# Fusion of Logically Concatenated Cross Binary Pattern and ε-Dragging Linear Regression for Face Classification across Poses

Kumud Arora and Poonam Garg

*Abstract*—**Face recognition across poses is a fundamental challenge in face recognition. The appearance of the face undergoes drastic changes with changes in the pose angle of the face. The popular uniform texture-based descriptor, Linear Binary Pattern (LBP), loses its appeal as the underlying intensity information changes. During pose changes, crucial facial key points get concealed and henceforth the rate of face recognition falls. In this paper, we propose utilizing an LBP-based texture descriptor, the Logically Concatenated Cross Binary Pattern, for extracting contour information from face images. The produced feature descriptors show a superior performance in terms of computational complexity. In order to retain only discriminant features, a balanced approach, like Marginalized Elastic-Net Regularization with ε-dragging, is fixed into the regularization term of the linear regression model. Classification is realized using regression residuals in the least-squares sense. In order to make the proposed approach robust to poses, Procrustes analysis is utilized at the preprocessing step when pose variations are significant. A comparative investigation of the proposed approach is presented by utilizing the UMIST database, the AT&T database, and the Indian Movie Face Database.**

*Index Terms*—**Discriminative elastic-net regularization, logically concatenated cross binary pattern, marginalized elastic-net regularized linear regression, negative dragging linear regression.**

## I. INTRODUCTION

In face recognition, there is an inherent challenge of pose changes. Any small change in the pose of the face given for recognition can cause a colossal change in the image's pixel information, thereby adversely affecting the recognition of information. Linear Binary Pattern (LBP) [1], a popular face descriptor, fails to generate a reliable recognition rate in the presence of pose changes, as it is based on the intensity information, which changes with face pose changes. To combat pose changes, sufficient discriminative information needs to be retained in order to overcome the information loss incurred during feature occlusion. Two approaches are commonly used to combat pose changes. The first approach is to craft a pose-invariant descriptor, and the second is to reconstruct the front face image or desired pose from the input pose image. The first approach of crafting a pose-invariant

descriptor is expected to have three essential features: retaining enough discriminative information to overcome information loss that occurs during feature occlusion, low dimension, and ease of computation. The second approach involves modeling techniques based on shape analysis and shape modeling. Procrustes analysis is one of the shape analysis approaches utilized for analyzing the given shape set distribution in terms of linear transformation (translation, scaling, reflection, and orthogonal rotation), and the analysis is used to attain the minimum value of the dissimilarity measure between the given shape and the required shape.

Driven by the need for a fast response time in face recognition, less computation complexity is desired at the feature extraction step. For front face image recognition, a plethora of feature descriptors have been investigated and applied. Liu *et al.* [2] mentioned that the highest classification accuracies are provided by the methods that combine LBP with other complementary features, after comparing 32 recent most promising LBP variants. However, in case of face recognition, a face image contains large informative content, especially around the key points of the face. Large information processing increases the computational complexity as well as the time complexity. In order to reduce both, there must be a compromise with the volume of information content processing. By retaining discriminative content, this type of compromise can be realized. A discriminative least-squares framework has been utilized effectively for image classification and feature selection [3]-[5]. The core idea is to apply slack variables to enhance the class margins under the abstract framework of least-squares regression. The idea of enlarging the class margin of the Discriminative Least-Squares Regression (DLSR) framework is satisfactory for face classification only when the distribution of training faces is in accordance with that of probe faces across different poses. When the distribution of training faces is different from that of the probe faces, the idea of minimizing the class margin of the DLSR framework works better.

In this paper, an LBP-based texture descriptor, Logically Concatenated Cross Binary Pattern (LC-CBP) [6], is used for extracting information content from a face image with a low computation complexity. In order to retain the most discriminative information content, Positive Dragging embedded in Linear Regression is utilized in the case of Procrustes transformation application to the training set and Negative Dragging embedded in Linear Regression is utilized when no Procrustes transformation is applied to the training set. The main contribution of this paper is the utilization of a simple descriptor, the LC-CBP, with a low

computational complexity and using it with the dragging approach of the discriminative least-squares framework. Further, a multimodal variant of the LC-CBP descriptor is proposed where texture values are used in conjunction with intensity values. This fusion improved the classification accuracy effectively. In Section II, work related to popular face descriptors and the elastic-net regularized linear regression framework is discussed. Section III comprises the description of the proposed framework utilizing the LC-CBP descriptor and the ε-dragging-based discriminative elastic-net regularized linear regression framework. Section IV presents the outcomes, along with their investigation. In the last section, the conclusion and future works are presented.

## II. RELATED WORK

Face images are organized as a two-dimensional (2D) pattern that can be analyzed in both the spatial and the frequency domains. 2D structures can be organized into different facial features, each having different discrimination abilities to recognize a person. The reliability of face classification depends heavily on the extraction of facial features. The primary challenge stems from the instances where pose variations are present, which may lead to the occlusion of key features. In addition, due to the demands for a short response time, there is a need for a short computational time for the extraction of features. The feature extraction process is closely entangled to the task of soft-discrimination decision integration defining the separation criterion between the probe face and the gallery faces intelligently. With Linear Discriminant Analysis (LDA), a small set of features with the most relevant information are utilized for classification. LDA explicitly tries to model the difference between the various subjects [7]. Fisher's Linear Discriminant, a representative class subspace analysis method [8], [9], distinguishes gallery face images of different classes by the high-dimensional feature space constructed from the intensity values of 2D gallery images. The most discriminative features are sought by maximizing the class separation and minimizing the interclass separation. In face classification, which is an inherent multiclass classification problem, the limitation of this method is the generation of linear decision boundaries between the different subjects. The linear decision boundaries cannot generate a reliable classification as the minor variations in view angle results in the nonlinear transformations which can't be represented accurately by linear decision boundaries. In the case of small face databases like the AT&T database [20], LDA suffers from a "small sample problem," which introduces the singularity issue of the covariance matrix. In order to overcome this issue, Dai and Yuen [10] proposed using regularization instead of optimizing the Fisher index. In the literature, regularization formulation performed with an $L_1$ or $L_0$ norm or a combination of both leads to sparse representation used abundantly in feature selection. Zhang, Dai, Xu, and Jordan [11] in their work proved the equivalence between regularized Fisher Discriminant Analysis and multivariate linear regression problems with mild restrictions on the regularization parameter. In order to utilize the residuals of ordinary least-squares regression in

the classification framework, Xiang, Nie, Meng, Pan, and Zhang [12] presented a framework of DLSR for enlarging the distance between different classes for multiclass classification and feature selection. They introduced the ε-dragging technique to enlarge the distance between classes. In their framework, they embedded class label information into the LSR formulation. The optimization model of the DLSR is defined as follows [12],

$$\min_{W,M} \left\| X^T \cdot W + e_n \cdot t^T - Y - B \bullet M \right\|_F^2 + \lambda \left\| W \right\|_F^2 \quad st. \quad M \geq 0 \quad (1)$$

where $c$ is the number of classes, $c_i \in \{1, 2, \ldots, c\}$, denoting the class label of the sample, and X is the set of "$n$" images, $X = [x_1, x_2, ., x_n]^T \in R^{n \times m}$. W represents the derived regression coefficients $[W = (XX^T)^\dagger. XY] \in R^{m \times c}$, and $t$ is a translation vector in $R^c$. B represents a constant matrix representing the dragging direction. $B_{ij} \in R^{n \times c}$ (ith row and jth column element) is defined as

$$B_{ij} = \begin{cases} +1 & if \ y_i = j \\ -1 & otherwise \end{cases} \quad (2)$$

$\lambda$ represents a positive regularization parameter and $e^n$ is a vector with all 1's = $[1, 1, \ldots, 1]^T \in R^n$. Y represents the label matrix for each class, with every column vector representing binary regression with the target value "+1" for the desired subject and target value "0" for the rest of the subjects; that is, if the $i$th sample image belongs to the $j$th class, then $Y_i$ is represented as $[0, 0, 1, 0, 0]^T$. $\bullet$ represents a Hadamard product operator of matrices and $\| \|_F$ represents the Frobenius norm of the matrix. In order to enhance the performance of DLSR in feature extraction and classification, Lu, Lai, Fan, Cui, and Zhu [13] in their work integrated the local geometrical structure into the data regularization terms. Manifold Discriminant Regression Learning computes the separation plane between various classes by computing an optimal subspace from the within-class graph and between-class graph. Both $L_1$ and $L_{21}$ norms are used as regularization terms to produce sparse projections for the extraction of feature sets and their classification. RMDRL utilizes the nuclear norm as a regularization term to learn a robust projection matrix.

Recently, Xie, Yang, Qian, Tai, and Zhang [14] and Yang, Luo, Qian, Tai, Zhang, and Xu [15] showed that the nuclear-norm-based matrix regression model (NMR) has great potential in dealing with structural noise. The objective function of the NMR is formulated as

$$\min_W \left\| X^T. W - Y \right\|_F^2 + \lambda \left\| W \right\|_* \quad (3)$$

where $\| W \|^*$ is the trace (or nuclear) norm of the regression coefficients (W). Tai *et al.* [16] proposed a 2D matrix-based error model by applying the nuclear-norm constraint on the error term to deal with face alignment and recognition. The proposed structural orthogonal Procrustes regression is utilized to handle pose variations existing in 2D face images. The objective function is optimized using an augmented Lagrangian multiplier. For face classification with mild pose variations, Peng *et al.* [4] used negative slack variables under a discriminative least-squares conceptual framework to

minimize the margin between the identity coupled pose-variant images. In order to deal with nonlinear variations of face images across pose variations, they utilized the "*kernel trick*" to transform the nonlinear face space to high-dimensional linear space. However, this becomes computationally challenging in the case of large image databases, like the Indian Movie Face Database (IMFDB) [17]. Formulating an efficient feature selection and representation model for multiclass classification problems is still a vital area to be further studied.

## III. PROPOSED WORK

While utilizing the framework of discriminative regression models, the primary challenge is to learn a model that can compactly utilize the extracted features in a discriminative manner. For classification purpose, the framework must be capable of generating a minimal residual error. In this paper, we propose utilizing LC-CBP with DLSR (Fig. 1). The goal of the proposed framework is to broaden the interclass distances as much as possible with positive ε-dragging and to reduce the interclass distance as much as possible with negative ε-dragging.

The proposed approach's framework is comprised of three sections:

(1) Face alignment using Procrustes analysis in the case when the pose variation across the yaw axis is more significant than 30°.
(2) Feature extraction using the LC-CBP descriptor.
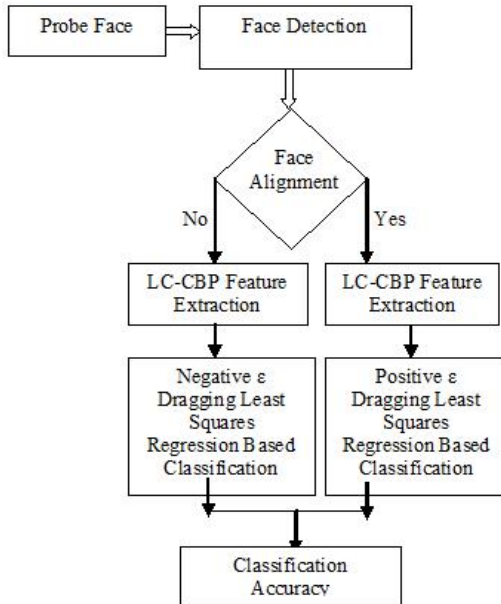(3) Face classification using the ε-dragging approach for discriminative linear regression.


Fig. 1. The proposed approach.

### A. Face Preprocessing

During this step, face detection is performed using the Viola–Jones algorithm [18]. The detected face is cropped from the input image, and then the cropped face image is used for the following steps.

### B. Alignment Using Procrustes Transformation

For reliable feature matching, there must be training and test features in the same feature space. Fig. 2 depicts the alignment of gallery face images with the poses to the probe image in order to have the corresponding key features of identity coupled different poses lie in the closer space.
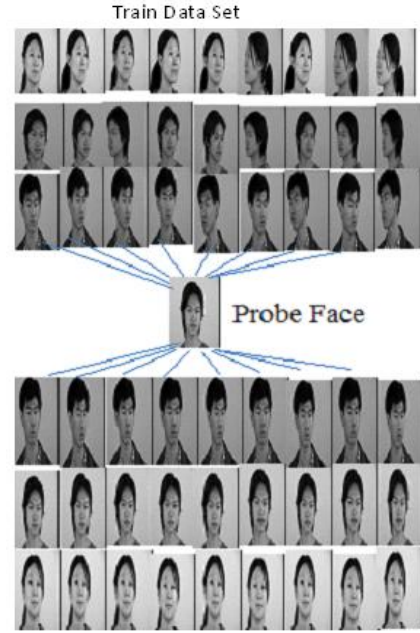

Fig. 2. Face alignment using procrustes transformation.

### C. LC-CBP Extraction

A pattern is produced by considering direct neighbors and extended neighbors of image cells of size 5 × 5 across the principal and non-principal diagonals of the cell [6]. The center pixel of the cell is encoded with the summation of the threshold function (*f*) applied to the difference of the center pixel intensity value and neighbor intensity value,

$$P_c= \sum_{i=0}^{7} f(P_i - P_c)\cdot 2^i,$$

$$\text{where } f(z)=\begin{cases}1, & \text{if} & z\geq 0 \\ 0, & \text{if} & z<0\end{cases} \qquad (4)$$

In order to extract maximum information from the diagonal pixels under consideration, their signed LBP is logically concatenated along with the magnitude pattern using AND/OR/XOR operators (Fig. 3).
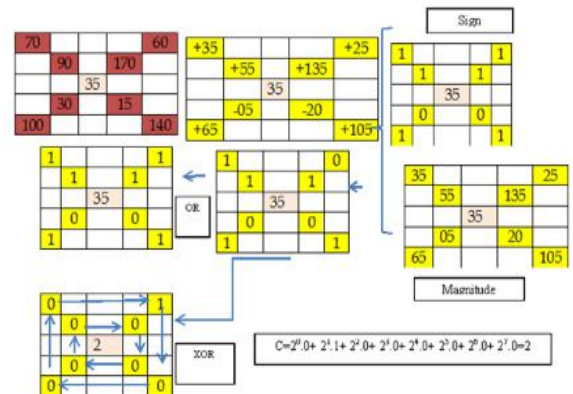

Fig. 3. Cross binary pattern generation by logical concatenation [6].

The magnitude and sign pattern of neighboring and extended diagonal pixels under consideration is described by

$$\text{XLBP} = \Sigma_{n=1}^{8} 2^{n-1} \cdot f(I_n - I_c) \qquad (5)$$

where $I_1$ is the top left pixel value of the immediate neighbor, and the anti-cyclic diagonal pixels are taken afterwards. Extended diagonal pixels are taken into consideration in the same fashion. The difference with the center pixel $I_c$ is encoded with a threshold function ($f$). Here the patterns for a sign and the magnitude patterns of the block are concatenated by a logical OR function; otherwise, other logical functions (AND, XOR, or NOT) can be used in the generation of the concatenated pattern.

This pattern facilitated reducing the feature size under consideration by 65% (approximately), inducing an increase in the computational speed.

### D. ε-Dragging Linear Regression

For classification purpose, there is usually a need to maximize the distance between the instances belonging to different classes. However, when the training samples' distribution is not in accord with the test samples' distribution instances, then minimization of interclass distances is sought. In order to broaden or to reduce the distance between classes, the DLSR model makes use of class label information as well as slack variables in the process of misclassification error approximation. The "$n$" instances of training samples is presented with tuple representation $\{x_i, y_i\}_{i=1}^{n}$, where the data point $X=\{x_i\} \in R^m$, and associated class label $Y=\{y_i\} \in R^c$. The binary regression targets for multiclass classification is represented by the $F_i$ function with only the $i$th element equal to one rest all zeros (e.g., $F_1 = [1,0,0\ldots0]^T$). The linear regression model aims to learn a function:

$$\min_{W,t} \left\| X^T W + e_n \cdot t^T - (Y + B \circ M) \right\|_F^2 + \lambda \|W\|_F^2 \qquad (6)$$

where $W \in R^{m \times c}$ is a transformation matrix that converts the training feature set $X$ into a class label binary matrix $Y$, $e_n$ is a vector with all 1's (i.e., $e_n = [1,1,1.1]^T$), $o$ is Hadamard's product operator, $\lambda$ is a positive regularization parameter, $t \in R^c$ is a translation vector, and $B$ is a dragging coefficient matrix. In case of negative dragging, the dragging coefficient matrix is defined by

$$B_{ij} = \begin{cases} -1 & if \quad Y_{ij} = 1 \\ 1 & if \quad Yij = 0 \end{cases} \qquad (7)$$

In case of positive dragging, the dragging coefficient matrix is defined by

$$B_{ij} = \begin{cases} 1 & if \quad Y_{ij} = 1 \\ -1 & if \quad Yij = 0 \end{cases} \qquad (8)$$

$M$ is the slack coefficient matrix defined as

$$\begin{bmatrix} \varepsilon_{11} & \varepsilon_{12} & \cdot & \cdot & \varepsilon_{1c} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \varepsilon_{n1} & \varepsilon_{n2} & \cdot & \cdot & \varepsilon_{nc} \end{bmatrix}, \varepsilon_{ij} \geq 0$$

In order to present the rationale for enlarging or minimizing the class margin, we consider an example of a $4 \times 4$ binary class label matrix $Y = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}$ and for negative dragging of slack variables is

$$M = \begin{bmatrix} -\varepsilon_{11} & \varepsilon_{12} & \varepsilon_{13} & \varepsilon_{14} \\ \varepsilon_{21} & -\varepsilon_{22} & \varepsilon_{23} & \varepsilon_{24} \\ \varepsilon_{31} & \varepsilon_{32} & -\varepsilon_{33} & \varepsilon_{34} \\ -\varepsilon_{41} & \varepsilon_{42} & \varepsilon_{43} & \varepsilon_{44} \end{bmatrix}, \quad s.t \quad \varepsilon_{ij \geq 0}$$

The modified label matrix is

$$\tilde{Y} = \begin{bmatrix} \mathbf{1}-\varepsilon_{11} & \varepsilon_{12} & \varepsilon_{13} & \varepsilon_{14} \\ \varepsilon_{21} & \mathbf{1}-\varepsilon_{22} & \varepsilon_{23} & \varepsilon_{24} \\ \varepsilon_{31} & \varepsilon_{31} & \mathbf{1}-\varepsilon_{33} & \varepsilon_{34} \\ \mathbf{1}-\varepsilon_{41} & \varepsilon_{42} & \varepsilon_{43} & \varepsilon_{44} \end{bmatrix}$$

The distance between the first two samples, represented by $Y$, is $\sqrt{2}$, whereas the distance between two samples ($D_{14}$) belonging to the same class for $\tilde{Y} = Y + B \bullet M$ with the "ε" constraint is

$$D_{14} = \sqrt{(\varepsilon_{11} - \varepsilon_{41})^2 + (\varepsilon_{42} - \varepsilon_{12})^2 + (\varepsilon_{43} - \varepsilon_{13})^2 + (\varepsilon_{44} - \varepsilon_{14})^2}$$

which is clearly $D_{14} \leq \sqrt{2}$ (i.e., the distance between two samples belonging to the same class is decreased).

For positive dragging,

$$M = \begin{bmatrix} \varepsilon_{11} & -\varepsilon_{12} & -\varepsilon_{13} & -\varepsilon_{14} \\ -\varepsilon_{21} & \varepsilon_{22} & -\varepsilon_{23} & -\varepsilon_{24} \\ -\varepsilon_{31} & -\varepsilon_{32} & \varepsilon_{33} & -\varepsilon_{34} \\ \varepsilon_{41} & -\varepsilon_{42} & -\varepsilon_{43} & -\varepsilon_{44} \end{bmatrix} \quad s.t \quad \varepsilon_{ij \geq 0}$$

$$\tilde{Y} = \begin{bmatrix} \mathbf{1}+\varepsilon_{11} & -\varepsilon_{12} & -\varepsilon_{13} & -\varepsilon_{14} \\ -\varepsilon_{21} & \mathbf{1}+\varepsilon_{22} & -\varepsilon_{23} & -\varepsilon_{24} \\ -\varepsilon_{31} & -\varepsilon_{31} & \mathbf{1}+\varepsilon_{33} & -\varepsilon_{34} \\ \mathbf{1}+\varepsilon_{41} & -\varepsilon_{42} & -\varepsilon_{43} & -\varepsilon_{44} \end{bmatrix}$$

$$D_{12} = \sqrt{(1-\varepsilon_{11}+\varepsilon_{21})^2 + (-\varepsilon_{12}-\varepsilon_{22}-1)^2 + (-\varepsilon_{13}+\varepsilon_{23})^2 + (-\varepsilon_{14}+\varepsilon_{24})^2}$$

Considering the "ε" constraint, clearly, $D_{12} \geq \sqrt{2}$ (i.e., the distance between two different samples is increased). In order to improve the flexibility of the ε-dragging approach, enforcing of a marginalized constraint makes the target space more distinct.

The objective function of a linear regression model with Marginalized Elastic-Net Regularization (MENR) and positive ε-dragging is framed as follows:

$$\arg\min_{W} \; f_W(X) = \min_{W,t} \left\| X^T W - \tilde{Y} \right\|_F^2 + \lambda_1 \|W\|_* + \lambda_2 \|W\|_F^2 \qquad (9)$$

Considering $\tilde{Y} = Y + B \bullet M$, the learned regression targets $\tilde{Y} = [\tilde{y}_1, \tilde{y}_2, \tilde{y}_3, \ldots \tilde{y}_n]^T \in R^{n \times c}$. The learned regression targets follow the constraint of the fixed marginal value "$K$" for setting a separation plane between true and false classes; that is,

$$\tilde{y}_{ip} - \max_{j \neq p} \tilde{y}_{ij} \geq K, \ i=1,2....n \qquad (10)$$

The $i$th sample from the $p$th class ($y_i = p$, target class) is expected to have a value larger than that of the rest of the learned targets by a fixed value "$K$." Similarly, the objective function of a linear regression model with MENR and negative ε-dragging is framed as follows:

$$\tilde{y}_{ip} - \min_{j \neq p} \tilde{y}_{ij} \leq K \qquad (11)$$

The $i$th sample from the $p$th class ($y_i = p$, target class) is expected to have a value smaller than that of the rest of the learned targets by a fixed value "$K$." $W$ can be expressed as $U\Sigma V^T$ or $\sum_{i=1}^{R} u_i \sigma_i v_i^T$, where $R$ is the rank of $W$. Therefore, elastic-net regularization terms can be expressed via singular values where the fundamental information resides. The term $\lambda_1 \|W\|_* + \lambda_2 \|W\|_F^2$ can be expressed as:

$$\sum_{i=1}^{R} |\sigma_i| + \sum_{i=1}^{R} \|\sigma_i\|^2 \qquad (12)$$

The use of singular values indirectly enables the selection of principal components and continuous shrinkage of redundant components [17]. Considering $W = U\Sigma V^T$,

$$\|W\|_* = \text{tr}(\Sigma) = \frac{1}{2}\text{tr}(\Sigma) + \frac{1}{2}\text{tr}(\Sigma)$$
$$= \frac{1}{2}\text{tr}(U\sqrt{\Sigma}\sqrt{\Sigma}^T U^T) + \frac{1}{2}\text{tr}(\sqrt{\Sigma} V^T V\sqrt{\Sigma}^T) \qquad (13)$$
$$= \frac{1}{2}\|W_1\|_F^2 + \frac{1}{2}\|W_2\|_F^2$$

where $W_1 = U\sqrt{\Sigma}$ and $W_2 = \sqrt{\Sigma}V^T$.

The reformulated objective function to be minimized can now be defined as

$$\arg \min_{W} f_W(X) =$$

$$\min_{W,t} \left\|X^T W - \tilde{Y}\right\|_F^2 + \frac{\lambda_1}{2}\left(\|W_1^2\|_F^2 + \|W_2^2\|_F^2\right) + \lambda_2 \|W\|_F^2 \ \text{s.t.}$$
$$W = W_1.W_2 \qquad (14)$$

During the training phase, the learned transformation matrix (W) is learned from the $\{x_i, y_i\}_{i=1}^{n}$ data points. During the testing phase, the probe face is estimated by simple product ($W^T.x_p$).

### E. Augmented Lagrange Multiplier (ALM) Strategy to Optimize MENR

Here, the ALM approach is utilized to resolve the optimization function f (14) by alternatively maximizing the dual of the original problem and minimizing the augmented Lagrangian of the original problem. The augmented Lagrangian function of problem (14)

$$\mathcal{L}(R, W, W_1, W_2, C_1) =$$
$$\left\|X^T W - \tilde{Y}\right\|_F^2 + \frac{\lambda_1}{2}\left(\|W_1^2\|_F^2 + \|W_2^2\|_F^2\right) + \lambda_2 \|W\|_F^2$$
$$+ <C_1, W - W_1 W_2> + \frac{\mu}{2}\|W - W_1 W_2\| \qquad (15)$$

where $C_1$ is a Lagrange multiplier, $\mu$ is a decorrelation penalty parameter, and $<a,b> = \text{tr}(a^T b)$.

The block coordinate descent scheme is utilized for finding the minimum points of $\mathcal{L}$ with respect to primal variables.

The following are the steps to minimize the augmented Lagrangian by updating one coordinate at each direction.

(a) Update $W_1$ by fixing other variables by solving

$$W_1^+ = \arg \min_{W_1} \ \frac{\lambda_1}{2}\left(\|W_1\|_F^2\right) + \frac{\mu}{2}\left\|W - W_1 \cdot W_2 + \frac{C_1}{\mu}\right\|_F^2$$

Being a least-squares problem with regularization, the solution is

$$W_1^+ = (C_1 + \mu \cdot W) \cdot W_2^T \cdot (\lambda_1 I + \mu \cdot W_2 \cdot W_2^T)^{-1} \qquad (16)$$

(b) Update $W_2$ by fixing other variables by solving

$$W_2^+ = \arg \min_{W_2} \ \frac{\lambda_1}{2}\left(\|W_2\|_F^2\right) + \frac{\mu}{2}\left\|W - W_1 \cdot W_2 + \frac{C_1}{\mu}\right\|_F^2 \qquad (17)$$

Being a least-squares problem with regularization, the solution is

$$W_2^+ = (\lambda_1 I + \mu \cdot W_1 \cdot W_1^T)^{-1} \cdot W_1^T \cdot (C_1 + \mu \cdot W). \qquad (18)$$

(c) Update $W$,

$$W^+ = \arg \min_{W} \left\|X^T W - \tilde{Y}\right\|_F^2$$
$$+ \frac{\lambda_2}{2}\left(\|W\|_F^2\right) + \frac{\mu}{2}\left\|W - W_1 \cdot W_2 + \frac{C_1}{\mu}\right\|_F^2 \qquad (19)$$

The optimal solution for $W$ can be derived by setting $\frac{\partial L}{\partial W} = 0$

$$W^+ = (2 \cdot X.X^T + (\lambda_2 + \mu) \cdot I)^{-1} \cdot (2X.\tilde{Y} + \mu W_1 W_2 - C_1) \qquad (20)$$

(d) Update "R" regression targets row-wise using the optimal solution derived in [5]. For the feature set of the $i$th probe image, the optimal target representing its actual class ($m$th) must satisfy the constraint for positive ε-dragging and negative ε-dragging, respectively:

$$\phi_j = r_{ij} + 1 - r_{im}, \phi \leq 0 \quad \forall j \neq m \qquad (21)$$

$$\phi_j = r_{ij} + 1 - r_{im}, \phi \geq 0 \quad \forall j \neq m \qquad (22)$$

An optimal solution for $R$ can be derived from the rows of the regression target of training data computed using $P = X^T.W^+ + e_n.t^T$ and its associated class labels.

For each row of P, compute:

$$\varphi_r = P_r + 1 - \text{replicate}(P_r(\text{class\_label}), 1, \text{num\_classes}) \qquad (23)$$

For each entry of class_id

For each false class_id

Compute the learning factor ξ by including all the entries of a row for false class_id's

$$\xi = \frac{\sum_{j \neq trueclass\_id} \varphi_j . \Pi((\varphi_r - \varphi_j) > 0)}{1 + steps \quad \forall((\varphi_r - \varphi_j) > 0)} \qquad (24)$$

End for

End for

Update

$$P_r = \begin{cases} P_r + \xi & \text{if } (class\_id)_r = true \ class\_id \\ P_r + \min(\xi - \varphi_r), & \text{if } (class\_id)_r \neq true \quad class\_id \end{cases} \quad (25)$$

For negative ε-dragging, the learning factor can be expressed as

$$\xi = \frac{\sum\limits_{j \neq trueclass\_id} \varphi_j . \Pi((\varphi_r - \varphi_j) < \mathbf{0})}{1 + steps \quad \forall((\varphi_r - \varphi_j) < \mathbf{0})} \quad (26)$$

and $P_r$ is updated as

$$P_r = \begin{cases} P_r + \xi & \text{if } (class\_id)_r = true \ class\_id \\ P_r - \max(\xi - \varphi_r), & \text{if } (class\_id)_r \neq true \quad class\_id \end{cases} \quad (27)$$

Repeat steps (a), (b), (c), and (d) until there is no convergence.

(e) Update Lagrange's multiplier $C_1$ by

$$C_1 = C_1 + \mu \cdot (W - W_1 \cdot W_2). \quad (28)$$

(f) Repeat steps (d) and (e) until $\left\| W^{k+1} - W^k \right\| \leq 10^{-4}$, where the projection matrix W forms the optimized output.

The calculated regression targets are classified using a simple nearest neighbor classifier.

### F. Complexity Analysis of the Proposed Approach

With LC-CBP descriptor computation, the number of operations required to encode the binary pattern is $O((k-1)*(k-1))$, with cell size $k \times k$ and $k>2$. If the image contains $T$ $(k \times k)$ cells, then its utilization as a descriptor will help in the reduction of the number of operations per cell by $\frac{(k-1)*(k-1)}{k^2} \approx$ approximately $(1/k^2)$ operations. Therefore, for the runtime complexity $O(2 \ mnc + 2 \ mcr + m)$ of the MENR with m-dimensionality, $n$ instances belonging to c classes will have now fewer operations, approximately fewer by $(1/k^2)$ operations, for every $T$ cell making up the m dimension of the sample image.

### IV. EXPERIMENTAL RESULTS

This section presents the results of the experiments that were conducted to evaluate the significance of the sign of ε-dragging coefficients in classification based on DLSR. Experiments were performed with MATLAB 2017 on a Core i5 2.3 GHz processor and 8 GB of DDR4 RAM. Table I describes the data sets utilized for validate the efficiency of proposed approach

For each experiment, there was random selection of input images for the training set. The probe sample was also selected randomly from the rest of the set. 10 runs were conducted for each experiment. Accounting for the random nature of the training set, the standard deviation is associated with the classification accuracy. The best value is considered as the accuracy value.

Utilizing the regression framework for the classification transforms to learning for predicting effectively the label of a query sample from the given training set and an estimator

function "f." This translates into finding a coefficient vector $\beta$ such that the residual error is minimized and the correlation with the actual class is maximized.

TABLE I: USED DATABASES

| Database used | Description |
|---|---|
| Subset of the IMFDB [17] | The subset of this database consists of 2,000 faces of 10 known female and male actors (200 faces per subject), collected from different Indian movies. *The whole database consists of 34,512 faces of 100 known actors collected from approximately 103 Indian movies. |
| UMIST [19] | This database is now known as the Sheffield Face Database. It consists of 564 images belonging to 20 subjects. Each subject has a range of poses varying from profile to frontal views, in separate directories labeled 1a, 1b, …, 1t. |
| AT&T [20] | This database contains 10 images of 40 subjects with minor lighting, facial expressions (open/closed eyes, smiling/not smiling), and yaw pose variations up to 30° of facial details (glasses/no glasses). All images have a dark uniform background. |



Fig. 4. LC-CBP images with XOR, AND, and OR operators.

TABLE II: CLASSIFICATION ACCURACY WITH VARYING $\lambda_2$.

| $\lambda_2$ | AT&T | | UMIST | |
|---|---|---|---|---|
| | MENRLR-LCBP+ | MENRLR-LCBP− | MENRLR-LCBP+ | MENRLR-LCBP− |
| .00001 | 86.67 ± 0.90 | 85.04 ± 0.87 | 81 ± 0.597 | 80.78 ± 0.85 |
| .0001 | 91.001 ± 1.50 | 89.87 ± 1.09 | 86 ± 0.90 | 86.56 ± 0.79 |
| **.001** | **94.64 ± 1.343** | **93.4 ± 0.890** | **90.04 ± 3.16** | **90.6 ± 1.12** |
| .01 | 87.664 ± 1.51 | 85.98 ± 1.90 | 87 ± 2.01 | 88.02 ± 1.67 |
| .1 | 83.67 ± 0.97 | 80 ± 1.34 | 80.034 ± 1.59 | 80.92 ± 0.98 |
| 1 | 81.89 ± 2.18 | 79 ± 1.83 | 81.67 ± 2.678 | 83 ± 1.35 |
| 10 | 78 ± 2.35 | 75.7 ± 1.890 | 75.45 ± 3.21 | 77.1 ± 1.29 |

In case of the preprocessing step, Procrustes transformation analysis is applied to align the images subject/class-wise as per the orientation of the probe sample. Fig. 4 highlights the effect of applying different logical operators (AND, OR, and XOR) amid the magnitudes and sign of the binary pattern. Discriminative information like silhouettes of key features is retained with different logical

operators of the LC-CBP descriptor. The conducted experiments stressed that, among the OR/XOR/AND logical operators, XOR fares better for the poses with deviation of the face more than 20° along the yaw axis.

The appropriate values of $\lambda_2$ and $\lambda_1$ are studied by observing the impact of variation in their values on the classification accuracy. The value of $\mu$ is taken as 0.25. Tables II and III demonstrate the enactment of $\lambda$ with different ranges of values on standard databases (AT&T, UMIST, and IMFDB).

TABLE III: Classification Accuracy with Varying $\lambda_1$.

| $\lambda_1$ | AT&T | | UMIST | |
| | MENRLR-LCBP$^+$ | MENRLR-LCBP$^-$ | MENRLR-LCBP$^+$ | MENRLR-LCBP$^-$ |
| --- | --- | --- | --- | --- |
| .00001 | 82.67 ± 2.02 | 82 ± 1.67 | 78 ± 1.497 | 78.97 ± 0.90 |
| .0001 | 89.91 ± 2.15 | 88.9 ± 1.38 | 84.89 ± 0.90 | 85.2 ± 1.02 |
| **.001** | **94.604 ± 1.343** | **93.89 ± 1.67** | **90.04 ± 3.16** | **91.89 ± 1.890** |
| .01 | 90.76 ± 2.51 | 89.3 ± 1.567 | 86.89 ± .976 | 88.02 ± 1.09 |
| .1 | 83.67 ± 0.97 | 82.78 ± 1.12 | 80.034 ± 1.77 | 83 ± 0.690 |
| 1 | 80.966 ± 0.21 | 78.89 ± 0.790 | 80.34 ± 0.445 | 81 ± .987 |
| 10 | 78 ± 2.35 | 76.6 ± 1.96 | 73.45 ± 3.25 | 73.6 ± 3.09 |

From the above tables (II and III), it was found that the best classification accuracy can be derived with $\lambda_1$ = .001 and $\lambda_2$ = .001. Tables IV, V, and VI present the comparison of the classification accuracy with regression-based classification approaches.

TABLE IV: Performance Comparison for the AT&T Database

| Method | Training/Testing splits of the database | | |
| | 40%/60% | 50%/50% | 60%/40% |
| --- | --- | --- | --- |
| Linear Regression Classification | 76.73 ± 1.07 | 78.5 ± 1.03 | 79.2 ± 1.05 |
| Ridge Regression | 82.02 ± 2.380 | 85.09 ± 2.35 | 89.78 ± 2.32 |
| Low-Rank Regression | 89.092 ± 4.65 | 92.31 ± 4.08 | 92.89 ± 4.55 |
| Discriminative Linear Regression | 87.431 ± 2.45 | 90.056 ± 1.991 | 90.094 ± 1.999 |
| Discriminative Elastic-Net Linear Regression | 88.62 ± 3.032 | 90.93 ± 2.78 | 92.1 ± 2.656 |
| MENRLR-LCBP$^+$ | **92.78 ± 1.171** | **94.664 ± 1.343** | **95.832 ± 1.210** |
| MENRLR-LCBP$^-$ | **92.48 ± 1.987** | **93.824 ± 2.11** | **94.008 ± 2.034** |

From the above comparisons, it was found that the performance of MENRLR is better than of the other approaches. Marginalized elastic-net regression with positive dragging (MENRLR-LCBP$^+$) fares better than marginalized elastic-net regression with negative dragging (MENRLR-LCBP$^-$) for the AT&T database. The underlying reason for the better performance with positive dragging can be attributed to the fact that the AT&T database has images with minor or no pose variations. Therefore, maximizing the separation planes between various classes of the database favors good classification accuracy.

TABLE V: Performance Comparison for the UMIST Database

| Method | Training/testing splits of the database | | |
| | 40%/60% | 50%/50% | 60%/40% |
| --- | --- | --- | --- |
| Linear Regression Classification | 71.35 ± 4.74 | 72.59 ± 4.75 | 74.0 ± 4.0 |
| Ridge Regression | 75.9 ± 2.3 | 77 ± 2.49 | 77.82 ± 2.38 |
| Low-Rank Regression | 78.05 ± .849 | 79.12 ± .956 | 81.45 ± .621 |
| Discriminative Linear Regression | 82.10 ± 2.76 | 83.31 ± 2.58 | 84.89 ± 3.02 |
| Discriminative Elastic-Net Linear Regression | 83.78 ± 2.1 | 84.97 ± 3.23 | 86.11 ± 2.78 |
| MENLR-LCBP+ | 87.62 ± 1.29 | 90.09 ± 1.28 | 91.21 ± .9656 |
| MENLR-LCBP− | 89.03 ± 1.043 | 91.89 ± 1.89 | 93.924 ± 1.987 |

From the above comparisons, it was found that the performance of MENRLR is better than of the other approaches. MENRLR-LCBP$^-$ fares better than MENRLR-LCBP$^+$ for the UMIST database. The underlying reason for the better performance with negative dragging can be attributed to the fact that UMIST has images with pose variations. Procrustes transformation application as per probe increases both the distance between the interclass images and the misclassification error.

TABLE VI: Performance Comparison for the IMFDB Database

| Method | Training/testing splits of the database | | |
| | 40%/60% | 50%/50% | 60%/40% |
| --- | --- | --- | --- |
| Linear Regression Classification | 28.42 ± 1.32 | 30.2 ± .83 | 31.79 ± 1.05 |
| Ridge Regression | 28.86 ± .880 | 31.5 ± .35 | 32.78 ± 1.032 |
| Low-Rank Regression | 30.06 ± .749 | 32.12 ± 1.01 | 32.95 ± 1.621 |
| Discriminative Linear Regression | 32.2 ± 1.84 | 34.31 ± 4.08 | 35.53 ± 1.79 |
| Discriminative Elastic-Net Linear Regression | 33.87 ± 0.45 | 36.956 ± 2.901 | 38.094 ± 1.675 |
| MENLR-LCBP$^+$ | **35.88 ± 2.032** | **37.93 ± 2.78** | **39.21 ± 3.15** |
| MENLR-LCBP$^-$ | **36.7 ± 2.034** | **38.89 ± 3.987** | **40.024 ± 2.11** |

From the above comparisons, it was found that the performance of MENRLR is better than of the other approaches. MENRLR-LCBP$^-$ fares better than MENRLR-LCBP$^+$ for the IMFDB. The underlying reason for the better performance with negative dragging can be attributed to the fact that IMFDB has images with huge pose variations. Therefore, minimizing the distance between the interclass images with pose variations favors clustering of pose-variant images of a class, and the separation planes between various classes of the database increase, hence favoring better classification accuracy. However, with positive dragging, along with the increase in the distance of the separation plane between classes, there is also an increase in the misclassification error.

## V. CONCLUSION

The experimental results obtained above clearly demonstrate that the sign of the ε-dragging coefficients has a

significant impact on the classification of pose-variant images. A DLSR model with MENR and LC-CBP is applied for efficient classification of compact characteristics of face images across pose variations. For frontal or near-frontal face images, it was found that positively signed ε-dragging coefficients along with Procrustes transformation, fare better in increasing the distance between the images of different classes. For half profile pose to nearly profile poses, negatively signed ε-dragging coefficients fare better, as they help indirectly in clustering identity coupled images thereby boosting the classification accuracy.

## CONFLICT OF INTEREST

The authors declare there is no conflict of interest with respect to research, authorship, and/or publication of this article.

## AUTHOR CONTRIBUTIONS

First author conceived the presented idea in this article, performed the computations and analyzed the results. Second author supervised the findings of this work. Both authors discussed the results and contributed to the final manuscript. Both authors had approved the final version.

## REFERENCES

[1] T. Ojala, M. Pietikäinen, and D. Harwood, "A comparative study of texture measures with classification based on featured distributions," *Pattern Recognition*, vol. 29, no. 1, pp. 51-59, 1996.

[2] L. Liu, P. Fieguth, Y. Guo, X. Wang, and M. Pietikäinen, "Local binary features for texture classification: Taxonomy and experimental study," *Pattern Recognition*, vol. 62, pp. 135-160, 2017.

[3] S. Xiang, F. Nie, G. Meng, C. Pan, and C. Zhang, "Discriminative least squares regression for multiclass classification and feature selection," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 23, no. 11, pp.1738-1754, 2012.

[4] Y. Peng, L. Zhang, S. Liu, X. Wang, and M. Guo, "Kernel negative ε dragging linear regression for pattern classification," *Complexity*, 2017.

[5] Z. Zhang, Z. Lai, Y. Xu, L. Shao, J. Wu, and G. S. Xie, "Discriminative elastic-net regularized linear regression," *IEEE Transactions on Image Processing*, vol. 26, no. 3, pp. 1466-1481, 2017.

[6] K. Arora, and P. Garg, "Two-dimensional canonical correlation analysis of the logically concatenated cross binary pattern for cross pose face recognition," *International Journal of Computer Vision and Image Processing (IJCVIP)*, vol. 8, no. 3, pp. 66-79, 2018.

[7] K.Etemad, and R. Chellappa, "Discriminant analysis for recognition of human face images," *Josa a*, vol. 14, no. 8, pp. 1724-1733, 1997.

[8] R. Rahim, T. Afriliansyah, H. Winata, D. Nofriansyah, and S. Aryza, "Research of face recognition with fisher linear discriminant," in *Proc. IOP Conference Series: Materials Science and Engineering*, IOP Publishing, January 2018, vol. 300, no. 1, p. 012037.

[9] P. N. Belhumeur, J. Hespanha, and D. J. Kriegman, "Eigenfaces vs. Fisherfaces: recognition using class specific linear projection," *PAMI*, vol. 19, pp. 711-720, 1997.

[10] D. Q. Dai and P.C. Yuen, "Face recognition by regularized discriminant analysis," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 37, no. 4, pp. 1080-1085, 2007.

[11] Z. Zhang, G. Dai, C. Xu, and M. I. Jordan, "Regularized discriminant analysis, ridge regression and beyond," *Journal of Machine Learning Research*, vol. 11, pp. 2199-2228, August 2010.

[12] S. Xiang, F. Nie, G. Meng, C. Pan, and C. Zhang, "Discriminative least squares regression for multiclass classification and feature selection," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 23, no. 11, pp. 1738-1754, 2012.

[13] Y. Lu, Z. Lai, Z. Fan, J. Cui, and Q. Zhu, "Manifold discriminant regression learning for image classification," *Neurocomputing*, vol. 166, pp. 475-486, 2015.

[14] J. Xie, J. Yang, J. J. Qian, Y. Tai, and H. M. Zhang, "Robust nuclear norm-based matrix regression with applications to robust face recognition," *IEEE Transactions on Image Processing*, vol. 26, no. 5, pp. 2286-2295, 2017.

[15] J. Yang,, L. Luo, J. Qian, Y. Tai, F. Zhang, and Y. Xu, "Nuclear norm based matrix regression with applications to face recognition with occlusion and illumination changes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 1, pp. 156-171, 2017.

[16] Y. Tai, J. Yang, F. Zhang, Y. Zhang, L. Luo, and J. Qian, "Structural orthogonal procrustes regression for face recognition with pose variations and misalignment," in *Proc. the 2016 SIAM International Conference on Data Mining*, June 2016, pp. 693-701.

[17] S. Setty, M. Husain, P. Beham, J. Gudavalli, M. Kandasamy, R. Vaddi, V. Hemadri, J. C. Karure, R. Raju and V. Kumar, "Indian movie face database: a benchmark for face recognition under wide variations," in *Proc. Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG)*, December 2013, pp. 1-5.

[18] P. Viola and M. J. Jones, "Robust real-time face detection," *International Journal of Computer Vision,* vol. 57, no. 2, pp. 137-154, 2004.

[19] UMIST Face Database [Online]. Available: https://www.sheffield.ac.uk/eee/research/face

[20] AT&T- The Database of Faces (formerly "The ORL Database of Faces") retrieved from https://www.cl.cam.uk/research/dtg/attarchive/facedatabase.html

**Kumud Arora** is an assistant professor at Inderprastha Engineering College, Ghaziabad, India, since 2007. She is also a part-time research scholar at Banasthali Vidyapith, Banasthali, India. She received her M.Tech degree in computer science from IETE, New Delhi, in 2014; her M.Phil degree in computer science from Alagappa University in 2004, and her masters degree in computer application with distinction from guru nanak dev university, Amritsar, Punjab, India, in 1999. She has published research papers in refereed international journals and conferences. Her area of research includes digital image processing, machine learning, and pattern recognition.

**Poonam Garg** is a professor at the Institute of Management Technology, Ghaziabad, India. She obtained her Ph.D. degree in computer science in 2006 from Banasthali Vidyapith, Banasthali, India. She has more than 25 years of experience in teaching, research, and consulting. Her current research interests are in the areas of enterprise resource planning, social media analytics, developing heuristics and metaheuristics, particularly genetic algorithms, Tabu search, and simulated annealing-based metaheuristics for various optimization problems, such as cryptanalysis of various encryption algorithms and project management. She has published many papers in different international/national journals and has 10 edited books. She attended many international conferences as a conference co-chair, conference track chair, and member of program committee. She is also serving as an external expert with the doctoral research committee for Ph.D. programs at the IIS University and MLS University, Indore, India.