similar to Y2F and F2Y the authors used the proposed structure on a private database that contains data of 220 persons' faces and unique fingerprints of the same individuals. Out of the database of 220 people, 100 people's data is chosen randomly containing ten sets of data for faces and 10 fingerprint images of every individual. Given this database, the authors arranged the informational collections for biometric integration as FACE2FINGER where the face is used for the training phase and fingerprints are used for the testing phase.

3.2 Cross-domain face recognition :

In the testing phase, the marking data of the testing pictures are invisible regarding each of the subspace learning algorithms. During the testing stage, the naming data of the reference pictures is accessible just for classification. Therefore the authors figured the separation between a testing picture and each reference picture and foresaw that the closest reference picture was the label of the testing picture. Figure 2 shows the recognition rates versus the dimensions of the subspace under the Y2F experimental setting.

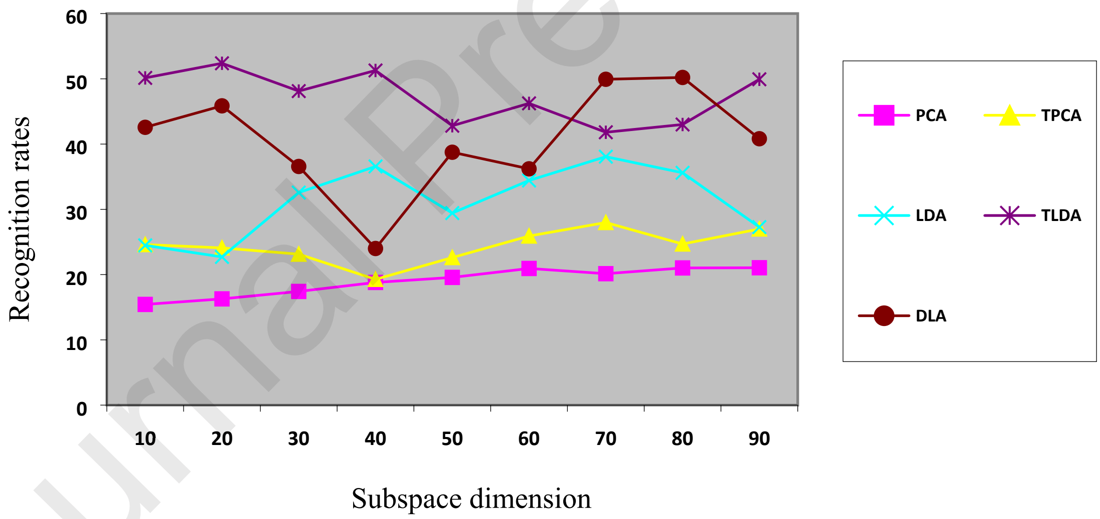

Figure 2: Recognition rates versus subspace dimensions for Y2F experimental settings

Figure 3 shows the recognition rates versus the dimensions of the subspace under the F2Y experimental setting.

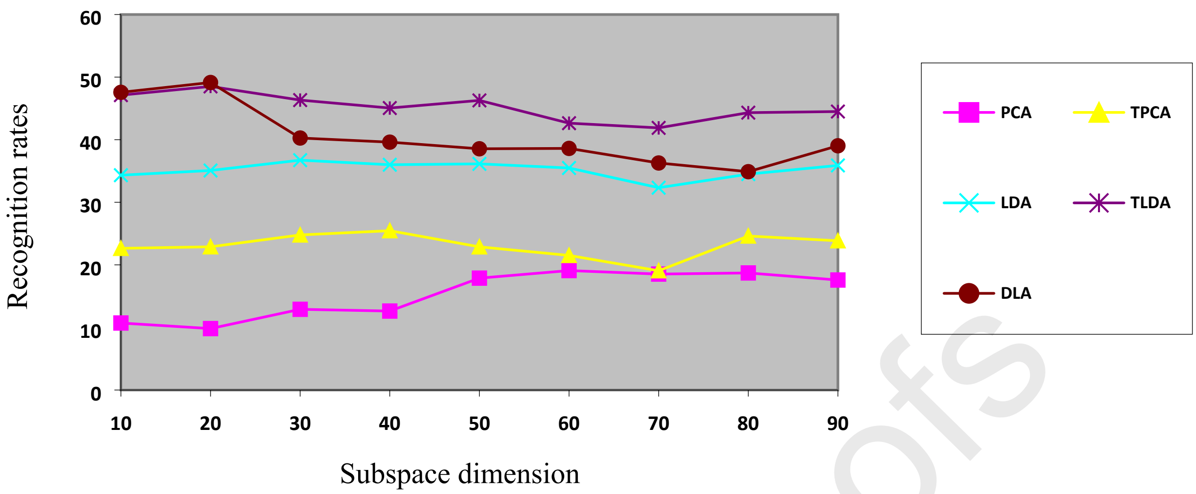

Figure 3: Recognition rates versus subspace dimensions for F2Y experimental setting.

3.3 Biometric Integration :

It is very crucial to have testing and training spaces share some normal data as processing on cross biometric integration. In this manner for a Face2Finger cross biometric integration plan, the procedure is to give the labeling data of the face and no labeling data of fingerprint images in the training phase; whereas, in the testing stage, the point is to discover the labeling of the fingerprint images of an unknown user. To start with, every calculation is applied to the sample to be used during training to take in the projection matrix for a specific subspace learning calculation. Secondly, every test to be used during the testing phase is projected onto a low-dimensional subspace using a projection matrix. Finally, the authors figure the separation between a testing test and every reference test and make use of the Nearest Neighbor (NN) classifier for predicting the label of this particular testing test. Figure 3 shows the recognition rates versus the dimensions of the subspace under the FACE2FINGER experimental setting and table 1 depicts the recognition rates for the different values of subspace dimensions for the FACE2FINGER experimental setting.

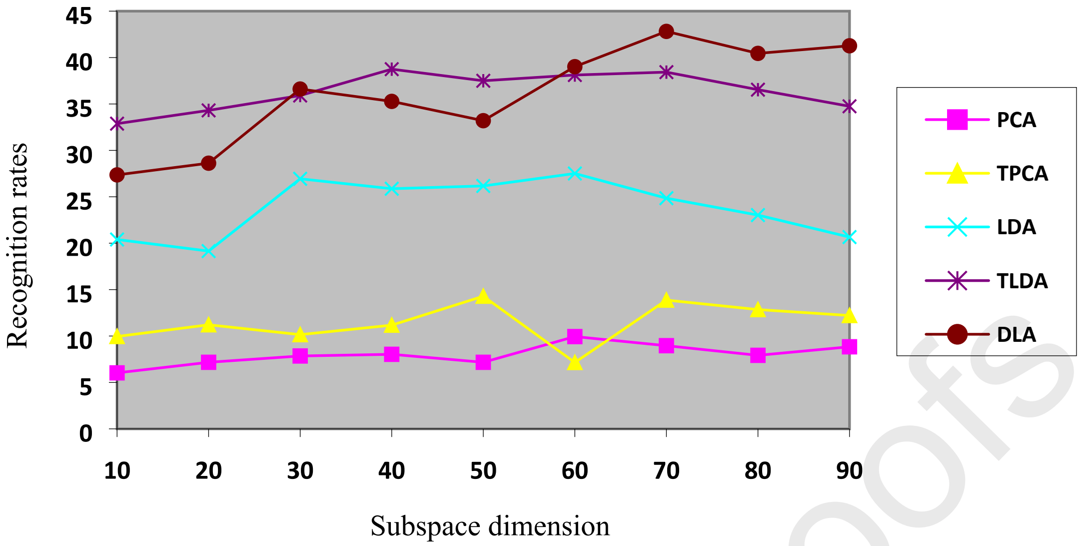

Figure 3 : Recognition rates versus dimensions of subspace for FACE2FINGER experimental setting

Finally, Table 1 shows the best recognition rate of various algorithms under three different experimental settings.

Table 1 : Best recognition rate of various algorithms under three experimental settings

Subspace dimension / Diff. Algorithm	Y2F	F2Y	FACE2FINGER
PCA	21.05(90)	19.09(60)	9.94(60)
TPCA	27.98(70)	25.46(40)	14.29(50)
LDA	38.07(70)	36.73(30)	27.51(60)
TLDA	52.38(20)	48.51(20)	38.75(40)
DLA	50.21(80)	49.14(20)	50(70)
TDLA	52.64(80)	50.21(50)	60(90)

The number in the ( ) represents the number of dimensions. Table 1 results depict that transfer DLA gives better performance as compared to other methods.

In Eq. (8), η is the learning rate and λ is the regularization parameter. The η and λ parameters control the transfer of information while bringing the subspace together in the same space. Making λ parameter high transfers more information however, very high values make most of the values of optimized subspace to zero. In contradiction to λ, the learning rate η controls the distribution difference between face and fingerprint domains. The small value of η keeps the distribution difference small however, very small values of η transfers less knowledge. Therefore, values of η and λ are determined by changing the value of one parameter and keeping another parameter constant. The best classification accuracy is obtained at $\eta = 0.01$ and $\lambda = 100$.

Conclusion

In most of the research done so far, different biometrics is not integrated. To handle this issue, transfer learning with Bregman divergence-based regularization is used and it minimizes the difference of the probability distribution between the face and fingerprint biometrics over an existing subspace learning algorithm like PCA, FLDA, DLA etc. and maps them in a mutual subspace. The given data set violates the i.i.d. assumption by using Bregman's regularization which transfers distinct information obtained from the training data to the testing data by eliminating the distribution difference and at the same time preserving the knowledge obtained from the training phase. Overall by using the proposed framework, the authors can map several features having different space representations in a single common subspace which can minimize the time and space complexity. This proposed structure helps in forensic science; by using even a single biometric other related biometric data can be predicted.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References:

Cappelli, R., Maltoni, D., Maio, D., Jain, A. K., Prabhakar, S. 2003, Synthetic Fingerprint Generation, Hand- book of Fingerprint Recognition 232 pages 203–203, 2003.

Chin, T. S. Y. J., Ong, B. Y., Hiew 2009. Integrating Palmprint and Fingerprint for Identity Verification. Proceeding of International Conference on Network and System Security, pages 437–442.

Dai, W., Yang, Q., Xue, G. R. 2007. Boosting for Transfer Learning. Proc. 24th Int'l Conf. Machine Learning, pages 193–200.

Gad, R., El-Fishawy, N., EL-SAYED, A., Zorkany, M. 2015. Multi-Biometric Systems: A State of the Art Survey and Research Directions. International Journal of Advanced Computer Science and Applications, 6(6).

Girish, G. N., Shrinivasanaika, C. L., Pradip, K., Das 2014. Face Recognition using MB-LBP and PCA: A Comparative Study. IEEE International Conference on Computer Communication and Informatics, pages 1–6.

Haibinyan, M. H. Poo, A. N. Cross -dataset Facial Expression Recognition, Proceeding of IEEE Inter- national Conference on Robotics and Automation 5985–5990 2011.

Huang, Y., Wang, Y., Tai, Y., Liu, X., Al, E., Huang Curricularface: adaptive curriculum learning loss for deep face recognition. proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020.

Jain, A. K. Klare, B. 2011. Matching Forensic Sketches and Mug Shots to Apprehend Criminals.

Computer, 44(5):94–96.

Kute, R. S., Vyas, V., Anuse, A. 2019. Component-based face recognition under transfer learning for forensic applications. Information Sciences, 476:176–191.

Kute, R. S., Vyas, V., Anuse, A. 2020. Association of Face and Facial Components Based on CNN and Transfer Subspace Learning for Forensics Applications. SN Computer Science, 1(5):287–287.

Lamia, H., Shehab, O. M., Fahmy, Safa, M., Gasser, M. S., El-Mahallawy,2020, An efficient brain tumor image segmentation based on deep residual networks (ResNets), Journal of King Saud University - Engineering Sciences, 2020.

Larrain, T., Bernhard, J. S., Mery, D., Bowyer, K. W. 2017. Face Recognition Using Sparse Fingerprint Classification Algorithm. IEEE Transactions on Information Forensics and Security, 12(7):1646–1657.

Maya, R., Hassan, B., Hassan, A. 2021. Develop an artificial neural network (ANN) model to predict construction projects performance in Syria. Journal of King Saud University - Engineering Sciences.

Meuwly, D. Veldhuis, R. 2012. Forensic biometrics: From two communities to one discipline. IEEE Proceed- ings of the International Conference of the Biometrics Special Interest Group, pages 1–12.

Minaee, S., Abdolrashidi, A., Su, H., Bennamoun, M., Zhang, D. 2019a, Biometric recognition using deep learning: A survey. arXiv preprint arXiv:1912.00271 2019a.

Minaee, S., Azimi, E., Abdolrashidi, A. 2019b, Fingernet: Pushing the limits of fingerprint recognition using convolutional neural network." arXiv preprint arXiv:1907.12956 2019b.

Monwar, M. M. Gavrilova, M. L. 2009. Multimodal Biometric System Using Rank-Level Fusion Approach. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 39(4):867–878. N, V. 1998.

Obozinski, G., Taskar, B., Jordan, M. I. Multi-Task Feature Selection, technical report, Dept. of Statistics, Univ. of California,Berkeley, 2006.

Pan, J. Yang, Q. 2010. A Survey on Transfer Learning. IEEE Transactions on Knowledge and Data Engg, 22(10):1345–1359.

Pan, S. J., Kwok, J. T., Yang, Q. 2008. Transfer Learning via Dimensionality Reduction. Proc. 23rd Nat'l Conf. Artificial Intelligence, pages 677–682.

Sagiroglu, S. Ozkaya, N. 2009. Artificial Neural Network Based Automatic Face Parts Prediction System from Only Fingerprints. Proceedings of IEEE Workshop on Computational Intelligence in Biomet- rics:Theory, Algorithms and Applications.

Shao, M., Kit, D., Fu, Y. Face Recognition using MB-LBP and PCA: A Comparative Study, IEEE International Conference on Computer Communication and Informatics (2014) 1–6.

Sharkawy, A.-N. Mostfa, A. A. 2021. Neural networks' design and training for safe human-robot cooperation. Journal of King Saud University - Engineering Sciences.

Si, S., Tao, D., Geng, B. 2010. Bregman Divergence-Based Regularization for Transfer Subspace Learning.
IEEE Transactions on Knowledge and Data Engineering, 22(7):929–942.

Si, S., Tao, D., Wang, M., Chan, K. P. 2012. Social image annotation via cross-domain subspace learning. Multimedia Tools and Applications, 56(1):91–108.

Song, H.-K., AlAlkeem, E., Yun, J., Kim, T.-H., Yoo, H., Heo, D., Chae, M., Yeun, C. Y. 2020. Deep user identification model with multiple biometric data. BMC Bioinformatics, 21(1):315–315.

Su, Y., Fu, Y., Tian, Q., Gao, X. 2010. Cross-database Age Estimation Based on Transfer Learning. Proceeding of IEEE International Conference on Acoustics, Speech and Signal Prcocessing, pages 1270– 1273.

Torrey, L., Shavlik, J., Natarajan, S., Kuppili, P., Walker, T. 2008. Transfer in Reinforcement Learning via Markov Logic Networks. Proc. AAAI Workshop Transfer Learning for Complex Tasks.

Xia, S., Shao, M., Luo, J., Fu, Y. 2012. Understanding Kin Relationships in a Photo. IEEE Transactions on Multimedia, 14(4):1046–1056.

Xu, Y., Fang, X., Wu, J., Li, X., Zhang, D. 2016. Discriminative Transfer Subspace Learning via Low- Rank and Sparse Representation. IEEE Transactions on Image Processing, 25(2):850–863.

Zadrozny, B. 2004. Learning and Evaluating Classifiers under Sample Selection Bias. Proc. 21st Int'l

Conf.
Machine Learning, pages 114–121.
Zhang, T., Tao, D., Yang, J. 2008. Discriminative Locality Alignment. Proc. 10th European Conf. Computer Vision, pages 725–738.