An Effective Combination of Textures and Wavelet Features for Facial Expression Recognition

Syed Muhammad Hassan
Department of AI and Mathematical Sciences
SMI University
Karachi, Pakistan
m.hassan@smiu.edu.pk

Abdullah Alghamdi
College of Computer Science and Information Systems
Najran University
Najran, Saudi Arabia
abdulresearch@hotmail.com

Abdul Hafeez
Department of Software Engineering
SMI University
Karachi, Pakistan
ahkhan@smiu.edu.pk

Mohammad Hamdi
College of Computer Science and Information Systems
Najran University
Najran, Saudi Arabia
mahamdi@nu.edu.sa

Imtiaz Hussain
Department of AI and Mathematical Sciences
SMI University
Karachi, Pakistan
imtiaz@smiu.edu.pk

Mesfer Alrizq
College of Computer Science and Information Systems
Najran University
Najran, Saudi Arabia
msalrizq@nu.edu.sa

***Abstract*—In order to explore the accompanying examination goals for facial expression recognition, a proper combination of classification and adequate feature extraction is necessary. If inadequate features are used, even the best classifier could fail to achieve accurate recognition. In this paper, a new fusion technique for human facial expression recognition is used to accurately recognize human facial expressions. A combination of Discrete Wavelet Features (DWT), Local Binary Pattern (LBP), and Histogram of Gradients (HoG) feature extraction techniques was used to investigate six human emotions. K-Nearest Neighbors (KNN), Decision Tree (DT), Multi-Layer Perceptron (MLP), and Random Forest (RF) were chosen for classification. These algorithms were implemented and tested on the Static Facial Expression in Wild (SWEW) dataset which consists of facial expressions of high accuracy. The proposed algorithm exhibited 87% accuracy which is higher than the accuracy of the individual algorithms.**

Keywords-ANN; FER; DWT; LBP; HOG; K-Nearest Neighbors

I. INTRODUCTION

Facial expressions are a way of sentiment expression and non-verbal correspondence. There are various systems that deal with human attitude and point of view recognition. Facial Expression Recognition (FER) transforms is one of the most discussed scientific areas nowadays. This issue is furthermore incredibly noteworthy in Human-Computer Collaboration (HCI) [1, 2]. FER is being utilized to provide a description for the mental state of human beings [3]. Meanwhile, modifications in the look of photos can occur by disturbances in the pixels. Illumination troubles might also occur in indoor or outdoor photos. The exploration indicates those issues and proposes a combination strategy for different accessible highlights that surpasses these issues [4].

II. LITERATURE SURVEY AND THEORETICAL FRAMEWORK

It is very difficult to identify human facial regions. In order to handle this efficiently, a technique should be implemented to recognize facial indicators. One clause that is vital to know is the dynamic angle on transferring video [5]. Lower and top face method extends the spatial pyramid histogram of edges which give 3-dimensional facial acknowledgment. Fundamentally in this method, elements are researched for cheerful and pity indicators [6]. LBP and Improved Local Binary Pattern have been applied alongside Coordinate Bunching Representation [7]. Face recognition using an optimized algorithm chain for both 2D and 3D images gives an accuracy about 96% with SVM classifier using LBP and PCA. Further testing on 2D and 3D images using LBP and PCA with FFBPNN (Feed Forward Back Propagation Neural Network) is less effective and efficient as compared to the SVM classifier [8]. Locality Preserving Projections (LPPs) have been used for manifold systems originated from Local Binary Pattern (LBP) subjects [9]. At first, a pyramid change is utilized to divide the test photographs into different areas. So, the goal pictures are isolated. After this, the ELBP is applied upon the little pictures to compute the ELBP pyramid and the community photo decided qualities are utilized to the little pictures from AWM which can ascertain the importance of the facts they got.

Corresponding author: Syed Muhammad Hassan

Finally, the AWELBPP highlight is assembled from the blend of the little ELBP pyramid and the AWM [10]. Support Vector Machine (SVM) has been applied in dispensing boisterous pictures for highlight extraction [11, 12].

Background Subtraction [13] showed good results by applying background subtraction on real-time feeds. In this work, a model based on Gaussian Mixture was used for unfolding the pixels of images and the variables of the pattern were calculated with the Expectation-Maximization (EM) algorithm. The shades were also spotted effectively. Background subtraction was also very effective and met the requirements of drowning detection. Authors in [14] reviewed earlier approaches and tried to cover up the issue of recognizing actions and behaviors and the problem of dealing with a moderate crowded situation with a good modeling technique. The conventional techniques where mixed and a Gaussian distribution was used to design the temporal changing of the background pixels in [15]. This has been proven to be insufficient for extremely non-stationary environment. However, the thresholding method with hysteresis dealt with the issue of choosing thresholds in the background subtraction context. Stationary cameras have also been used to find drowning persons in swimming pools [6, 17, 18]. In contrast to previous works based on geometrical and 3D Mahalanobis distance features, the presented method in [18] captured the temporal and spatial correlation of the swimmers along with color information using the Markov Random Field (MRF) context to give better performance. Promising outcomes for drowning detection were achieved using an exclusive functional link net which fused the descriptors of extracted swimmers optimally. An improved descriptor fusion technique associated with the hierarchical technique was proposed in [19]. The current drowning detection techniques can be broadly classified into the vision-based schemes and the systems based on wearable sensors [20-22]. On the other hand, the combination of aerial and underwater cameras to monitor the postures of FER was utilized in [23], whereas the CNN model achieved 99.78% accuracy [24]. An even more successful accuracy level was achieved in learning similarities and dissimilarities among the faces of dataset using FDREnet in [25].

III. SYSTEM METHODOLOGY

Usually, cameras are present in most areas for security purposes. The already installed cameras can be utilized for the purpose of monitoring and expression detection. So, few critical frames are extracted from the video or can be utilized. A facial expression video is divided into frames to be processed. The image frames extracted from the video are utilized for feature extraction. Then, classification is carried out.

A. Feature Extraction

The input dataset is very large to be handled and processed. It is supposed to be redundant (enough data, but not abundant information), so, the input dataset will be converted into a reduced depiction set containing features. This set is named as Features vector (Fv). This process is known as feature extraction. Therefore, taking out the prejudiced features from the images enhances the decline of the dimension of the Fv by removing the redundancy in images and squeezing the relevant data into the Fv to a much smaller size.

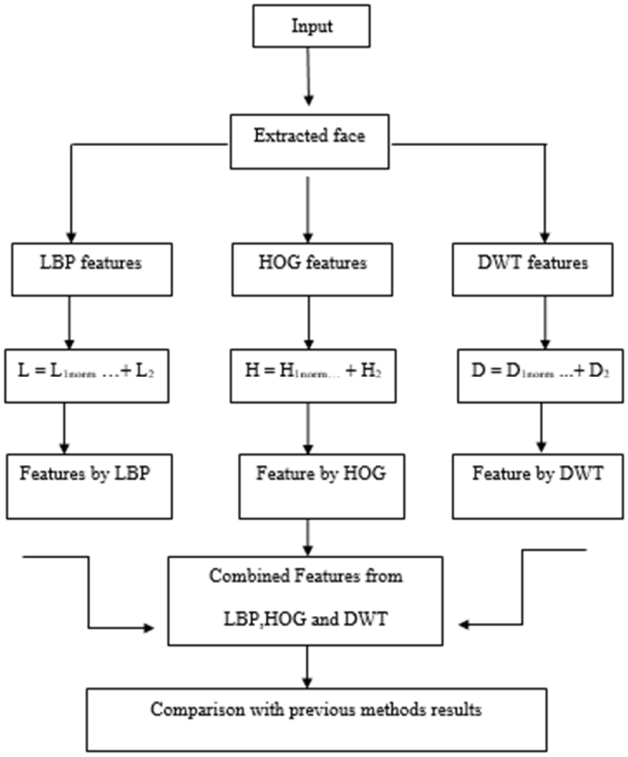

Fig. 1. The flowchart of the proposed method.

B. Feature Extraction via Discrete Wavelet Transform

Discrete Wavelet Transform (DWT) is utilized to extract features from an image on various levels of low pass (g), and high pass filtering (h). A signal x is calculated by passing through these series of filters, at first through the low pass filter (g) and then through the high pass (h):

$$h[n] = (x * g)[n] = \sum_{k=-\infty}^{\infty} x[k]g[n-k] \quad (1)$$

The low pass filter (g) and the high pass filter (h) with h cut-off frequency are described by:

$$y_{low}[n] = \sum_{k=-\infty}^{\infty} x[k]g[2n-k] \quad (2)$$

$$y_{high}[n] = \sum_{k=-\infty}^{\infty} x[k]h[2n-k] \quad (3)$$

The wavelet coefficients are the consecutive persistence of the estimate and the detail coefficients. The basic process of features extraction consists of:

- Mouldering the image using DWT in N-levels using decimation and filtering to get the detailed coefficients and approximation.
- Feature extraction using the DWT coefficients output.
- The features that were taken out from the DWT coefficients of the images are considered as the input to classifiers because of their operative representation.

The algorithmic steps for feature extraction from the dataset are:

- Step 1: The image data are decomposed into 4 detailed sub-bands by DWT.
- Step 2: The coefficients of approximation are further been decomposed by DWT to obtain localized data from the sub-band of the detailed coefficients of approximation (horizontal, vertical, and diagonal).
- Step 3: Aimed at processing and analyzing, all of the 4 levels detailed coefficients are calculated.
- Step 4: Finally, the features are analyzed and tabulated to be used as the input of the classifier.

C. Feature Extraction via Histogram of Gradients

The Histogram of Oriented Gradients (HOG) is the shape of the "function descriptor". The motive behind a feature descriptor is to generalize the item on a way that the same item (in this case a person) produces the same feature descriptor at the same time as considered under specific situations. This makes the class assignment simpler. Static Facial Expressions in the Wild (SFEW) has been utilized for selecting frames from AFEW. Regarding the block normalization for HOG, we consider $\boldsymbol{v}$ as the non-normalized vector containing all histograms in a given block, $\|v\|_k$ be its k-norm for k =1, 2, and e is some small constant. The normalized factor is defined as:

$$f = \frac{v}{\sqrt{\|v\|_2^2 + e^2}} \quad (4)$$

The dataset covers unconstrained facial expressions, numerous head poses, massive age variety, occlusions, numerous poses, and near actual global illumination. Frames had been extracted from AFEW sequences and were labeled based on the label of the series. Typically, SFEW includes seven-hundred snapshots which have been classified for six fundamental expressions: anger, disgust, fear, happiness, sadness, and surprise, and were categorized by unbiased labelers.

D. Feature Extraction via Local Binary Pattern (LBP)

The LBP method is applied on facial images in order to extract features that may be used to get a degree of similarity. Firstly, the pictures have been divided into several blocks. After that, the LBP histogram was calculated for each block. The value of the LBP code of a pixel (x_c, y_c) is considered as:

$$LBP_{(p,r)} = \sum_{p=0}^{p-1} s(g_p, g_c) 2^p \quad (5)$$

where $s(x) = \{1, \forall x \geq 0\}$ and $s(x) = \{0, \text{otherwise}\}$. The notation $LBP_{(p,r)}u2$ is used for the LBP operator, where (*p, r*) represents the neighborhood, and *u2* stands for uniform patterns and labeling all reaming patterns with a single label. The histogram for the image $f_1(x, y)$ is defined as:

$$H_i = \sum_{(x,y)} I\{f_i(x, y) = i\}, i = 0, ..., n-1, \quad (6)$$

The number of different labels produced by the LBP operator, and $I\{A\}$ is 1, if A is true and 0 if it is false. Further, the image patches whose histograms are to be compared must be normalized in order to get a coherent description:

$$N_i = \frac{H_i}{\sum_{j=0}^{n-1} H_j} \quad (7)$$

Then, the block LBP histograms were concatenated into an unmarried vector. The histograms have then been evaluated by using space similarity [16]. Moreover, each bin in histograms consists of the variety of its look within the region. Lastly, the feature vector is constructed with the useful data by concatenating the community histograms to one massive histogram.

IV. RESULTS AND DISCUSSION

In this study, the SFEW dataset was used for testing, which is close to real world environment, having 300 color images with 6 emotion categories, consisting of 50 pictures each with dimensions of 143×181 pixels. The classes are Surprise, Fear, Anger, Sadness, Disgust, and Happiness represented by SU, F, A, S, D, and H respectively. The results were evaluated with assessment metrics, including confusion matrix, precision, recall, and F1 score. To compute the overall precision, we used micro-averages to combine the consequences across the 6 categories. We divided our dataset into 80% training and 20% testing subsets. The sets were fed to the distinctive learning system which utilized algorithms such as K-Nearest Neighbor (KNN), Decision Tree (DT), Multilayer Perceptron (MLP), and Random Forest (RF). Our experimental model was divided into four parts. The mentioned machine learning algorithms were applied directly to the first part of the dataset. Table I shows the original dataset accuracies.

TABLE I. ORIGINAL IMAGES WITHOUT FEATURE EXTRACTION

Algorithm	KNN	DT	MLP	RF
Accuracy	27%	14%	22%	32%

TABLE II. LBP FEATURE EXTRACTION

Algorithm	KNN	DT	MLP	RF
Accuracy	50%	96%	22%	95%

Maximum accuracy was achieve by the RF and was only 32%. Then, all algorithm *accuracies* were computed using DWT, LPB, and HOG for the other parts of the dataset (Tables II-IV), and finally combined them and achieved 87% maximum accuracy with MLP and 29% minimum accuracy with KNN (Table V), which are respectively shown in the confusion matrices of Figures 2 and 3.

TABLE III. DWT FEATURE EXTRACTION

Algorithm	KNN	DT	MLP	RF
Accuracy	12%	17.5%	36%	21%

TABLE IV. HOG FEATURE EXTRACTION

Algorithm	KNN	DT	MLP	RF
Accuracy	12%	12%	37%	14%

TABLE V. COMBINATION OF LBP, DWT, AND HOG

Algorithm	KNN	DT	MLP	RF
Accuracy	29%	80%	87%	79%

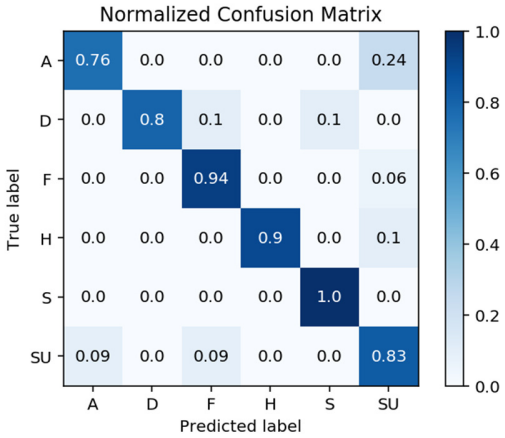

Fig. 2. MLP confusion matrix.

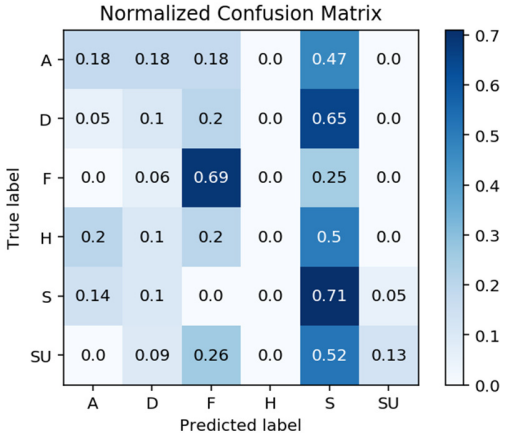

Fig. 3. KNN confusion matrix.

Further, we also calculated some edges of the face generated by DWT, LBP, and HOG. The original image is reconstructed using Harr DWT techniques.

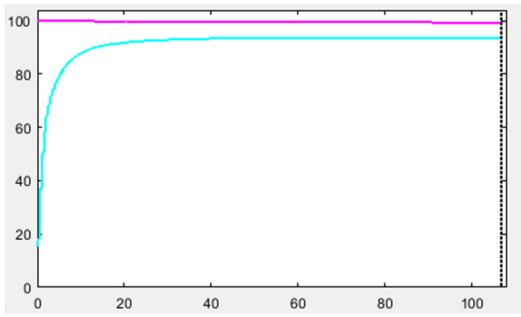

Fig. 4. Retained energy is 99.40%.

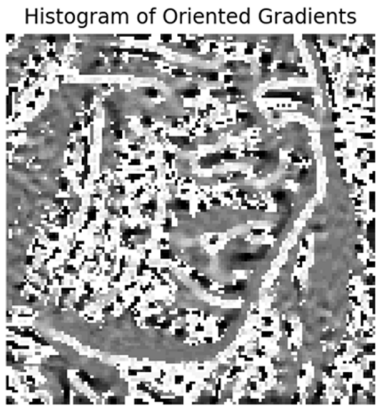

Fig. 5. HOG edges.

Fig. 6. LBP visualized surprised face.

V. CONCLUSION AND FUTURE WORK

The proposed model combines DWT, HOG, and LBP capabilities in a feature extraction technique with system learning algorithms in an excellent way of enhancing the accuracy of facial feature recognition. Six facial expressions from the SFEW database had been used for training and validation. The results indicated that the accuracy of the use of blended methods is 87%, which is higher from the individual accuracies of the combined algorithms. However, the proposed combination has the issue of generalization which may be addressed in our future work.

FER is one of the most well-known regions in image processing. Generally, FER has been given more attention nowadays. The proposed technique gives an exquisite overview of facial recognition methods. The extraction of functions is vital as it decreases the very massive amount of data to only a required set. Thus, it reduces the processing time of the machine and the results are more correct. In future work, the accuracy may be augmented by using more learning algorithms. A similar approach to the usage of the Convolution Natural Community can be combined with the prevailing support vector classifier.

ACKNOWLEDGEMENT

The authors would like to thank Dr. Abhinav Dhall, Australian National University for the provision of the SFEW dataset.

REFERENCES

[1] W.-L. Chao, J.-J. Ding, and J.-Z. Liu, "Facial expression recognition based on improved local binary pattern and class-regularized locality preserving projection," *Signal Processing*, vol. 117, pp. 1–10, Dec. 2015, https://doi.org/10.1016/j.sigpro.2015.04.007.

[2] F. Long and M. S. Bartlett, "Video-based facial expression recognition using learned spatiotemporal pyramid sparse coding features," *Neurocomputing*, vol. 173, pp. 2049–2054, Jan. 2016, https://doi.org/10.1016/j.neucom.2015.09.049.

[3] H. Fang *et al.*, "Facial expression recognition in dynamic sequences: An integrated approach," *Pattern Recognition*, vol. 47, no. 3, pp. 1271–1281, Mar. 2014, https://doi.org/10.1016/j.patcog.2013.09.023.

[4] J. Hussain Shah, M. Sharif, M. Raza, M. Murtaza, and S. Ur-Rehman, "Robust Face Recognition Technique under Varying Illumination," *Journal of applied research and technology*, vol. 13, no. 1, pp. 97–105, 2015.

[5] S. Arya, N. Pratap, and K. Bhatia, "Future of Face Recognition: A Review," *Procedia Computer Science*, vol. 58, pp. 578–585, Jan. 2015, https://doi.org/10.1016/j.procs.2015.08.076.

[6] M.-Y. Chen and C.-C. Chen, "The contribution of the upper and lower face in happy and sad facial expression classification," *Vision Research*, vol. 50, no. 18, pp. 1814–1823, Aug. 2010, https://doi.org/10.1016/j.visres.2010.06.002.

[7] A. Fernandez, O. Ghita, E. Gonzalez, F. Bianconi, and P. F. Whelan, "Evaluation of robustness against rotation of LBP, CCR and ILBP features in granite texture classification," *Machine Vision and Applications*, vol. 22, no. 6, pp. 913–926, Nov. 2011, https://doi.org/10.1007/s00138-010-0253-4.

[8] S. Shankar and V. R. Udupi, "Recognition of Faces – An Optimized Algorithmic Chain," *Procedia Computer Science*, vol. 89, pp. 597–606, Jan. 2016, https://doi.org/10.1016/j.procs.2016.06.020.

[9] R. K. Nagar, R. Manazhy, and P. Sankaran, "Sparse Manifold Discriminant Embedding for Face Recognition," *Procedia Computer Science*, vol. 89, pp. 743–748, Jan. 2016, https://doi.org/10.1016/j.procs.2016.06.050.

[10] T. Gao, X. L. Feng, H. Lu, and J. H. Zhai, "A novel face feature descriptor using adaptively weighted extended LBP pyramid," *Optik*, vol. 124, no. 23, pp. 6286–6291, Dec. 2013, https://doi.org/10.1016/j.ijleo.2013.05.007.

[11] K. Yu, Z. Wang, L. Zhuo, J. Wang, Z. Chi, and D. Feng, "Learning realistic facial expressions from web images," *Pattern Recognition*, vol. 46, no. 8, pp. 2144–2155, Aug. 2013, https://doi.org/10.1016/j.patcog.2013.01.032.

[12] R. A. Khan, A. Meyer, H. Konik, and S. Bouakaz, "Framework for reliable, real-time facial expression recognition for low resolution images," *Pattern Recognition Letters*, vol. 34, no. 10, pp. 1159–1168, Jul. 2013, https://doi.org/10.1016/j.patrec.2013.03.022.

[13] S. Ali Khan, A. Hussain, and M. Usman, "Facial expression recognition on real world face images using intelligent techniques: A survey," *Optik*, vol. 127, no. 15, pp. 6195–6203, Aug. 2016, https://doi.org/10.1016/j.ijleo.2016.04.015.

[14] S. Ali Khan, A. Hussain, A. Basit, and S. Akram, "Kruskal-Wallis-Based Computationally Efficient Feature Selection for Face Recognition," *The Scientific World Journal*, vol. 2014, May 2014, Art. no. e672630, https://doi.org/10.1155/2014/672630.

[15] C. Shan, S. Gong, and P. W. McOwan, "Facial expression recognition based on Local Binary Patterns: A comprehensive study," *Image and Vision Computing*, vol. 27, no. 6, pp. 803–816, May 2009, https://doi.org/10.1016/j.imavis.2008.08.005.

[16] W.-H. Chen, P.-C. Cho, P.-L. Fan, and Y.-W. Yang, "A framework for vision-based swimmer tracking," in *International Conference on Uncertainty Reasoning and Knowledge Engineering*, Bali, Indonesia, Aug. 2011, vol. 1, pp. 44–47, https://doi.org/10.1109/URKE.2011.6007835.

[17] D. Zecha, T. Greif, and R. Lienhart, "Swimmer detection and pose estimation for continuous stroke-rate determination," in *Multimedia on Mobile Devices 2012; and Multimedia Content Access: Algorithms and Systems VI*, California, United States, Jan. 2012, vol. 8304, Art. no. 830410, https://doi.org/10.1117/12.908309.

[18] S. S. Intille, J. W. Davis, and A. F. Bobick, "Real-time closed-world tracking," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, San Juan, USA, Jun. 1997, pp. 697–703, https://doi.org/10.1109/CVPR.1997.609402.

[19] K.-A. Toh, W.-Y. Yau, and X. Jiang, "A reduced multivariate polynomial model for multimodal biometrics and classifiers fusion," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, no. 2, pp. 224–233, Feb. 2004, https://doi.org/10.1109/TCSVT.2003.821974.

[20] M. Kharrat, Y. Wakuda, N. Koshizuka, and K. Sakamura, "Automatic waist airbag drowning prevention system based on underwater time-lapse and motion information measured by smartphone's pressure sensor and accelerometer," in *IEEE International Conference on Consumer Electronics*, Las Vegas, NE, USA, Jan. 2013, pp. 270–273, https://doi.org/10.1109/ICCE.2013.6486891.

[21] M. Kharrat, Y. Wakuda, S. Kobayashi, N. Koshizuka, and K. Sakamura, "Near drowning detection system based on swimmer's physiological information analysis," presented at the World Conference on Drowning Prevention (WCDP), May 2011.

[22] E. McAdams *et al.*, "Wearable sensor systems: The challenges," in *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Boston, MA, USA, Sep. 2011, pp. 3648–3651, https://doi.org/10.1109/IEMBS.2011.6090614.

[23] A. Ben-Hur, D. Horn, H. T. Siegelmann, and V. Vapnik, "Support Vector Clustering," *Journal of Machine Learning Research*, vol. 2, pp. 125–137, 2001.

[24] Y. Said, M. Barr, and H. E. Ahmed, "Design of a Face Recognition System based on Convolutional Neural Network (CNN)," *Engineering, Technology & Applied Science Research*, vol. 10, no. 3, pp. 5608–5612, Jun. 2020, https://doi.org/10.48084/etasr.3490.

[25] D. Virmani, P. Girdhar, P. Jain, and P. Bamdev, "FDREnet: Face Detection and Recognition Pipeline," *Engineering, Technology & Applied Science Research*, vol. 9, no. 2, pp. 3933–3938, Apr. 2019, https://doi.org/10.48084/etasr.2492.