ARUNAI PUBLICATIONS

Asian Journal of Applied Research
DOI: http://dx.doi.org/10.20468/ajar.2018.10.12

Research Article

# Gender Classification Using Histogram of Oriented Gradients–Support Vector Machine Classifier

**S. Gomathi Meena[1], K. G. Srinivasagan[2]**

*[1]Department of Computer Science, Vidhya Sagar Women's College, Chengalpattu, Tamil Nadu, India, [2]Department of Information Technology, National Engineering College, K. R. Nagar, Kovilpatti, Tamil Nadu, India*

## ABSTRACT

The classification of gender using face images is an area great significance and has been used extensively in surveillance, monitoring, and so on. This paper presents gender classification based on the combination of histogram of oriented gradients (HoG) with *k*-neural network and support vector machine classifiers in three different face regions - external, general, and internal. The performance analysis of the proposed method is conducted using two standard databases - FEI Face Database and Institute National de Recherche en Informatique et en Automatique. The performance is compared with the earlier techniques such as robust local binary pattern, robust local ternary pattern, and distributed robust local binary pattern.

**Address for correspondence:**
S. Gomathi Meena,
No. 67, 6th Street,
Priya Nagar 3rd Part,
Urapakkam,
Kanchipuram - 603210,
Tamil Nadu, India.

**Key words:**
DRLBP, Histogram of oriented gradients, LDN, Neural network, Robust local binary pattern, Robust local ternary pattern, Support vector machine

**Received:** 05th February 2018
**Accepted:** 20th September 2018
**Published:** 13th October 2018

## INTRODUCTION

Classification has emerged as a leading technique for problem solution and optimization. Classification has been used extensively in several problems domains. Automated gender classification is an area of great significance and has great potential for future research. It offers several industrial applications in near future, such as monitoring, surveillance, commercial profiling, and human-computer interaction. Different methods have been proposed for gender classification such as gait, iris, and hand shape. However, majority of techniques for gender classification are based on facial information. A comparative study of gender classification using different techniques is presented. The major emphasis of this work is on the critical evaluation of different techniques used for gender classification. The comparative evaluation has highlighted major strengths and limitations of existing gender classification techniques. Taking an overview of these major problems, our research is focused on summarizing the literature by highlighting its strengths and limitations. This study also presents several areas of future research in the domain of gender classification. In many social interactions, it is important to correctly recognize the gender. Researchers have addressed this issue based on facial images, ear images, and gait. The gender classification using facial images is based on sparse representation and basis pursuit. In sparse representation, the training data are used to develop a dictionary based on extracted features. Classification is achieved by representing the extracted features of the test data using the dictionary. For this purpose, basis pursuit is used to find the best representation by minimizing the [l 1] norm. In this work, Gabor filters are used for feature extraction. Experimental results are conducted on the FEI and INRIA datasets, and the obtained results are compared with other works in this area. The results show improvement in gender classification over existing methods.

This paper is organized as follows: Section 2 presents literature survey. Sections 3 and 4 present the texture classification using distributed robust local binary pattern (DRLBP) and LDN, respectively. Section 5 describes the feature selection. Section 6 deals with feature extraction using histogram of oriented gradients (HoG). Sections 7 and 8 deal with the prediction of support vector machine (SVM) and neural network (*k*-NN). Experimental results of the proposed work are presented in Section 9. Section 10 presents the summary and the scope of future work.

## LITERATURE SURVEY

In vision and learning, low computational complexity and high generalization are two important goals for video object detection. Low computational complexity here means not only fast speed but also less energy consumption. The sliding window object detection method with linear SVMs is a general object detection framework. The computational cost is herein mainly paid in complex feature extraction and inner product-based classification. The framework distributed object detection is developed by making the best use of spatial temporal correlation, where the process of feature extraction and classification is distributed in the current frame and several previous frames. In each framework, only sub feature vectors are extracted, and the response of partial linear classifier (i.e., sub-decision value) is computed.[1]

Sparse representation (or coding)-based classification (SRC) has gained great success in face recognition in recent years. However, SRC emphasizes the sparsely too much and overlooks the correlation information which has been demonstrated to be critical in real-world face recognition problems. Besides, some paper considers the correlation but overlooks the discriminative ability of sparsely. In general, the representation model is adaptive to the correlation structure that benefits from both 1-norm and 2-norm. Extensive experiments conducted on publicly available datasets verify the effectiveness and robustness of the proposed algorithm by comparing it with the state-of-the-art methods.[2]

Affective computing, the emergent field in which computers detect emotions and project appropriate expressions of their own, has reached a bottleneck where algorithms are not able to infer a person's emotions from natural and spontaneous facial expressions captured in video. While the field of emotion recognition has seen many advances in the past decade, a facial emotion recognition approach has not yet been revealed which performs well in unconstrained settings. A limitation of the current work is that the frames are processed in evenly sized segments, which may cause a boundary effect if an unlabeled apex is close to the segmentation boundary. However, this can be addressed using overlapped boundary segments.[3] Student engagement is a key concept in contemporary education, where it is valued as a goal in its own right. The automatic recognition of students' engagement from their facial expressions with the help of machine learning and the observers found that the discrimination at the low versus high degrees if engagement (Cohen's k ¼ 0:96). When fine discrimination is required (four distinct levels), the reliability decreases but is still quite high (k ¼ 0:56). Furthermore, the engagement labels of 10-s video clips can be reliably predicted from the average labels of their constituent frames (Pearson r ¼ 0:85), suggesting that static expressions contain the bulk of the information used by observers. Educational videos could be improved based on the aggregate engagement signals provided by the viewers. The signal not only indicates high or low engagement of the videos but also its parts to recognize the students' facial expressions. Our work underlines the importance of focusing on long-term field studies in real-life classroom environments. Collecting data in such environments are critical to train more reliable and ecologically valid engagement recognition systems. More importantly, sustained, long-term studies in actual classrooms are needed to gain a better understanding of the interplay between engagement and learning in real life.[4] Automatic facial expression analysis systems are aiming toward the application of computer vision techniques in human-computer interaction, emotion analysis, and even medical care through a space mapping between the continuous emotion and a set of discrete expression categories. The main difficulty with these systems is the inherent problem of facial alignment due to person-specific appearance. Besides the facial representation problem, the same displayed facial expression may vary differently across humans; this can be true even for the same person in different contexts. To cover all these variable factors, a prototype-based model is chosen as an anchor model and the expressions in the face images are mapped using the SIFT-flow registration. A set of prototype facial expression models is generated as a reference space of emotions on which face images are projected to generate a set of registered faces. Locality is an interesting property of SIFT which makes the descriptor more robust to occlusion and clutter. Accordingly, face regions that are important for facial expression recognition resist more to hair or hand occlusions in a realistic conversational context. A good example of such complex conditions exists in the GEMEP-FERA dataset which was considered to test the limits of face alignment. With a generalization performance of 83%, our SF/HoG obtained the second place <0.95% behind the winner of the FERA challenge and 6.6 times faster. Over the different evaluation protocols, our results have demonstrated that the SF/HoG descriptor may allow the construction of efficient generic facial expression recognition systems that can meet the real-time requirement. Finally, using shape prototypes together with appearance prototypes as a hybrid approach would certainly enhance the recognition accuracy. Future work will be targeted in this regard. Besides, much effort has to be done on the accurate detection and tracking of facial feature points for shape extraction because the performance will surely be very sensitive to their locations[5] for the task of robust face recognition, the focus is mainly on the corruption of training and test data which is occurred due to occlusion or disguise. Prior standard face recognition methods such as Eigen faces or state-of-the-art approaches such as sparse representation-based classification did not consider possible contamination of data during training, and thus, their recognition performance on corrupted test data would be degraded. A novel approach to face recognition algorithm is based on low-rank matrix decomposition to address the aforementioned problem. As a result, additional discriminating ability is added to the derived base matrices for improved recognition performance. The introduction of structural incoherence between low-rank matrices promotes the discrimination between different classes, and thus, the associated models exhibit excellent discriminating ability. The detailed derivations provided and showed that the proposed optimization problem can be solved by advancing augmented Lagrange multipliers. Our experiments on four face databases confirmed that our proposed methods are robust to severe illumination variations, occlusion, and random pixel noise corruptions, while our method has been shown to outperform state-of-the-art face recognition algorithms.[6]

SVM models are learnt on the Grossman followed by a voting-based strategy for classification. The theory of Rehashes been explored to adapt the SVM classifier for the

Grassmannian manifold. A new Grassmannian kernel function is also proposed. The performance of the system is tested on the largest publicly available 3D video database, BU4DFE. In comparison to previously published methods on BU4DFE database, our system shows a superior performance in terms of the classification accuracy. Moreover, the proposed system avoids the computationally expensive pre-processing steps for the establishment of a dense vertex level correspondence. Furthermore, it does not require any user intervention for manual annotation of facial landmarks. The system does not make any assumptions about the presence of all four expression segments in a video and performs equally well for all video types.[7]

Although facial expressions can be decomposed in terms of action units (AUs) as suggested by the facial action coding system, there have been only a few attempts that recognize expression using AUs and their composition rules. The preliminary synthesis of the experiment is to show the potential of the proposed algorithm. An important application for expression synthesis through AU composition is to synthesize some real expressions other than the universal one (e.g. frustration, empathy, contempt, interestedness, and boredom). Since having ground truth on these expressions is difficult, the proposed approach can be applied to synthesize such expressions. However, to synthesize such expressions we need to know the combination of action units and the Auks' building blocks to derive the efficacy and effectiveness of dictionary-based approaches for these kind of problems.[8]

The proposed gait feature extraction process is performed in the spatio-temporal domain. The space-time interest points (STIPs) are detected by considering large variations along both spatial and temporal directions in local spatio-temporal volumes of a raw gait video sequence. Thus, STIPs are allocated, where there are significant movements of the human body in both space and time. A HoG and a histogram of optical flow are computed on a 3D video patch in a neighborhood of each detected STIP, as a STIP descriptor. Then, the bag of words model is applied on each set of STIP descriptors to construct a gait feature for representing and recognizing an individual gait. When compared with other existing methods in the literature, it has been shown that the performance of the proposed method is promising for the case of normal walking and is outstanding for the case of partial occlusion caused by walking with carrying a bag and walking with varying a cloth type. This paper has proposed a new method for gait recognition. It constructs a new gait feature directly from a raw video without a pre-processing of foreground–background segmentation. The proposed gait feature is extracted in the spatio-temporal domain. The STIPs are detected from a raw gait video sequence. They represent significant movements of the human body along both spatial and temporal directions. Then, HoG and HoF are used to describe each detected STIP. Finally, a gait feature is constructed by applying bow on a set of HoG/HoF-based STIP descriptors from each gait sequence. It can be seen that the proposed gait feature relies on local motion information which is more robust to walking variations than global shape information used in most of existing methods for recognizing gaits. The proposed method has been reported to achieve the promising performance for the case of no variation and to achieve the significantly better performance for the case of large variations caused by clothing and carrying condition changes.[9]

The results showed that such perceptions have a deep impact on people's decisions: In Experiment 1, people cooperated more with virtual humans that showed cooperative facial displays (e.g., joy after mutual cooperation) than competitive displays (e.g., joy when the participant was exploited), but the effect was stronger with avatars (d ¼: 601) than with agents (d ¼: 360); in Experiment 2, people conceded more to angry than neutral virtual humans, but again, the effect was much stronger with avatars (d ¼ 1: 162) than with agents (d ¼: 066). Participants also showed less anger toward avatars and formed more positive impressions of avatars when compared to agents.[10]

The method is computationally efficient with a much higher rate of accuracy compared with existing gait recognition approaches. This paper describes a hierarchical method for frontal gait recognition using kindest depth and skeleton streams. A more detailed and discriminative feature is considered in the successive steps of the hierarchical classification. Such a step-wise classification scheme removes vastly dissimilar elements of the search space at each level, thereby enabling the final classification to be performed on a small percentage of elements selected from a much larger search space. The elimination of dissimilar elements at each level of hierarchy helps in achieving a better classification performance at the subsequent levels as there is a lower probability that the classifier gets biased toward an utterly different element of the search space. The search space reduction based on the cumulative match characteristic curves ensures that the correct element to be searched does not fall outside the reduced search space. In the proposed gait recognition procedure, the first two levels of hierarchy deal only with the skeleton joint information provided by Kindest SDK. For an accurate prediction, the final classification is done considering the detailed depth information of the test subject from the back view. The proposed method is quite accurate even in the presence of incomplete information in both the training and the test samples. The simple experimental set-u together with the reasonably accurate classification performance emphasizes the effectiveness of the proposed method in the considered application scenarios. The robustness of this approach could be derived by conducting more experiments with larger datasets. Reconstruction of the skeleton structures of the subject for an entire gait cycle from the partial information available from the front view and motion analysis using the reconstructed sequence would be a direction for future research.[11]

A two-stage sparse learning model is proposed to learn the locations of these patches based on the prior knowledge of facial muscles and Auks' a multiscale face division strategy is employed to obtain facial patches with different coverage area and eliminate the side effects from fixed patch size. The effectiveness of these patches is evaluated by facial expression recognition. Extensive experiments show that common patches can generally discriminate all the expressions and the recognition performance can be further improved by integrating specific patches. More comprehensive patches can also be selected to achieve better performance using multistage patch division strategy. The learned location information of

these patches also confirms the location knowledge of facial muscles in psychology.[12]

Sparse representation of signals for classification is an active research area. Signals can potentially have a compact representation as a linear combination of atoms in an overcomplete dictionary. Based on this observation, a SRC has been proposed for robust face recognition and has gained popularity for various classification tasks. It relies on the underlying assumption that a test sample can be linearly represented by a small number of training samples from the same class. However, SRC implementations ignore the Euclidean distance relationship between samples when learning the sparse representation of a test sample in the given dictionary. To overcome this drawback, the classification is performed using class-dependent sparse representation classifier (cusec) is proposed for hyperspectral image classification, which effectively combines the ideas of SRC and K-nearest neighbor classifier in a class-wise manner to exploit both correlation and Euclidean distance relationship between test and training samples. Toward this goal, a unified class membership function is developed, which utilizes residual and Euclidean distance information simultaneously? Experimental results based on several real-world hyperspectral datasets have shown that cusec not only dramatically increases the classification performance over SRC but also outperforms other popular classifiers, such as SVM. The HIS classification is performed using the classifier. In cusec, a test sample is represented in a way that exploits the correlation and Euclidean distance information between the test sample and the training samples in a class-wise manner. Through experimental results based on three real-world hyperspectral datasets, it is clear that cusec not only dramatically improves the performance of traditional SRC but also outperforms popular traditional classifiers, including two-regularized classifiers and SVM. Additional improvements in classification performance can be observed with the kernel variant of cusec.[13]

To evaluate the performance of our CoDe4D LST features and the complete system, the experiment is conducted using four benchmark color-depth human activity datasets, including UTK Action3-D, Berkeley MHAD, ACT42, and MSR daily activity 3-D data sets. Experimental results demonstrate the promising representative power of our CoDe4D features, which obtain the state-of-the-art performance on activity recognition from RGB-D visual data. We introduce a novel LST feature that is able to incorporate both color and depth information contained in a sequence of RGB-D frame activity 3-D color-depth activity datasets. Experimental results demonstrate that the proposed CoDe4D LST features present satisfactory representation power and achieve the state-of-the-art activity recognition performance.[14] There are four contributions to solve this problem: (1) A non-parametric algorithm is derived to utilize abundant unlabeled data to obtain an accurate quality measure for node splitting based on kernel-based density estimation and the law of total probability; (2) to adaptively select the optimal bandwidths for kernel-based density estimation of different categories, a multiclass version of the AMISE criterion was proposed; (3) to avoid the curse of dimensionality, the data points were projected from the original high-dimensional feature space onto a low-dimensional subspace before estimating the categorical distributions; and (4) a unified optimization framework was proposed to select a coupled pair of subspace and separating hyper plane for each node such that the smoothness of the subspace and the quality of the splitting are guaranteed simultaneously. Our method can be combined with many popular splitting criteria, and the experimental results showed that it brings obvious performance improvements to all of them. In the future, we would like to investigate the problem of constructing RFs without labeled data. A unified splitting framework that can handle both labeled and unlabeled data would be the extension.[15]

## TEXTURE CLASSIFICATION USING DRLBP

Robust local binary pattern (RLBP) is sensitive to noise and small pixel value fluctuations. LTP solves this problem using two thresholds that create three different states. It is more resistant to noise and small pixel variations. RLBP is used to solve the problem of LBP and show the difference in bright object against a dark background and vice versa. However, it is not applicable to ULBP and LLBP of LTP. DRLBP overcomes all these issues, and it retains both the edge and the texture information that is desirable to distinguish face images clearly.

$$DRLBP = \sum_{i=1}^{i=9} w(x,y) * RLBP(x,y) \quad (1)$$

Where $w(x,y)$ is calculated by gradient operator by finding the square root of the magnitude in $x$ and $y$ directions.

## TEXTURE CLASSIFICATION USING LDN

LDN encodes the directional information of the face images textures in a dense way producing a more discriminative code than current methods. The structure of each micro-pattern are figured out using the compass mask that encodes the directional information and encodes that information using the well-known direction indices along with its sign, which allows us to distinguish among similar structural patterns that have different intensity transitions. Then, the face images are divided into several regions and take out the distribution of the LDN features from them and then concatenate these features into a feature vector and used it as a face descriptor.

The descriptor performs consistently under illumination, noise, expression, and the time lapse variations. The LDN pattern is a six-bit binary code assigned to each pixel of an input image that represents the structure of the texture and its intensity transitions. As previous research indicates, edge magnitudes are largely insensitive to lighting changes. Accordingly, the pattern is created by computing the edge response of the neighborhood using a compass mask and by capturing the top directional numbers which are both positive and negative directions of those edge responses.

The positive and negative responses give valuable information of the structure of the neighborhood as they expose the gradient direction of bright and dark areas in the neighborhood. Thus, this distinction between dark and bright responses allows LDN to discriminate between blocks

with the positive and the negative directions swapped which are equivalent to swap the bright and the dark areas of the neighborhood by generating a different code for each case, while other methods may mistake the swapped regions as one. These transitions also occur repeatedly in the face image, for example, the top and the bottom edges of the eyebrows and mouth have different intensity transitions.

The code is generated using LDN, by analyzing the edge response of each mask {M0.,M7} that represents the edge significance in its respective direction and by combining the dominant directional numbers. Given that the edge responses are not equally important than the presence of a high negative or positive value signals in a prominent dark or bright area. The regions are encoded using the implicit use of sign information to assign a fixed position for the top positive directional number as the three most significant bits in the code and the three least significant bits are the top negative directional numbers.

$$LDN(x,y)=8i_{x,y}+j_{x,y} \quad (2)$$

Where (x,y) is the central pixel of the neighborhood being coded, $i_{x,y}$ is the directional number of the maximum positive response, and $j_{x,y}$ is the directional number of the minimum negative response defined by:

$$i_{x,y}=\arg\max i\{II_i(x,y)\,|\,0\le i\le 7\} \quad (3)$$

$$j_{x,y}=\arg\max j\{II_i(x,y)\,|\,0\le j\le 7\} \quad (4)$$

Where $\prod_i$ is the convolution of the original image, I, and the $i^{th}$ mask, Mi is defined by:

$$IIi=I*Mi \quad (5)$$

To create the LDN code, a compass mask is used to compute the edge responses. The proposed code is analyzed using two different asymmetric masks: Kirsch and derivative-Gaussian. Both masks operate in the gradient space that reveals the face image structure. Gaussian smoothing is used to stabilize the code in the presence of noise using the derivative Gaussian mask. The Kirsch mask is rotated apart to obtain the edge response in eight different directions. This indicates the use of this mask to produce the LDN code by LDNK. Kirsch mask uses the derivative of a skewed Gaussian to create an asymmetric compass mask that is used to compute the edge response on the smoothed face. This mask is strong against noise and illumination changes while producing strong edge responses. Therefore, the Gaussian mask is defined by:

$$G_{\bar{A}}(x,y)=\frac{1}{2\pi\sigma^2}\exp\left(-\frac{x^2+y^2}{2\sigma^2}\right) \quad (6)$$

Where x and y are location positions and their width of the Gaussian bell; this defines the mask as: Mσ(x,y) = Gσ(x+k,y) * Gσ(x,y) where y is the derivative of Gσ with respect to x, σ is the width of the Gaussian bell, * is the convolution operation, and k is the offset of the Gaussian with respect to its center. A compass mask is generated, {M0σ.,M7σ}, by rotating Mσ 45° apart in eight different directions. Thus, a set of masks is obtained.

## FEATURE SELECTION

This section demonstrates the role and effect of similar patterns, which are generally discarded in most of the existing LBP approaches because this also contains discriminative information. In general, similar patterns are known to exhibit high discriminative patterns, and it does not generalize across different datasets because, for complex images, similar patterns may not be the most frequently occurring ones. To overcome this kind of drawback, the fusion of similar patterns of DLRBP and LDN performed and it contributes to both reducing the length of the feature vector and improving the performance of classifiers. In the proposed algorithm, the face image is divided into three regions - external, general, and internal. The first region called external region is extracted from the input images and is scaled into 128 × 128 pixels. In few cases, if borders are encountered with the objects such as hand and hat, this kind of images is simply removed from the data. The next region called general region extracted from the input images and is scaled into 64 × 64 pixels, and it is smaller than the external region. The major part of the region covers the forehead, cheek, and chin. The final region called internal region which is also extracted from the input images is scaled into 32 × 32 pixels, and this mainly covers eyebrow, nose, and lip.

## FEATURE EXTRACTION USING HOG

Local object appearance and shape can often be characterized rather well by the distribution of local intensity gradients or edge directions, even without precise knowledge of the corresponding gradient or edge positions. This statement leads to the definition of the HoG technique that has been used in its mature form in Scale Invariant Features Transformation, and it has been widely exploited in human detection. HoG descriptor is based on the accumulation of gradient directions over the pixel of a small spatial region referred as “cell” and in the subsequent construction of a 1D histogram whose concatenation supplies the features vector to be considered for further purposes. Let *R* be an intensity (gray scale) function describing the image to be analyzed. The image is divided into cells of size $N \times N$ pixels and the orientation $O_{x,y}$ of the gradient in each pixel is computed by means of the following rule:

$$O_{x,y}=\tan^{-1}\frac{R(x,y+1)-R(x,y-1)}{R(x+1,y)-R(x-1,y)}$$

Successively, the orientations $O_i^j, i=1,\ldots,N^2$, i.e., belonging to the same cell *j* are quantized and accumulated into a M-bins histogram. Finally, all the achieved histograms are ordered and concatenated into a unique HoG histogram that is the final outcome of this algorithmic step, i.e., the features vector to be considered for the subsequent processing.

## PREDICTION OF SVM

The HoG feature vectors are then given as input to a group of SVMs. SVM is a discriminative classifier defined by a separating hyperplane. Given a set of labeled training data (supervised learning), the algorithm computes an optimal hyperplane (the trained model) which categorizes new examples in the right class. Given the training vectors $x_i \in \Re^n, i=1,\ldots,l$ and the corresponding set of *l* labels $y_i \in \{1, -1\}$, the following primal optimization problem is solved:

**Table 1:** Test results of FEI database

| Algorithm/parameter | Using SVM classifier with 100 images | | | | | |
|---|---|---|---|---|---|---|
| | External feature (128×128 Size) | | General feature (64×64 size) | | Internal feature (32×32 size) | |
| | *k*-NN | SVM | *k*-NN | SVM | *k*-NN | SVM |
| RLBP | 69 | 71 | 65 | 69 | 70 | 72 |
| RLTP | 68 | 70.5 | 68.5 | 71.4 | 69.6 | 71 |
| DRLBP | 73.7 | 76.9 | 72.8 | 74.8 | 76.9 | 77.4 |
| LDN | 71.7 | 73.6 | 72.4 | 74 | 75 | 77.8 |
| RLBP+LDN | 74.7 | 75 | 74.8 | 75.7 | 76 | 78.8 |
| RLTP+LDN | 84.6 | 86 | 83.8 | 86 | 87.5 | 88 |
| DRLBP+LDN | 87.6 | 88.9 | 85.9 | 89 | 90.8 | 93.8 |
| HoG | 85 | 87 | 84 | 86 | 89 | 90 |
| RLBP+LDN+HoG | 86 | 88 | 85 | 86 | 88 | 89 |
| RLTP+LDN+HoG | 85 | 89 | 88 | 89 | 87 | 93 |
| DRLBP+LDN+HoG | 87 | 90 | 92 | 93 | 94 | 96 |

SVM: Support vector machine, HoG: Histogram of oriented gradients, NN: Neural network, RLBP: Robust local binary pattern, RLTP: Robust local ternary pattern, DRLBP: Distributed robust local binary pattern

**Table 2:** Test results of INRIA database

| Algorithm/parameter | Using SVM classifier with 100 images | | | | | |
|---|---|---|---|---|---|---|
| | External feature (128×128 size) | | General feature (64×64 size) | | Internal feature (32×32 size) | |
| | *k*-NN | SVM | *k*-NN | SVM | *k*-NN | SVM |
| RLBP | 67 | 69 | 63 | 64 | 69 | 69 |
| RLTP | 66 | 69.5 | 66.5 | 69.4 | 67.6 | 69 |
| DRLBP | 71 | 72.3 | 70.8 | 73.4 | 74.9 | 75.4 |
| LDN | 70 | 71.3 | 70.4 | 72 | 73 | 75.8 |
| RLBP+LDN | 73.5 | 73 | 72.4 | 73.7 | 74 | 76.8 |
| RLTP+LDN | 82.4 | 84 | 81.8 | 84 | 85.5 | 86 |
| DRLBP+LDN | 85.2 | 86.9 | 83.9 | 86 | 89.8 | 91.8 |
| HoG | 83 | 85 | 82 | 84 | 87 | 89 |
| RLBP+LDN+HoG | 84 | 86 | 83 | 84 | 86 | 87 |
| RLTP+LDN+HoG | 85 | 86 | 86 | 87 | 85 | 91 |
| DRLBP+LDN+HoG | 87 | 88 | 89 | 91 | 91 | 94 |

SVM: Support vector machine, HoG: Histogram of oriented gradients, NN: Neural network, RLBP: Robust local binary pattern, RLTP: Robust local ternary pattern, DRLBP: Distributed robust local binary pattern

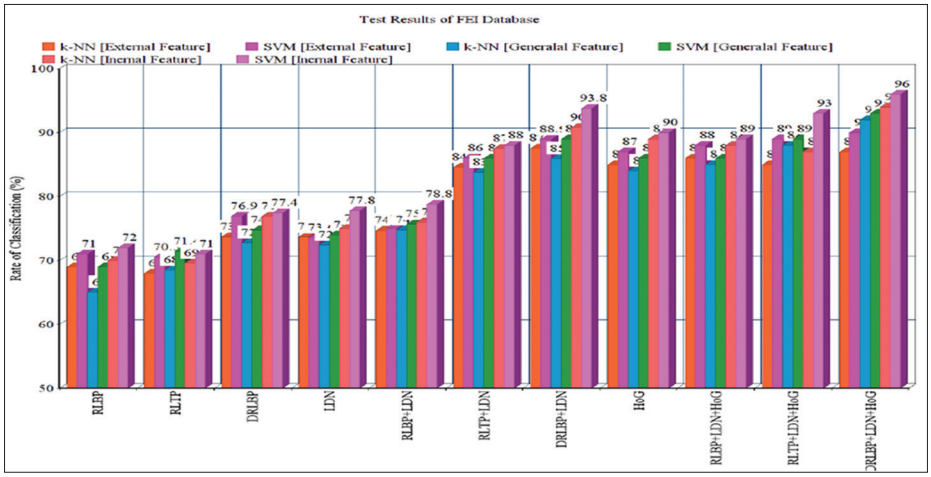


**Figure 1:** Test results of FEI database

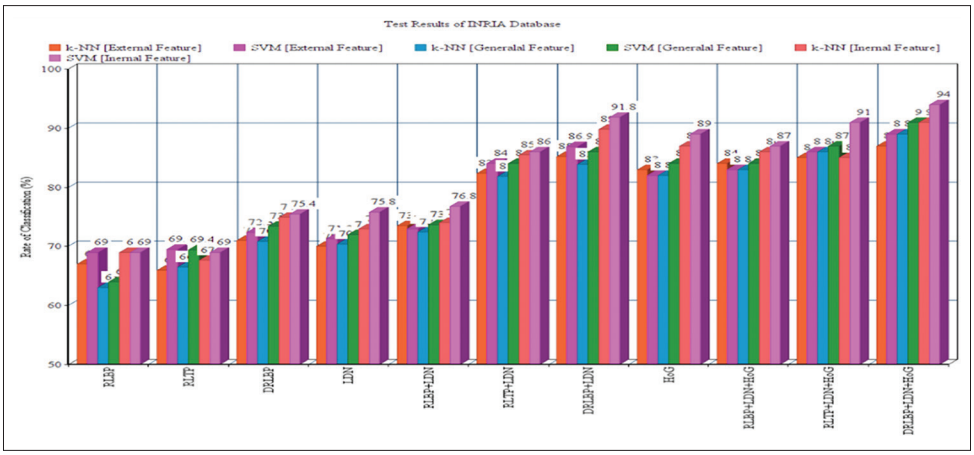


**Figure 2:** Test results of INRIA database

$$\min_{w,b,\xi} \frac{1}{2} w^T w + C \sum_{i=1}^{l} \xi_i$$

Subject to $y_i(w^T \phi(x_i) + b) \geq 1 - \xi_i$, $\xi_i = 1, \ldots l$

Where $\xi_i$ is the misclassification error for the $i^{th}$ training vector; $\xi$ is the total misclassification error; $w$ is the normal vector to the hyperplane; $b/||w||$ determines the offset of the hyperplane from the origin along the normal vector $w$ (with $||$ $||$ the norm operator); $\varphi(x_i)$ maps $x_i$ into a higher-dimensional space; and $C>0$ is the regularization parameter.

## *k*-NN classification

*k*-NN is the simplest non-parametric technique, and it belongs to the domain of supervised learning. The classification of objects is calculated using the Euclidian distance from its nearest neighbors. The cost of computation is bit expensive, and the classification of objects of the need to be computed for "N" number of face images available in the database. Even if the class boundaries are less distinct, the amount of noise level reduced much.

## EXPERIMENTAL RESULTS

An empirical study is done extensively to evaluate the existing approaches and the proposed techniques with various variables and conditions. The proposed method is tested on two different texture datasets - FEI Face Database and INRIA Database.

The first is the well-established FEI Face Database containing 2800 images from 14 different texture classes under different lighting conditions, rotations, and pose variations. The testing set consists of 2220 images, and the training set consists of 14 × 30 = 420 images, whereas the second one is the well-known database called INRIA containing 12,180 windows which are sampled randomly from 1218 negative training photos under different lighting conditions, rotations, and pose variations. The testing set consists of 11,440 images, and the training set consists of 740 images.

Based on the objective, the experiment dealt with recognition of face images in invariant conditions. However, there are some measures which are insensitive to variation under different illumination conditions. The position of the edges is also insensitive to changes in the illumination respect to background objects along with its reflection on the face but nor for smooth surfaces. The experimental results clearly showed that the impact of rotations and pose variations on external and general regions than the internal region. It can be observed that the modifications in the proposed descriptors have achieved highest accuracy in the internal region 96% (DRLBP+LDN+HoG) for FEI database and 94% (DRLTP+LDN+HoG) for INRIA database compared to RLBP+LDN+HoG and robust local ternary pattern (RLTP)+LDN+HoG for SVM classifier than *k*-NN. It clearly shows that the internal region supersedes the external and general regions and the summarized results are shown in Tables 1 and 2, and its corresponding histograms are shown in Figures 1 and 2, respectively.

## CONCLUSION

The fusion of DLRBP and LDN with HoG is analyzed as a descriptor for texture classification. The complete information about the neighborhood is analyzed by addressing the uniform patterns because it distinguishes similar-structured texture patterns. The obtained results demonstrate that the proposed descriptor gives better performance in the internal region and the rate of recognition is 96% for FEI Database and 94% for INRIA Database, respectively.

## REFERENCES

1. Yanwei P, Zhang K, Yuan SY, Wang K. Distributed object detection with linear SVMs. IEEE Trans Cybern 2014;44:2122-33.
2. Jing W, Lu C, Wang M, Li P, Yan S, Hu X. Robust face recognition via adaptive sparse representation. IEEE Trans Cybern 2014;44:2368-78.
3. Albert CC, Bhanu B, Thakoor NS. Vision and attention theory based sampling for continuous facial emotion recognition. IEEE Trans Affect Comput 2014;5:418-31.
4. Jacob W, Serpell Z, Lin YC, Foster A, Movellan JR. The faces of engagement: Automatic recognition of student engagement from facial expressions. IEEE Trans Affect Comput 2014;5:86-98.
5. Mohamed D, Meunier J. Prototype-based modeling for facial expression analysis. IEEE Trans Multimedia 2014;16:1574-84.
6. Chia-Po W, Chen CF, Wang YC. Robust face recognition with structurally incoherent low-rank matrix decomposition. IEEE

Trans Image Proc 2014;23:3294-307.
7. Munawar H, Bennamoun M. An automatic framework for textured 3D video-based facial expression recognition. IEEE Trans Affect Comput 2014;5:301-3.
8. Sima T, Qiu Q, Chellappa R. Structure-preserving sparse decomposition for facial expression analysis. IEEE Trans Image Process 2014;23:3590-603.
9. Worapan K. Recognizing gaits on spatio-temporal feature domain. IEEE Trans Inform For Secur 2014;9:1416-23.
10. Celso MM, Gratch J, Carnevale PJ. Humans versus computers: Impact of emotion expressions on people's decision making. IEEE Trans Affect Comput 2015;6:1-1.
11. Pratik C, Sural S, Mukherjee J. Frontal gait recognition from incomplete sequences using RGB-D camera. IEEE Trans Inform For Secur 2014;9:1843-56.
12. Lin Z, Liu Q, Yang P, Huang J, Metaxas DN. Learning multiscale active facial patches for expression analysis. IEEE Trans Cybern 2015;45:1499-510.
13. Minshan C, Prasad S. Class-dependent sparse representation classifier for robust hyperspectral image classification. IEEE Trans Geosci Remote Sens 2015;53:1592-606.
14. Hao Z, Parker LE. 'CoDe4D: Color-depth local spatio-temporal features for human activity recognition from RGB-D videos. IEEE Trans Circuits and Syst Video Technol 2016;26:541-55.
15. Xiao L, Song M, Tao D, Liu Z, Zhang L, Chen C, Bu J. Random forest construction with robust semisupervised node splitting. IEEE Trans Image Process 2015;24:471-83.

**Cite this article:** Meena SG, Srinivasagan KG. Gender Classification Using Histogram of Oriented Gradients–Support Vector Machine Classifier. Asian J Appl Res 2018;4(2):61-68.

**Source of Support:** Nil, Conflict of Interest: None declared.