Int. J. of Computer Vision and Signal Processing, 4(1), 29-38(2014) ORIGINAL ARTICLE

Dynamic Blocks for Face Verification

Mohammad Ibrahim*, Tajkia Rahman Toma, Md. Iftekharul Alam Efat, Shah Mostafa Khaled, Md. Shariful Islam, Mohammad Shoyaib

Institute of Information Technology
University of Dhaka, Dhaka 1000, Bangladesh

Abstract

Face recognition system is a computer based biometric information processing for automatically identifying or verifying a person from a digital image or a video frame. The significance of this research area is increasing day by day. Although the existing methods of face verification system perform well under certain conditions, there are still challenging problems due to several issues such as pose variation, facial expression variation, occlusion, imaging conditions, illuminations, size variations, age variations, orientations, etc. This paper addresses the problem of recognizing human faces despite the variations of pose and size. To handle these problems, we mainly focus on dynamic block size. Instead of uniform block, we propose Dynamic Size Blocks (DSB) considering most prominent face features such as eye, eyebrow, nose, mouth, chin, cheek, fore-head, etc., based on facial landmarks. In this feature based approach, we use a Dynamic Local Ternary Pattern (DLTP) for extracting facial feature information from each dynamic block. Then we perform a square-root of Chi-Square distance for similarity measurement of each block. We use a Support Vector Machine (SVM) classifier for face verification. We performed a comprehensive experimental evaluation on Labeled Faces in the Wild (LFW) dataset with restricted settings original images. Our proposed method has achieved an accuracy of 74.08% on all test images and 82.26% on dataset images excluding extreme pose variations.

Keywords: Dynamic Block, Face recognition, Face verification

ISSN: 2186-0114
http://www.cennser.org/IJCVSP

Article History:
Received: 30 August 2014
Revised: 8 November 2014
Accepted: 26 November 2014
Published Online: 30 November 2014

1. Introduction

Face recognition system is a biometric information processing of facial image. It is a computer based system for automatically identifying or verifying a person in real life uncontrolled condition such as from a digital image or a video frame. We can perform this recognition by comparing selected facial features from the given image and a dataset images. In advancement of computers power and recent research works in pattern recognition, face recognition systems can now perform in real life images and achieve satisfying performance under controlled conditions, leading to many potential applications.

Everyday millions of facial images are being uploaded in the Internet through Facebook, Picasa, Google Plus, etc. Furthermore, millions of surveillance images are being stored in different databases throughout the globe everyday. Facial image analysis has, therefore, an increasing demand in these social media to ensure security. Face analysis covers face detection, face identification, face verification, facial expression recognition, etc. In verification mode, the main concern of applications are access control, such as computer or mobile device log-in, building gate control, digital multimedia data access. Face verification has many advantages for example the biometric signature, video surveillance, information retrieval, multimedia data management, human computer interaction, personal settings identification [1].

A face recognition system can be used in two modes: one is verification (or authentication) and is identification. Most of the cases, verification and identification share the same classification algorithms. Identification is the process of identifying a person from a given picture having a prominent face on it and decides the identity of the face comparing to a set of faces. It attempts to establish the identity of a given person out of a pool of people in a database. On the

*Corresponding author
Email addresses: m.ibrahim.se@gmail.com (Mohammad Ibrahim), tajkiatoma@gmail.com (Tajkia Rahman Toma), iftekhar.efat@gmail.com (Md. Iftekharul Alam Efat), khaled@univdhaka.edu (Shah Mostafa Khaled), shariful@univdhaka.edu (Md. Shariful Islam), shoyaib@du.ac.bd (Mohammad Shoyaib)

other hand, confirming or denying the identity claimed by a person (one-to-one matching) is known as face verification. In verification, two pictures of two faces are provided and the decision is made, whether the two faces represent the same person or not.

Recognition system takes standard input images and the detection algorithms detect the faces and extract face images from the input image. The extracted features include eyes, eyebrows, nose, and mouth, etc. [2]. This makes the overall algorithm more complex than single detection or verification algorithm in face recognition. Generally a face recognition or verification system includes few steps such as image acquisition, face detection and cropping, recognition or verification and identity. In most of the previous works Viola and Jones face detector [3] is used for face detection, which works for frontal and near frontal faces. However, a verification system should be able to manage pose variations. The popular features such as Local Binary Pattern (LBP) [4], Local Ternary Pattern (LTP) [5] Dynamic Local Ternary Pattern (DLTP) [6], Gabor [7], etc. are used to extract features information from the detected face region. Finally person identity is used as a result of recognition part. An illustration of the steps for the face recognition system is given in Fig. 1.

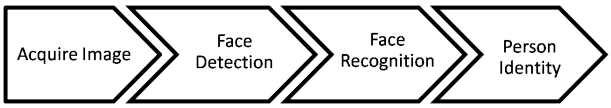


Figure 1: Steps of face recognition system

Although a good number of algorithms have been proposed for face recognition and verification. Many of these algorithms have been performed very well under certain controlled environments. But in real life uncontrolled environments, face recognition and verification is still a challenging problem due to several issues such as pose variations, facial expression, occlusion, illumination, imaging conditions, age variations, size and orientation, etc. Some examples of such unconstrained settings are given in Fig. 2.

Figure 2: Images from the same person may look quite different due to pose (upper left), expression (upper right), illumination (lower left), and occlusion (lower right)

In face recognitions systems, the most frequently used block size is uniform block. In uniform block, the whole image is divided into fixed number of blocks and every block is same size. If dataset images become frontal and in same alignment then uniform blocks perform well. But in pose variant images uniform block has some limitations.

Fig. 3 shows uniform block with pose variation. In this case, each face image is divided into 100 uniform blocks, 1 to 100. In first image Right eye is divided into 8 blocks where block positions are 46, 47, 48, 49 and 56, 57, 58, 59. But in second image, the same person's right eye is divided into 6 blocks due to pose where block positions are 38, 39, 40 and 48, 49, 50. In these two image, for right eye feature the total number of blocks is not same and these blocks position is also not same. So when we extract features from these two images and perform a similarity measurement using block-by-block matching then there is more probability for misclassifying these two images although both image person are same.

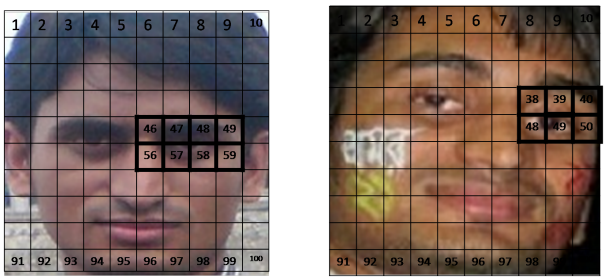


Figure 3: Uniform block with pose variation. Right eye is divided in 8 blocks (Left image), Right eye is divided in 6 blocks (Right image) and their position is different

This paper focuses to construct a fully machine based face verification system which works in real-time images. The system must be robust enough to handle pose and size variations in order to be used in a real-life applications. To resolve the problems of uniform blocks in pose and size variation we propose dynamic size blocks (DSB) based on landmark as in [8]. We use landmark for dynamically locating most prominent face features such as Eye, Eyebrow, Nose, Mouth, forehead etc. The state-of-the-art face detection and landmark localization [9] is used for this purpose. A family of novel face-image descriptor such as Dynamic Local Ternary Pattern (DLTP) is designed to capture local patches information. We performed a square root of *Chi-Square* distance for similarity measurement.We used a Support Vector Machine (SVM) [10] classifier to deal with challenges in face recognition. The proposed approach has been tested on Labeled Faces in the Wild (LFW) database and the results are found to be highly encouraging.

This paper is organized as follows: Section 2 contains related research on face recognition and verification. Section 3 presents proposed method for improving performance of the system with DSB in different pose and size variations. Section 4 presents detail about experimental data

and the results obtained. Finally Section 5 concludes the paper with future research direction.

2. Related Research

Although face recognition systems are known for few decades, there has been a number of relevant works addressing the problem of face recognition and verification. This section gives an overview about the most significant works in face recognition.

The basic question relevant to face classification is that what type of structural code should be considered to achieve face recognition. To address these recognition challenges for machine identification of human faces a number of algorithms has been proposed. These algorithms can be categorized in three main categories: Holistic approach, Feature based approach and Hybrid approach. We elaborate the related works according to their categories in the following sub-sections.

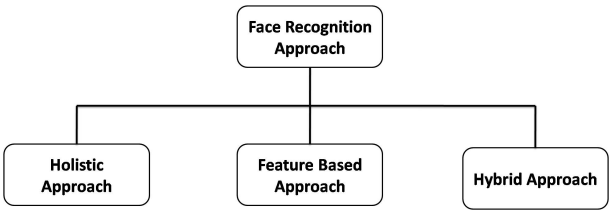


Figure 4: A hierarchy of face recognition approaches

2.1. Holistic Approaches

In holistic based approach the facial features are extracted from the entire face images and used as a single vector for classification. In this approach, the face image is usually divided into number of blocks. Different types of features such as: Gabor jet, LBP, LTP, DLTP, etc. are extracted and used as a whole for verification and/or recognition purposes. In holistic approach statistical methods such as Principal Component Analysis (PCA), Independent Component Analysis (ICA) etc. are applied to identify facial features to the whole image.

Cao et al. [11] presented a learning based (LE) descriptor and pose adaptive face matching technique to solve the representation and matching problem in pose variant face recognition. This learning based encoding method is an unsupervised learning method which encodes the local microstructures of the face into a set of discrete codes having uniform distribution. They applied PCA technique for the dimension reduction in code histogram by compressing the concatenated histogram. They also used a simple normalization after PCA compression which can enhance overall performance of the recognition system. At their proposed framework, they first extracted face landmark using a fiducial point detector and based on these landmarks. They detected nine different face components. They performed a difference of Gaussian (DoG) pre-processing on these identified face components to remove both low and high frequency illumination variations. They used a learning based encoder to extract feature vector from each component image. Then the final component representation is produced by the concatenated patch histogram of the encoded features after PCA reduction and normalization. The resulting nine component similarity scores are used into the pose adaptive classifier. This pose-adaptive matching method is used in pose specific classifiers to handle the large pose variation. They showed their experiments on both the LFW and Multi-PIE benchmark with a leading performance. In this approach, although the face micro-pattern encoding is learned but the pattern sampling is designed manually.

Tan and Triggs [5] proposed a local texture based feature extraction method called Local Ternary Pattern (LTP) which mainly follows Local Binary Pattern (LBP) [4]. This LTP is more discriminate and less sensitive to noise in uniform regions where LBP features are more sensitive to noise in near-uniform image region. They introduced a new bit to manage the intensity fluctuation which makes it a ternary code. They split a LTP code into two binary codes (upper pattern and lower pattern) to reduce the size of the feature vector. They also built separately two histograms for the two types of codes. They also incorporated illumination normalization based on Histogram of Oriented Gradient (HOG) features, distance transformation based matching and kernel based PCA feature extraction method for dimensionality reduction. They achieved a face verification rate of 88.1% on the challenging FRGC-204 dataset and achieved state-of-the-art result. However they used uniform blocks in their feature extraction method which has limitations to handle pose variations.

Brunelli et al. [12] proposed a template matching algorithm as the optimal strategy for face recognition. In their study, they compared two techniques: geometric feature based technique and template matching based system. In template matching, they automatically selected a set of four features template for all the available faces. The four features include eye, nose, mouth, and the whole face. Within each template, the given image region is evaluated with the same region of the database image through a normalized cross correlation using a distance metric. They used a data compression method for dimensionality reduction. The recognition decision was made using total matching scores. One major problem of template matching lies in the description of these templates. In geometric feature based technique, they used integral projection technique for the extraction of facial features. This projection technique is very useful in determining the position of features to construct low dimensional projections of a high dimensional point. They used Bayesian classifier in recognition process. The comparison between the results of geometric features and template matching shows that a template based approach performs better over geometric features based approach.

2.2. Feature Based Approaches

Feature based approaches use local face features such as eyes, nose, mouth, cheek, chin and head outline, etc. These features can be used to uniquely identify the individuals. However, the major challenge in feature based approach is that the recognition process is generally affected by the error-proneness of the features.

Arashloo et al. [13] used a Markov Random Field (MRF) matching algorithm for resolving the problem of pose-invariant recognition of faces. They mainly focused on reducing the processing time of MRF image matching. In this respect, they focused on parallel processing algorithm and used the well known decomposition approach [14]. Processing model in pixel level is computationally demanding and highly vulnerable to noise and local minima. So they proposed multi-resolution analysis based on Re-normalization Group Transform (RGT) [15] , efficient message passing using distance transformation and the incremental sub-gradient approach to resolve pixel level complexity. This work used multi-resolutions LBP operator and Uniform patterns for texture representation. They also applied a PCA transformation to reduce the dimensionality of the feature vectors at each region. Then the resulting feature vectors are compared and a match score is produced for each pair of regions using the cosine similarity metric. Then the final similarity score of two faces is defined by taking a classifier fusion approach. The experimental evaluation of this work is performed via three databases of XM2VTS, FERET and LFW dataset with unseen pair-matching paradigms. This work achieved better performance in pose-invariant recognition of faces and performed well using a single feature classifier. But they used LFW-a (the version of LFW-aligned images using a trained commercial system) dataset for supervised setting and LFW funneled and aligned dataset for restricted setting evaluation.

Hua et al. [16] proposed a probabilistic elastic matching method for handling pose variation in real world face recognition. They used part based demonstration by extracting local features with LBP from multi-scale image areas. They trained a Gaussian Mixture Model (GMM) to find out the spacial-appearance distribution of all face images. Then a three layer Gaussian image pyramid is built. They extracted overlapping image patches from each level of image pyramid. They used each mixture component to make a relationship between the effect of the appearance and location terms. Then a connection is built to match features between two images using each mixture component. For classification, they used a SVM classifier which is trained with vector, concatenating the absolute difference vectors of all the corresponding feature pairs. Using this SVM classifier, they easily decided a matched or mismatched image pair. For better accuracy and better model of verification in different pose variation, they also proposed a joint Bayesian adaptation algorithm to adjust the trained GMM. In their method, they did not fully utilize the compressed Fisher vector encoding. But for the most appropriate matching they kept all extracted features. They argued that their experimental results on LFW dataset with restricted protocol and YouTube video face database shows better performance.

Gang Hua et al.[17] presented a part based face representation to enable elastic and partial matching distance metric for face recognition. They used overlapping and densely sampled image patches for extracting N local image descriptors. In part based face representation, they first used the Viola-Jones face detector for face detection. The detected face image is passed into an eye detector for locating left and right eye position. After that, they used a similarity transformation as geometric rectification that places the left and right eyes into canonical positions. They densely partitioned the face image into N overlapping patches. Then they extracted the features from each patch using local image descriptor and finally represented the whole face image. They utilized both elastic and partial matching distance metric to calculate the distance between two face representation. In their work, they designed a distance metric using the fundamental principle of Hausdorff distance. With this distance metric each descriptor is matched against its spatial neighbourhood in the other face. The minimal distance is recorded and stored in a distance list. Finally the minimal distance list is sorted in ascending order and final distance is selected. They also argued that a simple Difference of Gaussian (DoG) filter shows better performance in difficult lighting variations.

Ho et al. [9] focused on handling pose variations between the probe and gallery face images. They presented a method where a given non frontal face image is reconstructed into a frontal view image using Markov Random Fields (MRFs) and a variant of the belief propagation algorithm for image reconstruction. They divided the input face image into a grid of overlapping patches. They also estimated a globally optimal set of local warps to synthesize the patches at the frontal view. Every patch is acquired by aligning it with images from a training database of frontal faces. They used an extension of the Lucas-Kanade algorithm to perform the alignment process efficiently in handling illumination variations. After reconstructing the non-frontal pose images into frontal face image, it can be used for any face recognition technique. In this method they automatically selected facial landmarks or head pose estimation. They did not use any global geometric transformation for face reconstruction. For classifying an input image as frontal or non-frontal they used SIFT descriptors, PCA based Random projections for dimensionality reduction and SVM classification algorithm. They also proposed a generalized classification algorithm for both frontal and non-frontal pose images. They provided their experimental results using the FERET, CMU PIE, and Multi-PIE databases. In their work, they only synthesized the probe image to the frontal pose without considering the other viewing pose.

Sanderson et al. [18] proposed a scalable face matching algorithm. In their work, multi-region probabilistic his-

togram of visual words is used to describe each face. They divided each face image in adjacent regions. These regions are again divided into blocks or patches. Each block overlaps neighbour blocks by 75%. From these blocks, descriptive features are extracted via the 2D Discrete Cosine Transform (DCT). They calculated a normalized distance between the histograms of two faces and compared the distance. In order to reduce the sensitivity of threshold selection, the distance is additionally normalized. They also proposed a fast histogram approximation method which dramatically reduces the computational burden with minimal impact on discrimination performance. Furthermore, the use of multiple regions improves accuracy in most cases, especially when dealing with illumination changes and very low resolution images. They showed experimental results on the recent LFW and FERET dataset. The results demonstrated that the normalized distance can considerably improve the robustness of both multiple and single histogram systems.

2.3. Hybrid Approaches

Hybrid approach is a combination of holistic and feature based approaches. This approach uses both geometric local features and the whole face region to recognize a face.

Kim et al. [19] proposed a hybrid approach based on expert fusion to explicitly model the pose variations. Their hybrid method combined four different systems in a single framework to tolerate pose variations in face images for recognition. Among these, the first method is obtained based on a linear pose transformation using PCA features which are then classified using linear discriminant analysis. The second system concurrently trains linear transformation matrix and the linear discriminant analysis (LDA) system and uses raw image data without the previous PCA feature extraction. The third system applies a non-linear radial basis function as pose transformation which is called Generalized Discriminant Analysis (GDA). The fourth system applies a pose transformation lookup table which stores facial feature generated by rotating 3D face shape. Among these four systems, the first two systems belong to the subcategory of linear pose-tolerant feature extraction; the third system belongs to non-linear pose-tolerant feature extraction (i.e., kernel-based method); and the last system is classified in 2D transformation using a 3D generic face model. Finally, these four systems are merged to make a single classification decisions in Euclidean distance assuming they are mutually independent. They performed their experiments on XM2VTS face dataset. Their system was trained using 170 people among the total 300 people of the dataset and the proposed fused method achieved 70% accuracy with considering 30° rotated faces using single frontal views.

3. Proposed Method

From our background study we have known about some existing methods of face recognition. We have identified some areas of improvement. This section proposes an improved method for face recognition system in pose and size variation.

There are five common steps in face verification system which is shown in Fig. 5. These are: Image acquisition (from a digital image or a video frame), Face detection from the given image, Block size definition, Feature extraction and finally Face verification. Among these, we mainly focus on Face Detection and Block size definition steps to handle pose and size variations.

Our proposed method consists of four steps: face detection and landmark localization, block and sub-block formation, feature extraction and similarity based face verification. The goal of this method is to develop a learning based system that classifies a pair of images to decide similarity or dissimilarity. The details of the process is described in the following subsections.

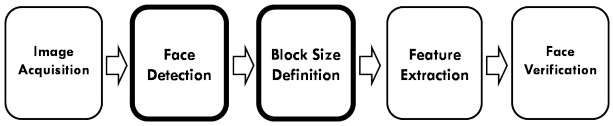


Figure 5: Steps of face verification system

3.1. Face and Landmark Localization

In our proposed system, we firstly perform face detection and landmark localization operation on an input image to dynamically identify most prominent face features. This phase of the proposed method takes an image I as input. It produces face and a vector ν of landmark points as the output of this phase. We adapt the method proposed by Zhu et al. [20] for this localization. The model of Zhu et al. uses mixture of trees with a shared pool of parts V to model facial landmarks and incorporates global mixtures to accommodate topological changes due to viewpoint difference. It uses a pictorial structure with tree $T_m = (V_m, E_m)$, with $V_m \subseteq V$ where m indicating a mixture. With pixel location $l_i = (x_i, y_i)$ of part i, configuration of parts $L = \{l_i : i \in V\}$ is scored by the following equations:

$$S(I, L, m) = App_m(I, L) + Shape_m(L) + \alpha^m \quad (1)$$

$$App_m(I, L) = \sum_{i \epsilon V_m} w_i^m \cdot \phi(I, l_i) \quad (2)$$

$$Shape_m(L) = \sum_{ij \epsilon E_m} a_{ij}^m dx^2 + b_{ij}^m dx + c_{ij}^m dy^2 + d_{ij}^m dy \quad (3)$$

where α^m is a scalar bias associated with viewpoint mixture m, w_i^m is a template for part i, tuned for mixture

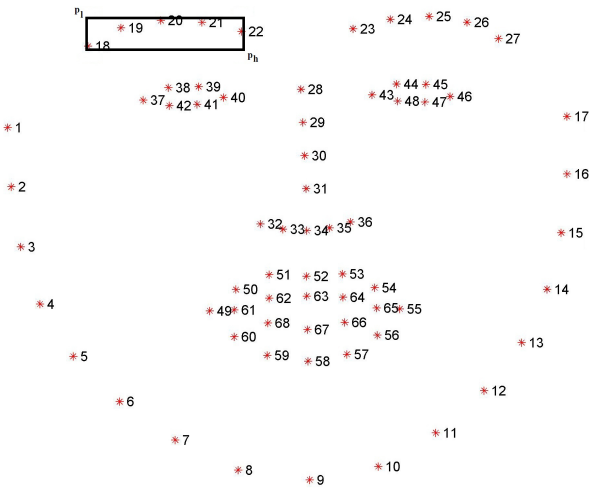
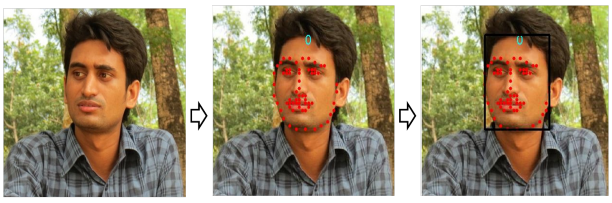


Figure 6: Template image for landmark detection using 68 point for frontal image [21]

m, $\phi(I, l_i)$ is a feature vector extracted from pixel location l_i in image I, $dx = x_i - x_j$ and $dy = y_i - y_j$ are the displacement of the ith part relative to the jth part, and (a, b, c, d) are parameters specifying the rest location and rigidity of each spring.

Based on the aforementioned method we get the landmark points on a face such as eye inner and outer corner, mouth corner, nose corner for different pose variations. These points are grouped to identify face features such as eyes, nose and mouth that are presented as blocks. A typical output of landmark detection for frontal face is shown in Fig. 6. Fig. 7 shows an output image with landmark localization and face detection.



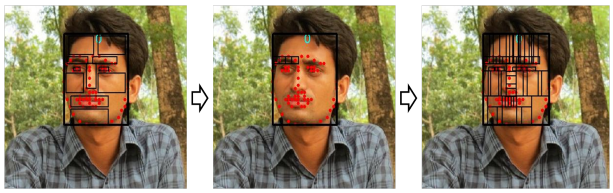
Original Image Landmark Detection Face Detection

Figure 7: Face Detection and Landmark localization of an input image

3.2. Block and Sub-block Formation

After face detection and landmark localization, we perform block operation. For each prominent face feature, we extract the lowest point and highest point of each feature region. Then we formulate a rectangular block from highest to lowest point. In this way, we define rectangular regions of all prominent face features such as mouth, nose, eye, etc.

The vector ν contains j landmark points. We group together these points to construct k face feature groups ν^kcorresponding to eye, nose, mouth, etc. We use these face feature groups ν^k for formulating face feature block B_k. For each group $v_i \in \nu^k$ for $i = 1..k$ we extract the lowest point $p_l = (x_l, y_l)$ and highest point $p_h = (x_h, y_h)$ for formulating a rectangular block from p_l to p_h. Rectangular portion of Fig. 6 depicts the feature points $p \in v_i$ in an feature group v_i , and the lowest and highest coordinates of v_i have been used to form the block B_k, which in this case corresponds to an eyebrow.



Main Face Features Blocks Left Eyebrow Sub-blocks All Sub-blocks

Figure 8: Block and Sub-block formation of an input image

After block formation, we divide each main block into multiple sub-blocks. These main feature blocks B_k are processed again with a sub-block process to generate a set of sub-blocks B'_k. This process considers three pieces of information apart from B_k. These are ratio (%) of block overlapping (ε), number of sub-blocks (η) in the main block, and ratio (%) of external block region (κ) to be appended to construct the sub-blocks. We need to define how much sub-blocks (η) will be created in each main feature block . We also append a ratio of external block region (κ) in each main feature block to avoid the missing part of total feature region. And we also use a small block overlapping (ε) between sub-blocks to avoid the missing part of image feature on the boundary of a sub-block. This process constructs sub-blocks B'_k from B_k with given ε% overlap and given κ% external block region. Fig. 8 shows an image with block and sub-blocks.

3.3. Feature Extraction

After block and sub-block formation, we extract feature's information from each sub-block. For feature extraction, we use dynamic Local Ternary pattern (DLTP)[6] which is based on Human vision constant value. This DLTP can be defined as follows:

$$DLTP(a) = \begin{cases} +1, & a \geq (x_c + t_{HVC}) \\ -1, & a \leq (x_c - t_{HVC}) \\ 0, & otherwise \end{cases} \quad (4)$$

where x_c is the center pixel value, a is a neighbouring gray level value of the center pixel x_c and t_{HVC} is the dynamic threshold for the x_c. We use Algorithm 1 to process blocks B_k of image I produced by steps described in Subsection 3.2 to produce feature vector Γ of length k.

Algorithm 1 Feature Extraction

Input: Blocks $B_i(I_A)$, $B_i(I_B)$ corresponding to I_A and I_B for $i = 1, ..., k$ obtained from Subsection 3.2; $B_i(I_A)$ and $B_i(I_B)$ contains set of sub-blocks $B'_i(I_A)$, $B'_i(I_B)$.

Output: Classification-feature vector Γ

Begin

Step 1. Calculate histograms $H_i(I_A)$, $H_i(I_B)$ for each block $B'_i(I_A)$, $B'_i(I_B)$ respectively using $DLTP$, for $i = 1, ..., n$

Step 2. Calculate the square-root of χ^2 distances between histograms $H_i(I_A)$ and $H_i(I_B)$ $(i = 1, ..., k)$ to obtain classification-feature vector Γ of length k.

End

3.4. Similarity Based Face Verification

After defining sub-blocks we create a sub-block pool. we use our Dynamic Local Ternary Pattern (DLTP) for each sub-block. From each sub-block we find feature's information such as how much lines, how much corners, how much edges, spots, etc. exists in the sub-block which will be displayed as histogram value. Then we perform a square root of chi-square distance for similarity measurement of each block using these features information. Finally we perform a Support Vector Machine (SVM)[10] classification operation.

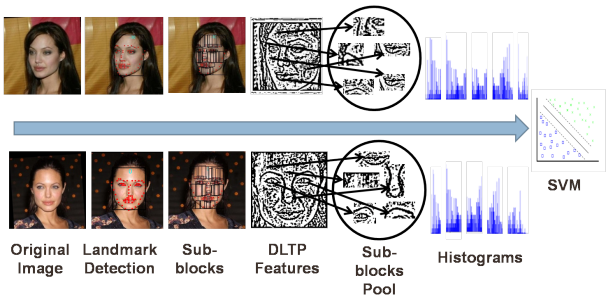


Figure 9: Similarity measurement and classification method for face verification

With a database $\mathcal{T}$ containing m images $P_i = (P_A, P_B)$, $i = 1, ..., m$, we have feature vectors Γ_i for each person $P_i \in \mathcal{T}$ and the respective classification information ϕ_i for $i = 1, ..., m$. This gives us a set ξ containing tuples $\{\Gamma_i, \phi_i\}$ for the matched and mismatched pairs of $\mathcal{T}$. This set ξ is used to train a Support Vector Machine (SVM) for classifying Γ_is in accordance with ϕ_i for $i = 1, ..., m$.

The test data comprises of a pair of images $\{Q_A, Q_B\}$. We applied methods of Subsection 3.1 - Subsection 3.4 on these images to produce classification-feature vector Γ. We use the SVM trained on ξ to classify Γ and produce boolean decision σ describing whether Q_A and Q_B belong to the same person or not. Fig. 9 shows a combined diagram with similarity measurement and classification process.

4. Experimental Results and Discussion

In this section we present a comprehensive experimental evaluation of our proposed method using dataset for studying face recognition in unconstrained environments.

4.1. Labeled Faces in the Wild (LFW)

There are many face databases widely used for face related studies. Among these, Labeled Faces in the Wild or LFW is a more challenging dataset. In this database, images vary in all possible ways due to pose, lighting conditions, resolution, facial expression, illumination, occlusions, clothing, size , background and quality, etc.[22]. This database contains great ethnicity variance. It also contains faces wearing glasses/sunglasses or faces with beards. For this reason, LFW is currently considered as the most challenging and large protocol in the image processing field.

The LFW dataset contains 13,233 face images which are collected from the web. All these images have been labelled manually with the name of the pictured person. Among these images, 1680 people have two or more distinct photos in the dataset. This database organizes its images into two "Views", or "Groups". View 1 and View 2. View 1 is used for developing algorithm and general experimentation, prior to formal process evaluation. This might also be called a model selection or validation view. View 2 is used for performance reporting and final evaluation of a method. The purpose of this methodology is to use the final test sets before evaluation reporting and comparing performance. In testing phase, each test image set is used only once.

There are two ways to use training dataset images. These are: *Image-Restricted Training* and *Unrestricted Training*. The actual state of affairs behind the image restricted setting is that the experimenter should not use the name of an individual to infer the equivalence or non-equivalence of two face pictures that do not seem to be expressly given within the training set. Under the Image-Restricted training setting, the experimenter ought to discard the particular names related to a combination of training pictures, and retain solely the knowledge regarding whether or not a combination of pictures is matched or mismatched. The concept behind the Unrestricted training setting is that one may form as several pairs of matched and mismatched pairs as desired from a collection of pictures labelled with people names. To build our results comparable with other established methods, we tend to additionally use the Image-Restricted training setting in our experiments [23].

4.2. Performance Analysis

We have evaluated our proposed algorithm on the LFW dataset and compared the results with previous approaches. There are three evaluation protocols on this database: the image unrestricted setting, the image restricted setting and the unsupervised setting. Images are categorized into three types: original, funneled, and LFW-a where original means images without any change, funneled means aligned images and LFW-a means the version of LFW-aligned images using a trained commercial system. We have evaluated our proposed method on the image restricted settings with original image.

4.2.1. Performance using Uniform Block Size

First, we have performed a verification operation on LFW dataset images using different size and number of uniform blocks. In this case, we observed that a prominent feature may divide into more than one block. Although we need to divide each prominent features into the same number of blocks in all images where system's comparison will be more accurate.

Table 1 shows a verification result on LFW dataset using uniform blocks. First, we have divided each face image only in one uniform block and found accuracy 47.23%. When we increased the number of uniform blocks then the accuracy also increased. After a certain number of blocks the accuracy is decreasing. In this way, we have found best accuracy 71.80% in 63 number of uniform blocks. We have found best accuracy in this case, because block-by-block alignment and matching operations fit appropriately for LFW dataset images. Fig. 10 shows the graphical representation of Table 1.

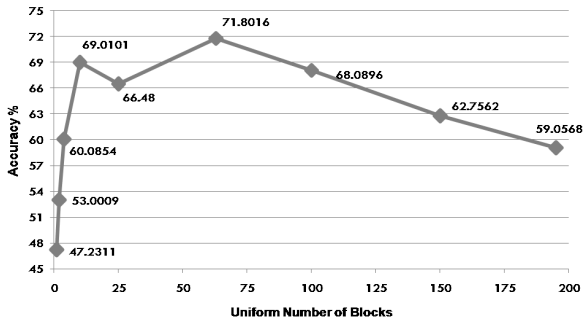


Figure 10: Results of using different size and number of uniform block

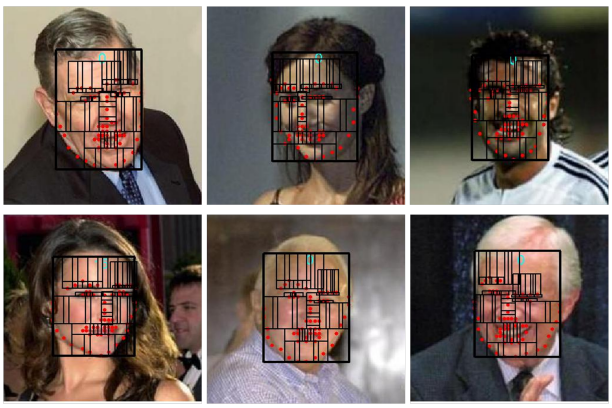


Figure 11: Some dataset images with dynamic size blocks

Our main goal in this research is to find out the performance of our proposed dynamic size blocking mechanism in the real life images. We have divided a face region into 11 main feature block such as Left Forehead, Right Forehead, Left Eyebrow, Right Eyebrow, Left Eye, Right Eye, Nose, Left Cheek, Right Cheek, Mouth, and Chin. Then we have divided these block into multiple sub-blocks according to sub-block formation. Fig. 11 describes some dataset images with DSB.

4.2.2. Performance using Dynamic Size Blocks

We have experimented our proposed method with 1-5 sub-blocks in each face feature region and calculated the accuracy. Table 2 describes the result of this process. It shows that when the number of sub-blocks are small the accuracy is less. For example, when the sub-block number is one the accuracy is 26.23%. However, the increasing number of sub-blocks also increased the accuracy and it decreased again when the number goes high. In this way, we have found the best accuracy when the number of sub-blocks are three and the performance is 74.08%. It is noteworthy to mention here that as long as we are comparing two faces based on comparing their histograms, we have to use fixed number of sub-blocks. We fixed three sub-blocks in our experiments as it gives higher accuracy. This happen because, using more or less number of sub-blocks the less facial region find similarity to compare the simple pattern frequency such as number of line, corners, edges and spots etc. in each block and varied more dramatically in different images. Furthermore, usually a histogram handles image alignment problem to some extent. Even though we used landmark to define block with increased number of sub-block, alignment problem contribute to fall the accuracy. In our experiment we use small block overlapping between sub-blocks to avoid the missing part of DLTP features on the boundary of a sub-block. As mentioned in section 3.2, we append a ratio of external block region in each main feature block to avoid the missing part of total feature region.

We have also done another experiment for avoiding images with extreme pose variation of the LFW dataset. These images are also difficult for human to recognize them manually. From the empirical analysis we have found a very small amount (around 5%) of extreme pose variant images in LFW dataset. Avoiding these images, we have achieved a very good result shown in Table 2. Fig. 12 is the graphical representation of Table 2.

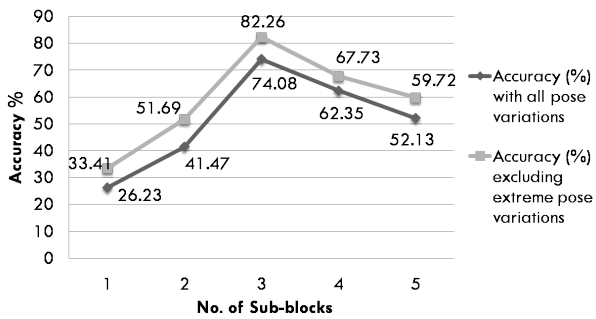


Figure 12: Results of using dynamic size blocks

We have considered the results of the state-of-the-art

Table 1: Results of Uniform block size

Number of uniform blocks	Size of uniform blocks (height×Weight)	Accuracy (%)
1	150 × 130	47.2311
2	150 × 65	53.0009
4	75 × 65	60.0854
10	75 × 26	69.0101
25	30 × 26	66.4888
63	18 × 23	71.8016
100	15 × 13	68.0896
150	10 × 13	62.7562
195	15 × 13	59.0568

Table 2: Results of DSB for different number of sub-blocks in each main block

All pose variations		Excluding extreme pose variations	
No. of Sub-blocks	Accuracy (%)	No. of Sub-blocks	Accuracy (%)
1	26.23	1	33.41
2	41.47	2	51.69
3	74.08	3	82.26
4	62.35	4	67.73
5	52.13	5	59.72

methods that are illustrated in LFW website on restricted image setting. Their results along with our proposed method are represented in Table 3. These results show that the incorporation of Dynamic size blocks mechanism based on face detection with landmark localization, significantly improves the result of face verification even with a single feature extraction method. Fig. 13 shows the graphical representation of the comparative analysis with state-of-the-art accuracy.

5. Conclusion and Future Research

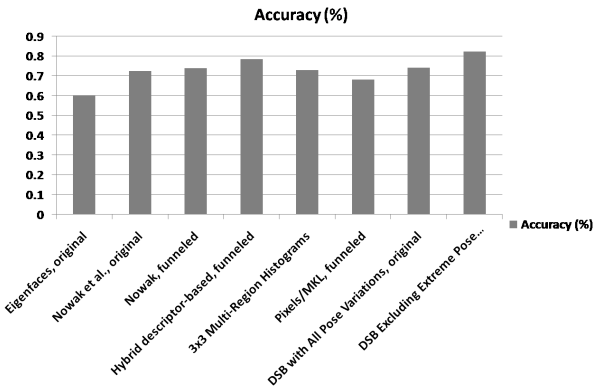


Figure 13: Comparative analysis with state-of-the-art accuracy

We have considered the problem of recognizing human faces despite of variations in pose and size using real life uncontrolled images. Our main contribution is on Dynamic Size Blocks (DSB) definition based on landmark detection. Instead of uniform block, we have proposed DSB considering most prominent face features such as eye, eyebrow, nose, mouth, chin, cheek, fore-head, etc. Our proposed method consists of four parts. Firstly, we perform face detection and landmark localization on an input image to dynamically identify most prominent face features. We perform this work using mixture of trees with a shared pool of parts based on the model of Zhu et al. [20]. Next, we perform block operation and define rectangular block in all prominent face features region such as mouth, nose, eye, etc. After block formation, we divide each main block into multiple sub-blocks using the parameters of sub-block formation. Then, we use a Dynamic Local Ternary Pattern (DLTP) for extracting facial features information from each sub-block of a main block. We perform a square-root of Chi-Square distance for similarity measurement to find out classification feature vector corresponding two blocks of the two images. Finally, we use a Support Vector Machine (SVM) classifier for face verification. This dynamic size blocking method provides promising performance in the real life challenging images on Labeled Faces in the Wild (LFW).

Further research on this work may include advanced image descriptors and classifiers for handling more complex pose variations such as fully side-view images. We can also use this Dynamic Size Blocking method in different texture based recognition and classification problems such as content based object detection, defect detection, medical image diagnosis and facial expression recognition.

Table 3: Some State-of-the-art accuracy on LFW restricted images

Approach/Method	Accuracy(%) or Ratio
Eigenfaces, original [24]	0.6002 ± 0.0079
Nowak et al.,original [25]	0.7245 ± 0.0040
Nowak, funneled [26]	0.7393 ± 0.0049
Hybrid Descriptor-based, funneled [27]	0.7847 ± 0.0051
3x3 Multi-Region Histograms (1024) [18]	0.7295 ± 0.0055
Pixels/MKL, funneled [28]	0.6822 ± 0.0041
DSB with All Pose Variations, original	0.7408 ± 0.0041
DSB Excluding Extreme Pose Variations, original	0.8226 ± 0.0054

References

[1] C. Aiping, P. Lian, T. Yaobin, N. Ning, Face detection technology based on skin color segmentation and template matching, in: Education Technology and Computer Science (ETCS), 2010 Second International Workshop on, Vol. 2, IEEE, 2010, pp. 708–711.

[2] J. Ruan, J. Yin, Face detection based on facial features and linear support vector machines, in: Communication Software and Networks, 2009. ICCSN'09. International Conference on, IEEE, 2009, pp. 371–375.

[3] P. Viola, M. Jones, Robust real-time face detection, International journal of computer vision 57 (2) (2004) 137–154.

[4] T. Ojala, M. Pietikainen, T. Maenpaa, Multiresolution gray-scale and rotation invariant texture classification with local binary patterns, Pattern Analysis and Machine Intelligence, IEEE Transactions on 24 (7) (2002) 971–987.

[5] X. Tan, B. Triggs, Enhanced local texture feature sets for face recognition under difficult lighting conditions, Image Processing, IEEE Transactions on 19 (6) (2010) 1635–1650.

[6] M. Ibrahim, M. I. A. Efat, H. K. Shamol, S. M. Khaled, M. Shoyaib, M. Abdullah-Al-Wadud, Dynamic local ternary pattern for face recognition and verification, International Conference on Computer Engineering and Applications (8).

[7] B. S. Manjunath, W.-Y. Ma, Texture features for browsing and retrieval of image data, Pattern Analysis and Machine Intelligence, IEEE Transactions on 18 (8) (1996) 837–842.

[8] M. Ibrahim, M. Efat, I. Alam, S. M. Khaled, M. Shoyaib, Face verification with fully dynamic size blocks based on landmark detection, in: Informatics, Electronics & Vision (ICIEV), 2014 International Conference on, IEEE, 2014, pp. 1–5.

[9] H. T. Ho, R. Chellappa, Pose-invariant face recognition using markov random fields, Image Processing, IEEE Transactions on 22 (4) (2013) 1573–1584.

[10] C.-C. Chang, C.-J. Lin, Libsvm: a library for support vector machines, ACM Transactions on Intelligent Systems and Technology (TIST) 2 (3) (2011) 27.

[11] Z. Cao, Q. Yin, X. Tang, J. Sun, Face recognition with learning-based descriptor, in: Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on, IEEE, 2010, pp. 2707–2714.

[12] R. Brunelli, T. Poggio, Face recognition: Features versus templates, IEEE transactions on pattern analysis and machine intelligence 15 (10) (1993) 1042–1052.

[13] S. R. Arashloo, J. Kittler, Efficient processing of mrfs for unconstrained-pose face recognition, in: Biometrics: Theory, Applications and Systems (BTAS), 2013 IEEE Sixth International Conference on, IEEE, 2013, pp. 1–8.

[14] N. Komodakis, N. Paragios, G. Tziritas, Mrf energy minimization and beyond via dual decomposition, Pattern Analysis and Machine Intelligence, IEEE Transactions on 33 (3) (2011) 531–552.

[15] B. Gidas, A renormalization group approach to image processing problems, Pattern Analysis and Machine Intelligence, IEEE Transactions on 11 (2) (1989) 164–180.

[16] H. Li, G. Hua, Z. Lin, J. Brandt, J. Yang, Probabilistic elastic matching for pose variant face verification, in: Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on, IEEE, 2013, pp. 3499–3506.

[17] G. Hua, A. Akbarzadeh, A robust elastic and partial matching metric for face recognition, in: Computer Vision, 2009 IEEE 12th International Conference on, IEEE, 2009, pp. 2082–2089.

[18] C. Sanderson, B. Lovell, Multi-region probabilistic histograms for robust and scalable identity inference, Advances in Biometrics (2009) 199–208.

[19] T.-K. Kim, J. Kittler, Design and fusion of pose-invariant face-identification experts, Circuits and Systems for Video Technology, IEEE Transactions on 16 (9) (2006) 1096–1106.

[20] X. Zhu, D. Ramanan, Face detection, pose estimation, and landmark localization in the wild, in: Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on, IEEE, 2012, pp. 2879–2886.

[21] http://ibug.doc.ic.ac.uk/media/uploads/images/300 w/figure_1_68.jpg. [link]. URL `http://ibug.doc.ic.ac.uk/media/uploads/images/300-w/figure_1_68.jpg`

[22] G. Huang, M. Mattar, T. Berg, E. Learned-Miller, Labeled faces in the wild: A database forstudying face recognition in unconstrained environments, in: Workshop on Faces in'Real-Life'Images: Detection, Alignment, and Recognition, 2008.

[23] S. Shan, Y. Chang, W. Gao, B. Cao, P. Yang, Curse of mis-alignment in face recognition: Problem and a novel mis-alignment learning solution, in: Automatic Face and Gesture Recognition, 2004. Proceedings. Sixth IEEE International Conference on, IEEE, 2004, pp. 314–320.

[24] M. Turk, A. Pentland, Face recognition using eigenfaces, in: Computer Vision and Pattern Recognition, 1991. Proceedings CVPR'91., IEEE Computer Society Conference on, IEEE, 1991, pp. 586–591.

[25] E. Nowak, F. Jurie, Learning visual similarity measures for comparing never seen objects, in: Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on, IEEE, 2007, pp. 1–8.

[26] G. Huang, V. Jain, E. Learned-Miller, Unsupervised joint alignment of complex images, in: Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on, IEEE, 2007, pp. 1–8.

[27] L. Wolf, T. Hassner, Y. Taigman, Descriptor based methods in the wild, in: Workshop on Faces in'Real-Life'Images: Detection, Alignment, and Recognition, 2008.

[28] N. Pinto, J. DiCarlo, D. Cox, How far can you get with a modern face recognition test set using only simple features?, in: Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on, IEEE, 2009, pp. 2591–2598.