

## Local-Tetra-Patterns for Face Recognition Encoded on Spatial Pyramid Matching

Khuram Nawaz Khayam<sup>1</sup>, Zahid Mehmood<sup>2,\*</sup>, Hassan Nazeer Chaudhry<sup>3</sup>, Muhammad Usman Ashraf<sup>4</sup>, Usman Tariq<sup>5</sup>, Mohammed Nawaf Altouri<sup>6</sup> and Khalid Alsubhi<sup>7</sup>

<sup>1</sup>Department of Software Engineering, University of Engineering and Technology, Taxila, 47050, Pakistan

<sup>2</sup>Department of Computer Engineering, University of Engineering and Technology, Taxila, 47050, Pakistan

<sup>3</sup>Department of Electronics, Information and Bioengineering, Politecnico di Milano, Milano, 20122, Italy

<sup>4</sup>Department of Computer Science, University of Management and Technology, Sialkot, Pakistan

<sup>5</sup>College of Computer Engineering and Science, Prince Sattam bin Abdulaziz University, Alkharj, 16278, Saudi Arabia

<sup>6</sup>Department of Computer Science, Imam Mohammad Ibn Saud Islamic University, Riyadh, 13318, Saudi Arabia

<sup>7</sup>Department of Computer Science, King Abdulaziz University, Jeddah, 21589, Saudi Arabia

\*Corresponding Author: Zahid Mehmood. Email: zahid.mehmood@uettaxila.edu.pk

Received: 04 May 2021; Accepted: 03 August 2021

**Abstract:** Face recognition is a big challenge in the research field with a lot of problems like misalignment, illumination changes, pose variations, occlusion, and expressions. Providing a single solution to solve all these problems at a time is a challenging task. We have put some effort to provide a solution to solving all these issues by introducing a face recognition model based on local tetra patterns and spatial pyramid matching. The technique is based on a procedure where the input image is passed through an algorithm that extracts local features by using spatial pyramid matching and max-pooling. Finally, the input image is recognized using a robust kernel representation method using extracted features. The qualitative and quantitative analysis of the proposed method is carried on benchmark image datasets. Experimental results showed that the proposed method performs better in terms of standard performance evaluation parameters as compared to state-of-the-art methods on AR, ORL, LFW, and FERET face recognition datasets.

**Keywords:** Face recognition; local tetra patterns; spatial pyramid matching; robust kernel representation; max-pooling

### 1 Introduction

The face recognition field always stayed an active topic in the computer vision research area. Researchers have introduced many techniques that were used to increase the recognition accuracy with lesser time and lesser processing cost [1]. More specifically, during recent years, facial recognition and other biometric verification systems have developed greatly. In the current era, the dependency of protection on a single authorization method is out of the question [2]. Since the cyber threats and attacks are also getting stronger, the security methods can be improved with the provision of introducing multiple factor authorization. Facial recognition is a very promising



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

means of improved security and immaculate access. This also provides a smart mechanism for authentication to personal systems because the other methods such as passwords, PINs, etc. are difficult to remember, are prone to forgery attacks, and require individual participation. Thus, facial recognition is still superior to other methods in such scenarios. The facial recognition system is also being used in the areas such as Access Control, closed-circuit television (CCTV) observation, Persistent Computing, etc. These systems are under consistent development with the target of achieving more robustness and accuracy to provide better security for privacy and assets.

Face recognition systems encounter many challenges including face misalignment, illumination changes, rotational changes, pose changes, covered faces, and blurred faces [3]. Moreover, factors such as aging, expressions, and face changes due to the mustache, beards, and hairstyle, etc. also present problems in face recognition. Proposing a robust algorithm that handles all aforementioned problems is of chief concern in the latest researches. Various techniques have been proposed for this purpose, which can broadly be categorized into two approaches: ‘computer-vision-based approaches’ and ‘deep-learning-based approaches.’

The deep-learning-based approach for face recognition is a relatively new but fast-growing area [3]. Deep learning models can extract features themselves and learn to classify them based on extracted features. In recent years, several advancements have been made in deep learning techniques for face recognition. Although a tremendous amount of research has been done in deep learning and increasing a lot of accuracy in recent years. Yet, there are some problems in the implementation of these approaches [4]. Firstly, high computational power and training time is required for the implementation of such techniques [5]. Moreover, a huge dataset is required to achieve significant results. If a small to medium dataset is used for such techniques, there is a high chance of overfitting. Furthermore, the complexity of developing deep learning models is high. These approaches are also reported to face problems when the number of classes increases [6]. Outstanding face recognition techniques of many computer-vision-based recent studies [7,8] prove their proficiency in various challenging situations such as covered faces and pose variations, etc. However, the performance of these methods can further be improved through investigation.

In most of the existing techniques of the computer-vision-based approaches, face recognition is performed in three steps: pre-processing, selection of features, and classification [9]. The pre-processing step involves face detection, alignment, rotation, correction, scaling, and noise removal. The feature extraction process is concerned with extracting discriminatory facial features that play important role in the unique identification of faces. Finally, the extracted features are classified among database image features. Yet it is more interesting to perform robust face recognition without performing any preprocessing. Our research study is aimed to improve the existing technique proposed by Yang et al. [7], in which Local Binary Pattern (LBP) was used followed by multipartition max-pooling and Robust Kernel Representation (RKR).

We propose an efficient feature extraction technique using Spatial Pyramid Matching (SPM) and second-order Local Tetra Pattern (LTrP) features. This method suffices in handling all the aforementioned problems like pose variation, illumination changes, and covered faces, etc., rather than using extra preprocessing techniques. In our proposed methodology, firstly a pattern map of the input image is created using LTrP, then the image partitioning is performed using SPM, then the histograms of the resultant image are created, afterward, max-pooling is executed on the obtained histograms to create the final feature vector, the last step is the classification based on obtained feature vector. Experimental results on four benchmark datasets show the superiority of our approach over other state-of-the-art techniques. Our research contributes to the literature regarding face recognition by presenting a technique utilizing SPM and second-order

LTrP. A combination of these techniques provides better results as compared to various existing computer-vision-based face recognition techniques.

The rest of the paper is organized as follows: Section 2 outlines a review of the latest techniques relevant to the proposed approach. In Section 3, the detail of the proposed methodology is presented. Experimental results are elucidated in Section 4. Lastly, Section 5 concludes the proposed method and presents future directions.

## 2 Related Work

Face recognition is a vast research field where researchers have put a lot of effort to form a better and smarter World. It has been researched a lot to bring a lot of techniques that have a high impact in this field. In this section, we have provided a review of existing techniques implemented for feature extraction, dimension reduction, and classification. LBP is a local face descriptor and can be used greatly in most of the image's feature extraction phase [10]. The concept of LBP was first introduced by Ojala where a  $3 \times 3$  neighborhood of image pixels was considered while the center pixel of the image was compared with each pixel as a threshold [11]. The value of the center pixel is the decimal equivalent of the resulting binary number. In general, LBP is calculated by comparing the surrounding pixels from the referenced pixels and encoding the resulting bits 1 and 0 based upon the compared result. A binary pattern is obtained by setting the output matrix's pixel to 1 for those whose pixel value is greater than the center and the rest of the output values are set to 0. Then, a binary code is calculated by concatenating the numbers clockwise. The final LBP feature can be obtained by concatenating all the binary or decimal codes in a resultant matrix. Several variants of LBP have been proposed by researchers. Truong et al. [12] proposed a variant of LBP called 'weighted statistical binary pattern by direction.' In this technique, the descriptors used straight-line topologies along with different directions. By dividing the input image into mean and variance moments, and subsequently computing weighted histograms of the sign and amplitude components, robust facial feature representation was achieved. Local Derivative Pattern (LDP) is a variant of LBP, in which directional pattern features are encoded based on variations in the local derivative. It was first employed by [13] for the problem of face recognition. It works by extracting higher-order local information through the encoding of individual spatial relationships encompassed in the local region. After detailed experimentation, [14] concluded that LDP exhibited robustness under all noisy conditions, and it performed better with different illumination conditions and rotation angles as compared to LBP and its six other variants.

Local Ternary Patterns (LTP) were developed by Tan et al. [15]. LTP is an extended form of LBP. In LTP, a three-pair code is extended from LBP and a sign function  $s(x)$  is replaced by a three-pair function. In the resultant values, there are three types of values: positive, negative, and zero. The ternary code formed by this formation is calculated ideally from the top-left cell clockwise. As an example, it can form a code like 01(-1)011(-1)0. This matrix is further split up into two matrices separated by positive and negative values. After this, the individual histograms of both these matrices are calculated and then the final feature descriptor is obtained by concatenating all these histograms. The higher-order LTrP was first introduced by Murala et al. [16]. It is adopted by the idea of other local patterns, including LBP, LDP [13], and LTP [15]. The general use of LTrP is for texture analysis due to its improved ability to extract definite information from a particular image. It is calculated by taking the surrounding pixels from the referenced pixel by taking their first-order derivative in a horizontal and vertical direction. LTrP is very effective for content-based images. It is mostly used in Content-Based Image Retrieval (CBIR). It was found

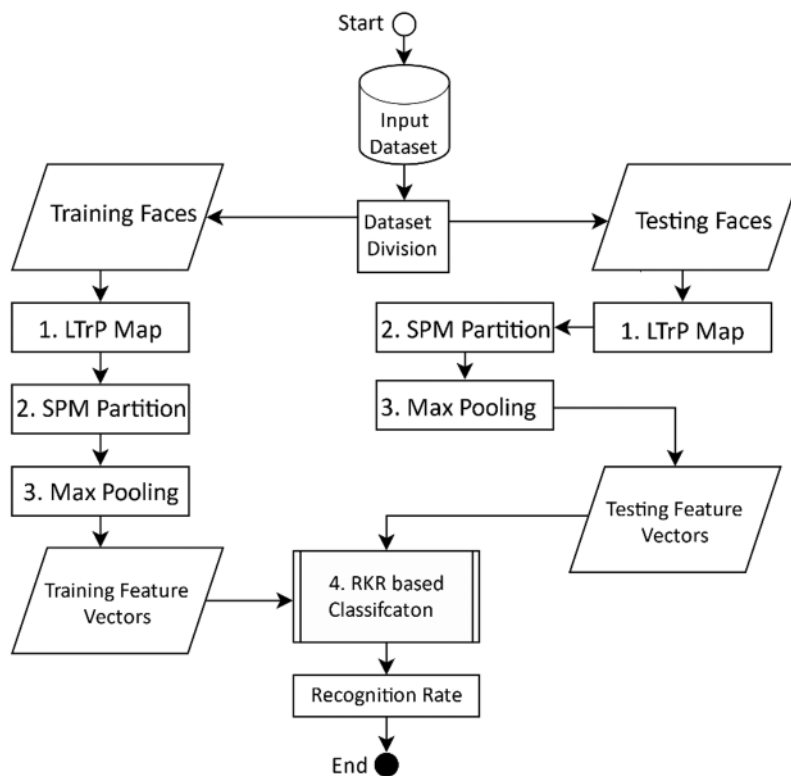
that the results obtained by second-order LTrP are better than higher-order LTrP because the sensitivity to noise increases as the order gets increased [17]. Mehmood et al. [18] presented a novel image representation technique based on the Bag of Visual Words (BoVW) model in which an image is partitioned into two rectangular regions and then histograms of those regions are calculated. During the construction of histograms, this spatial information is stored in the BoVW model. This representation of an image can be used as a feature in image classification and so it can be used for face recognition. For dimension reduction, pooling techniques are being used, such as sum pooling [19,20] and max-pooling [20–22]. Spatial pooling has been adopted widely in image classification for the extraction of invariant features meanwhile reducing dimensionality and processing cost. In face recognition, various researchers adopt pooling techniques as a way to enhance feature encoding [23,24]. Most used pooling methods include sum pooling, average pooling, and max-pooling. The maximum response is preserved using max-pooling whereas the average response is preserved by using average pooling. Many researchers proved the superiority of max-pooling over other pooling techniques for the face recognition problem. This preeminence of max-pooling is probably due to its robustness to local spatial variations [22,25]. The performance of each pooling method is a lot dependent on the block division. In this respect, an effective approach is a variable-sized multi-partitioning scheme [7].

Various classification techniques are proposed by researchers as the final step of face recognition. Most popular are the three approaches including Collaborative Representation-based Classification (CRC), Sparse Representation-based Classification (SRC), and kernel representation-based techniques. Both SRC and CRC are termed dictionary learning methods because the test samples are represented using a dictionary. In SRC, the training images are coded in a sparse matrix such that only the required elements or atoms of the images are taken, and the rest of the atoms (usually non-important) are discarded. Then the chosen elements play a much important role in the discrimination of test images from training images [26]. In CRC based technique, training images are coded collaboratively that represent the complete training set [27]. The classification in this technique is done by comparing the query image with the training set having minimal distance. The representation coefficient in this technique is generally obtained through  $l_2$  regularization. Wang et al. [28] presented a comprehensive review of the facial feature extraction methods in which local features and discriminative representation was brought together. The review enlightened several common practices and motives to the facial recognition framework interchangings with exemplifications. Song et al. [8] used block-weighted LBP along with CRC for face recognition and achieved better results as compared to the simple CRC technique and variants of the SRC technique. Ma et al. [29] proposed a robust face classification technique using a spatial pyramid structure with weights. The traditional spatial pyramid structure evaluates each partition of the image equally which is less effective. This model adds weight to each partition based upon its self-adaptive method. The algorithm is robust against misalignments, pose variations, and expressions. The spatial pyramid matching technique is not robust against rotated images. In other words, it is not rotation invariant. Karmakar et al. [30] proposed a modified spatial pyramid matching technique that is robust against rotated images. The model is proposed using a weight function that plays important role in making the technique rotation invariant. The weighted spatial pyramid technique was proposed by Choi [31] that is based on the division and subdivision of multiple finer grain partitions on each level of the pyramid. After that, on each pyramid level, the calculated sum of each different partition in the partition set is used for recognition. The weights of the spatial pyramid are determined on each pyramid level using the discriminative power of the feature class. Kernel representation based classifiers provide a better way to classify the non-linearly separable features. It maps the features into a high dimensional

feature space and then the linear classifiers can do a better performance to classify. Yang et al. [7] presented an RKR method that performed robust face recognition using statistical local features. In that technique, multi-partition max-pooling was used to extract the local features and the RKR method was used to exploit maximum discrimination information stored in local features. Occluded faces were handled by robust regression.

### 3 Proposed Methodology

The face recognition process in our methodology is divided into two phases. The first phase is feature extraction and formation. Initially, the input image undergoes the LTrP operation. Afterward, the SPM is used to convert the individual image into a different number of partitions. Finally, the histogram for each block is determined and the max-pooling method is applied over histogram features to produce a final feature vector. In the second phase of the methodology, RKR based classifier is employed for feature classification. This classification step identifies the class of the input image based on the analysis of extracted features. Results are compiled by the percentage of the correctly identified number of test images which is known as recognition accuracy. The individual sub-steps of the proposed methodology are discussed in the next subsections. Fig. 1 shows the basic methodological demonstration of the face recognition process.



**Figure 1:** Proposed model illustration with four steps

### 3.1 Local Tetra Patterns Formulation

This is the first step of the feature extraction phase, where a pattern map of LTrP of the input image is calculated. The mathematical implementation is described below in detail. The basis for the development of LTrP is the preliminary local pattern which are LBP [32], LDP [33], and LTP [34]. LTrP shows a well-formed spatial structure of basic patterns taken from either textures or face data (as in our case). This spatial structure is derived from the direction of the central pixel. Consider an image  $I$ , the derivative operation of first-order on the directions of 0 and 90 degrees are denoted as  $I_{\theta}^1(g_p)|_{\theta=0^{\circ}, \theta=90^{\circ}}$ . Let us suppose that  $g_c$  is the value of a central pixel in image  $I$ , and  $g_v, g_h$  denote neighboring pixels in vertical and horizontal position relative to  $g_c$  respectively. The first-order derivative concerning central pixel  $g_c$  can be expressed as:

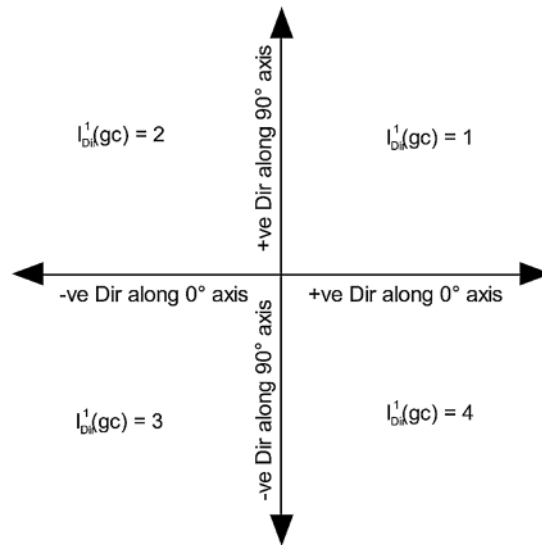
$$I_{0^{\circ}}^1(g_c) = I(g_h) - I(g_c) \quad (1)$$

$$I_{90^{\circ}}^1(g_c) = I(g_v) - I(g_c) \quad (2)$$

The central pixel's direction calculation can be performed by

$$I_{Dir}^1(g_c) = \begin{cases} 1 & I_{0^{\circ}}^1(g_c) \geq 0 \text{ and } I_{90^{\circ}}^1(g_c) \geq 0 \\ 2 & I_{0^{\circ}}^1(g_c) < 0 \text{ and } I_{90^{\circ}}^1(g_c) \geq 0 \\ 3 & I_{0^{\circ}}^1(g_c) < 0 \text{ and } I_{90^{\circ}}^1(g_c) < 0 \\ 4 & I_{0^{\circ}}^1(g_c) \geq 0 \text{ and } I_{90^{\circ}}^1(g_c) < 0 \end{cases} \quad (3)$$

Eq. (3) assures that the central pixel has the possibility of four values (1, 2, 3, and 4). Thus, each pixel of the image gets converted to four directions (based on values). Depending on the first-order derivative values along the x and y-axis, the output value of the image's pixels is selected. This process of output value selection is illustrated in Fig. 2.



**Figure 2:** Four directions of LTrP as 1, 2, 3 and 4

The second-order LTrP is calculated by utilizing the values obtained from Eq. (3). It can be expressed as:

$$LTrP^2(g_c) = f_3(I_{Dir}^1(g_c), I_{Dir}^1(g_1)), f_3(I_{Dir}^1(g_c), I_{Dir}^1(g_2)), \dots, f_3(I_{Dir}^1(g_c), I_{Dir}^1(g_p)) \tag{4}$$

where p is the pixel index (ranges from 1 to 8 in this case). And the function  $f_3$  is defined in Eq. (5).

$$f_3(I_{Dir}^1(g_c), I_{Dir}^1(g_p)) = \begin{cases} 0, & I_{Dir}^1(g_c) = I_{Dir}^1(g_p) \\ I_{Dir}^1(g_p), & \text{else} \end{cases} \tag{5}$$

Using Eqs. (4) and (5), 8-bit tetra patterns are obtained for each central pixel. Since there are four values for four directions (1, 2, 3, 4) so four groups are formed in which each respective direction's tetra pattern is stored. These tetra patterns are then converted into binary patterns (three patterns for each direction). For a case, if the value of the central pixel  $I_{Dir}^1(g_p)$  is 1, then the second order LTrP can be evaluated by setting it apart into three binary patterns. The output of the LTrP feature is extracted using Eq. (6).

$$LTrP^2|_{dir=2,3,4} = \sum_{p=1}^P 2^{p-1} \times f_4(LTrP^2(g_c))|_{dir=2,3,4} \tag{6}$$

$$f_4(LTrP^2|(g_c))|_{dir=\phi} = \begin{cases} 1 & LTrP^2(g_c) = \phi \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

where  $\phi = 2, 3, 4$ . An illustration of the LTrP pattern over a face image can be visualized in Fig. 3.



**Figure 3:** An illustration of the LTrP map

### 3.2 Spatial Pyramid Matching

In the previous step, we calculated the pattern map of LTrP, now we will divide this pattern map into multiple partitions based on SPM and then calculate histograms as feature vector sets from each partition. These types of partitions will add advanced features to our proposed technique, making it rotation and pose invariant. This is because partitioning an image into multiple blocks will carry information from different regions of the image making it invariant to rotation and pose changes.

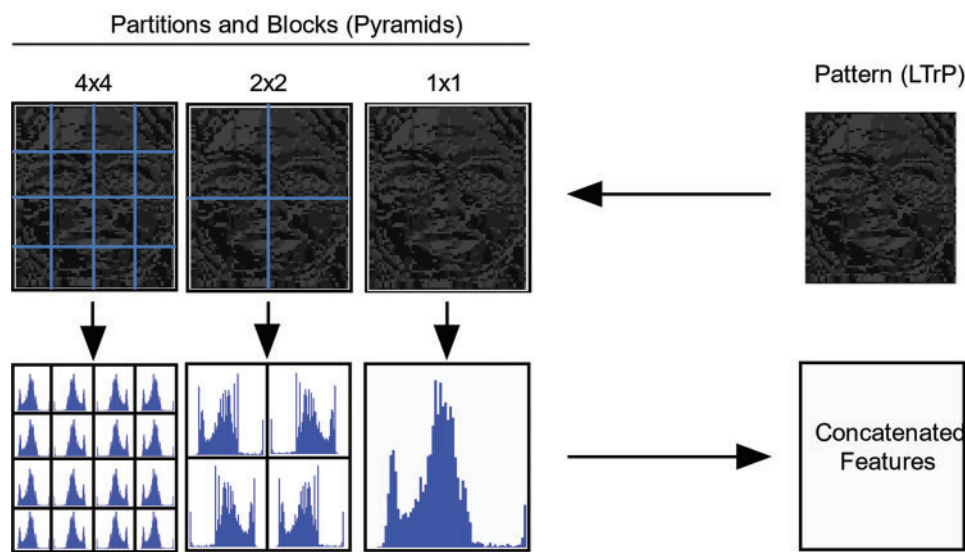
Initially, the Pyramid Matching (PM) kernel was used to perform matching feature collections [35]. This leads to the calculation of an intersection based on weighted histograms in the already existing multi-resolution histograms. But a major drawback of this approach is that it does not consider the spatial information about the images. More precisely, the resultant features lack discriminative spatial information leading to inaccuracy. Thus, SPM was proposed to consider spatial information.

SPM computes the histogram distribution over a diverse spatial resolution while considering images having the same dimensions. SPM kernel computation is performed by performing a matched sum of corresponding values in different channels of a feature. For  $m$  number of channels, the SPM kernel can be represented as:

$$K(X, Y) = \sum_{c=1}^m k(X_c, Y_c) \quad (8)$$

where the function  $k$  represents PM kernel and  $X_c$  and  $Y_c$  are histogram distribution of the feature  $c$  over all the spatial parts.

One of the key benefits of using SPM is that the image's spatial discrimination information can be robustly obtained. It is since SPM divides the image into multi-scale regions in different orders such as  $1 \times 1$ ,  $2 \times 2$ , and  $4 \times 4$  as shown in Fig. 4. These form a total of 21 blocks. We expect that the partition of the image with SPM to form 21 blocks can perform better when combined with LTrP.



**Figure 4:** Illustration of SPM partitions and concatenation of calculated histogram features



A visualization of SPM based on a sample image is demonstrated in Fig. 4. The input image is partitioned as  $1 \times 1$ ,  $2 \times 2$ , and  $4 \times 4$  blocks. After that, features are extracted from each sub-block and concatenated to form a feature vector set.

In every block of each partition, bins are prepared for storing pattern values of that block. This is the process of creating a feature collection in each block. There is a total of 21 blocks, so we get 21 feature vector sets. These features are concatenated together in another matrix for further processing.

In the next Section 3.3, we tried to reduce the dimensions of feature vectors and pick only better and use non-redundant values to enhance discriminative power by using max-pooling.

### 3.3 Feature Vector Formation

In the previous section, the concatenated histogram features contain much spatial information needed for classification it is still in our interest to reduce that information for a better outcome. The first benefit of reducing information is that the redundant information will be removed which does not benefit the output. The second benefit is that better information will make the feature more discriminative. And the last benefit is that the classification process is enhanced due to the smaller size of the feature vector. Suppose we have concatenated feature vectors as  $f_c = \{f_1, f_2, \dots, f_{21}\}$ . These are 21 features from each block. Suppose that each block's feature ranges from 0 to 256. Then the pool of feature  $f_c$  can be seen in Eq. (9):

$$f_c = \begin{bmatrix} f_{1,1} & f_{1,2} & \dots & f_{1,J} \\ f_{2,1} & f_{2,2} & \dots & f_{1,J} \\ \dots & \dots & \dots & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ f_{n,1} & f_{n,2} & \dots & f_{n,J} \end{bmatrix} \quad (9)$$

Here  $n$  is the total number of features concatenated and  $J$  is the total number of indices of each concatenated feature. Then, this pool of feature  $f_c$  can be reduced by picking up the maximum value only from each column i.e., the  $j^{\text{th}}$  column and finally we will get only one-row matrix which will be served as the final feature vector  $f$  in Eq. (10). The final feature  $f$  can be obtained as:

$$\{f\} = \max\{|\{f_n\}_1|, |\{f_n\}_2|, \dots, |\{f_n\}_J|\} \quad (10)$$

We have chosen max-pooling because it performs better dimension reduction as well as keeping better discrimination information [13]. The next Section 3.4 is phase two of our technique which describes the classification method.

### 3.4 Robust Kernel Representation (RKR)

We used the RKR method for classification purposes. This is a much better technique for face recognition with local features [7]. Because the kernel-based techniques can do better mapping for features that cannot be separated nonlinearly, so it can add a lot of discriminative power in features, and then they can be classified better in feature space.

### 3.5 Pseudocode of Proposed Model

The algorithm of the proposed solution is also summarized in two phases. The first phase is feature extraction described in the algorithm in Tab. 1 and the second phase is classification described in Tab. 2. Feature extraction in algorithm Tab. 1 is based on three major steps. The first step is an initialization, in which the input query face image is taken for processing. The second step is calculations that are needed to get features. In calculations, the first LTrP as defined in Section 3.1 is calculated from the input image. LTrP is briefly described in Eq. (6). Then the calculated LTrP is partitioned into 3 partitions and 21 blocks for SPM calculation as defined in Section 3.2. The next step is about processing where each block from each partition is taken and a histogram is calculated for each block. Finally, max-pooling is applied for dimension reduction as defined in Section 3.3 to achieve the final feature vector as output.

**Table 1:** Face extraction procedure of the proposed model

---

**Step 1: Initialization**

Input Image:  $I$  = face image

**Step 2: Calculations**

Calculate LTrP as defined in Eq. (4):  $ltrp = \text{CalculateLtrp}(I)$

Create SPM partitions of 21 blocks:  $partition = \text{SPM\_Blocks}(ltrp)$

Creating feature variable to store feature:  $f = \text{Array}()$

**Step 3: Processing**

**While**  $partitions$  as each  $part$  **do**

    getting blocks from each partition:  $blocks = \text{CreateBlocks}(part)$

    creating  $block\_feature\_pool$  variable to store features:  $block\_feature\_pool = \text{Array}()$

**While**  $blocks$  as each  $block$  **do**

        Calculate Histogram of each block:  $histogram = \text{hist}(block)$

        Concatenate in  $block\_feature\_pool$  variable:  $block\_feature\_pool += histogram$

**end**

    Applying max-pooling over to get final block feature:

$block\_feature = \text{Max\_Pooling}(block\_feature\_pool)$

    Concatenate in  $f$  variable to update final feature vector:  $f += block\_feature$

**end**

output  $f$

---

**Table 2:** Face classification procedure of the proposed model

---

**Step 1: Initialization**

input feature vector calculated in Tab. 1

$f$  = face feature

input face dataset with identities:  $db$  = database

**Step 2: Classification**

Calculate identity of input feature from  $db$  images' features according to [7]:

output  $id$

---

The algorithm in [Tab. 2](#) is about classification where inputs are a feature vector  $f$  that needs to be classified, and a dataset of trained images. Classification is done by RKR [7] also described in Section 3.4. Finally, the identity of the input image is calculated as  $id$ . This identity represents the identity of that person who is recognized by the algorithm.

### 3.6 Datasets

We have tested our technique on four benchmark datasets including Aleix Martinez (AR) [36], Olivetti Research Ltd. (ORL) [37], Labeled Faces in the Wild (LFW) [38], and Face Recognition Technology (FERET) [39] dataset. The AR dataset consists of 120 subjects out of which we selected 100 subjects with 7 clear and non-occluded images of each subject as training data and two types of testing data occluded with glasses and scarf. There are 3 faces of each subject in testing data of both categories. There are also two sessions of each category. The AR dataset provides a challenging task to recognize faces with occlusion. An example of the AR dataset can be seen in [Figs. 5](#) and [6](#).



**Figure 5:** Example of AR test dataset

The ORL dataset consists of 40 subjects with 10 face images in each subject making a total of 400 images. These images were taken at different time instances and consist of expression and illumination variations. We used the cross-validation technique to pick the training data and testing data. Here we also used the same division as 30% testing data and 70% training data. Some samples of the ORL dataset can be seen in [Fig. 7](#).



Figure 6: Example of AR train dataset



Figure 7: Example of ORL dataset

The LFW dataset consists of 13,233 images having 5748 subjects. It is a challenging benchmarking dataset having a lot of images with huge diversity including pose variation, illumination variation, expression variation, makeup, and occlusions. This dataset is very huge and contains a different number of images for each subject. We have chosen a way to select those subjects which have at least 10 samples for proper testing and validation making a total of 157 subjects. After the selection of subjects, we then used the cross-validation train test split technique to split the training and testing data. 30% testing data and 70% training data were used for the LFW dataset experiment. Some samples of the LFW dataset can be seen in [Fig. 8](#).



**Figure 8:** Example of LFW dataset

The FERET dataset comprises 14,126 images which are from 1,199 subjects. It was designed to evaluate the performance of state-of-the-art face recognition algorithms. The dataset has variations in terms of pose and expression. For extensive experimentation, we have taken 200 subjects with a maximum of 7 samples as train data of each subject. In this dataset, we also used the same cross-validation technique to mix the train data and testing data for appropriate results. An example of the FERET dataset can be seen in [Fig. 9](#).

### 3.7 Evaluation Metrics

We have used a standard evaluation metric for the measurement of accuracy. The metric used is recognition rate (RR), which can be formulated as:

$$RR = \frac{I^C}{I^T} \times 100 \quad (11)$$

where  $I^C$  represents the number of testing images correctly identified and  $I^T$  represents the total number of testing images. The recognition rate is a success percentage and can be useful for measurement in percentage.



**Figure 9:** Example of FERET dataset

#### 4 Experimental Results and Discussions

We conducted our experiments in a controlled environment with two datasets AR [36] and ORL [37]. For each dataset, we verified our results on 3, 5, and 7 training samples. A training sample is defined as an image of a unique person from training data. For example, 3 training samples means that 3 pictures of each unique person are used in training. The proposed technique provides much better results compared to the other state-of-the-art methods, including SLC-ADL [40], LRC [41], SRC [42], CRC [27], ESRC [43], TPTSR [44], CLDA [45], LPP [46], CRC-ADL [8], Two-Step LSRC [47], RRC [48], RCR [49], Homotopy [50] and FISTA [51].

##### 4.1 AR Dataset Results

We have achieved competitive results against other state-of-the-art methods with much higher accuracy. In all 3 types of training samples, we achieved better results as compared to other methods. AR dataset results can be seen in Tab. 3.

##### 4.2 ORL Dataset Results

Our method achieved improved results against three training samples on the ORL dataset as compared to other existing techniques. Yet for the 2 and 4 training samples, some methods showed better results as compared to our approach. Tab. 4 shows the results of face recognition on the ORL dataset.

##### 4.3 LFW Dataset Results

Experimentation with the LFW dataset showed better results with our method as compared to existing techniques. The experimentation has been performed over 2, 3, and 4 training samples to

perform the face recognition. [Tab. 5](#) shows the comparison of results of our method with subject to LFW dataset.

**Table 3:** Recognition rate % on AR dataset with different methods

Techniques	No. of training samples		
	3	5	7
SLC-ADL [40]	59.86	63.27	66.59
LRC [41]	54.60	50.87	64.71
SRC [42]	49.68	69.70	73.86
CRC [27]	50.00	51.17	54.48
ESRC [43]	60.43	65.84	68.00
TPTSR [44]	50.69	62.89	67.76
CRC-ADL [8]	64.40	72.26	77.38
<b>Proposed method</b>	<b>83.33</b>	<b>88.33</b>	<b>95.33</b>

**Table 4:** Recognition rate % on ORL dataset with different methods

Techniques	No. of training samples		
	2	3	4
SLC-ADL [40]	71.87	77.14	83.33
LRC [41]	73.75	81.25	83.75
SRC [42]	76.25	80.00	80.00
CRC [27]	76.25	83.75	90.0
ESRC [43]	77.25	79.83	80.23
TPTSR [44]	83.12	88.57	93.75
CLDA [45]	79.47	83.39	86.75
LPP [46]	80.93	87.78	91.33
CRC-ADL [8]	85.00	88.75	93.75
<b>Proposed method</b>	<b>82.50</b>	<b>90.00</b>	<b>90.00</b>

**Table 5:** Recognition rate % on LFW dataset with different methods

Techniques	No. of training samples		
	2	3	4
CRC [27]	11.58	14.03	15.63
LRC [41]	7.87	10.89	13.01
ESRC [43]	14.16	17.23	19.97
TPTSR [44]	14.86	19.59	23.81
SLC-ADL [40]	7.69	10.84	13.82
Two-Step LSRC [47]	11.00	14.50	18.00
CRC-ADL [8]	17.09	21.36	24.53
<b>Proposed method</b>	<b>17.77</b>	<b>23.13</b>	<b>26.25</b>

#### 4.4 FERET Dataset Results

Considering FERET Dataset, extensive experimentation has been performed by taking 3, 4, and 5 samples. The results obtained by our method outperform the previous techniques in terms of recognition rate. Tab. 6 shows the comparison of face recognition results by our method with previous methods for the FERET Dataset.

**Table 6:** Recognition rate % on FERET dataset with different methods

Techniques	No. of training samples		
	3	4	5
CRC [27]	51.75	64.40	66.50
LRC [41]	54.46	68.74	70.33
ESRC [43]	54.13	71.33	76.50
RRC [48]	42.93	53.74	70.21
RCR [49]	45.12	51.02	59.82
TPTSR [44]	56.68	68.34	72.07
SLC-ADL [40]	49.75	68.33	73.75
Homotopy [50]	54.14	72.67	77.45
FISTA [51]	38.90	50.54	58.95
CRC-ADL [8]	74.00	77.00	78.50
<b>Proposed method</b>	<b>74.53</b>	<b>76.36</b>	<b>80.13</b>

#### 4.5 Discussions

In the AR dataset, the result of the proposed model has much higher accuracy as compared to other techniques. Using 3 train samples, we acquired 83.33% accuracy in terms of facial recognition. Using 5 training samples, we acquired 88.33% accuracy and using 7 train samples, we acquired 95.33% accuracy. These results are much better than other state-of-the-art methods in the literature. While in the ORL dataset results, the acquired accuracy could not be improved much higher. Using 2 train samples, it is 82.5% which is less than two other techniques, while using 3 train samples, the accuracy achieved is 90%, which is better than all other techniques. Yet, when using 4 train samples on the ORL dataset, an acquired accuracy of 90% is obtained which is less than 3 other techniques. This is because of some of the major variations of the pose in the dataset. These results can be improved by preprocessing dataset images by aligning them correctly to reduce pose variations. Regarding the LFW dataset, it is challenging since several complexities are added to the problem due to variations in terms of pose, expression, illumination, make-up, and occlusions. Using 2 train samples, we have achieved 17.77% accuracy and for 3 train samples, we acquired 23.13% accuracy. Using 4 train samples, the achieved accuracy is 26.25%. These results are better than other literary techniques. In the color FERET dataset, the results are much better than previous techniques. Using 3 train samples, the accuracy came out to be 74.53% whereas, using 4 train samples, the accuracy is 76.36%. Using 5 train samples, the accuracy is 80.13%. These obtained results are very promising as compared to previous literature techniques and therefore, the proposed method outperforms other state-of-the-art literature methods in terms of accuracy of facial recognition.



## 5 Conclusion and Future Work

In this research work, an efficient feature extraction approach and a classification technique have been utilized for robust face recognition. The process is composed of two main parts including feature extraction and classification. Initially, a feature descriptor LTrP is utilized to extract local patterns. Then, SPM is utilized to partition the patterns into 21 blocks. After that, max-pooling is utilized to construct features with the most discriminative information. Finally, the classification is done using an RKR method to fully exploit the discriminative power for robust performance. Due to the worldwide spread of COVID-19, people wear the mask in workplaces and public places. Hence there is a need for a robust face recognition system that recognize faces with and without mask accompanying with other factors such as illumination changes and poster changes etc. In the future, we would conduct extensive experiments in which eyes, as well as nose, are covered, to formulate a robust face recognition system. We intend to use an extended version of LTrP by modifying it in a way that it collects more powerful and discriminative information as a feature vector from the input image.

**Funding Statement:** This project was funded by the Deanship of Scientific Research (DSR) at King Abdul Aziz University, Jeddah, under Grant No. KEP-10-611-42. The authors, therefore, acknowledge with thanks DSR technical and financial support.

**Conflicts of Interest:** The authors declare that they have no conflicts of interest to report regarding the present study.

## References

- [1] K. Chengeta and S. Viriri, "A survey on facial recognition based on local directional and local binary patterns," in *2018 Conf. on Information Communications Technology and Society (ICTAS)*, Durban, South Africa, pp. 1–6, 2018.
- [2] T. Sabhanayagam, V. P. Venkatesan and K. Senthamaraiannan, "A comprehensive survey on various biometric systems," *International Journal of Applied Engineering Research*, vol. 13, no. 5, pp. 2276–2297, 2018.
- [3] I. Masi, Y. Wu, T. Hassner and P. Natarajan, "Deep face recognition: A survey," in *2018 31st SIBGRAPI Conf. on Graphics, Patterns and Images (SIBGRAPI)*, Parana, Brazil, pp. 471–478, 2018.
- [4] G. Guo and N. Zhang, "A survey on deep learning based face recognition," *Computer Vision and Image Understanding*, vol. 189, pp. 102805, 2019.
- [5] M. Lal, K. Kumar, R. H. Arain, A. Maitlo, S. A. Ruk *et al.*, "Study of face recognition techniques: A survey," *International Journal of Advanced Computer Science and Applications*, vol. 9, no. 6, pp. 42–49, 2018.
- [6] J. M. Pandya, D. Rathod and J. J. Jadav, "A survey of face recognition approach," *International Journal of Engineering Research and Applications (IJERA)*, vol. 3, no. 1, pp. 632–635, 2013.
- [7] M. Yang, L. Zhang, S. C.-K. Shiu and D. Zhang, "Robust kernel representation with statistical local features for face recognition," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 24, no. 6, pp. 900–912, 2013.
- [8] X. Song, Y. Chen, Z.-H. Feng, G. Hu, T. Zhang *et al.*, "Collaborative representation based face classification exploiting block weighted LBP and analysis dictionary learning," *Pattern Recognition*, vol. 88, pp. 127–138, 2019.
- [9] M. Sharif, K. Ayub, D. Sattar, M. Raza and S. Mohsin, "Enhanced and fast face recognition by hashing algorithm," *Journal of Applied Research and Technology*, vol. 10, no. 4, pp. 607–617, 2012.
- [10] Z. Guo, L. Zhang and D. Zhang, "A completed modeling of local binary pattern operator for texture classification," *IEEE Transactions on Image Processing*, vol. 19, no. 6, pp. 1657–1663, 2010.

- [11] T. Ahonen, A. Hadid and M. Pietikainen, "Face description with local binary patterns: Application to face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 12, pp. 2037–2041, 2006.
- [12] H. P. Truong, T. P. Nguyen and Y.-G. Kim, "Weighted statistical binary patterns for facial feature representation," *Applied Intelligence*, vol. 51, pp. 1–20, 2021.
- [13] B. Zhang, Y. Gao, S. Zhao and J. Liu, "Local derivative pattern versus local binary pattern: Face recognition with high-order local pattern descriptor," *IEEE Transactions on Image Processing*, vol. 19, no. 2, pp. 533–544, 2009.
- [14] A. Suruliandi, K. Meena and R. R. Rose, "Local binary pattern and its derivatives for face recognition," *IET Computer Vision*, vol. 6, no. 5, pp. 480–488, 2012.
- [15] X. Tan and B. Triggs, "Enhanced local texture feature sets for face recognition under difficult lighting conditions," *IEEE Transactions on Image Processing*, vol. 19, no. 6, pp. 1635–1650, 2010.
- [16] S. Murala, R. P. Maheshwari and R. Balasubramanian, "Local tetra patterns: A new feature descriptor for content-based image retrieval," *IEEE Transactions on Image Processing*, vol. 21, no. 5, pp. 2874–2886, 2012.
- [17] K. Juneja, A. Verma and S. Goel, "An improvement on face recognition rate using local tetra patterns with support vector machine under varying illumination conditions," in *Int. Conf. on Computing, Communication & Automation*, Greater Noida, India, pp. 1079–1084, 2015.
- [18] Z. Mehmood, S. M. Anwar and M. Altaf, "A novel image retrieval based on rectangular spatial histograms of visual words," *Kuwait Journal of Science*, vol. 45, no. 1, pp. 54–69, 2018.
- [19] S. Lazebnik, C. Schmid and J. Ponce, "Beyond bags of features: spatial pyramid matching for recognizing natural scene categories," in *2006 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition (CVPR'06)*, New York, NY, USA, vol. 2, pp. 2169–2178, 2006.
- [20] Z. Jiang, Z. Lin and L. S. Davis, "Learning a discriminative dictionary for sparse coding via label consistent K-sVD," in *CVPR 2011*, Colorado Springs, CO, USA, pp. 1697–1704, 2011.
- [21] J. Yang, K. Yu and T. Huang, "Supervised translation-invariant sparse coding," in *2010 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition*, San Francisco, CA, USA, pp. 3517–3524, 2010.
- [22] J. Yang, K. Yu, Y. Gong and T. Huang, "Linear spatial pyramid matching using sparse coding for image classification," in *2009 IEEE Conf. on Computer Vision and Pattern Recognition*, Miami, FL, USA, pp. 1794–1801, 2009.
- [23] E. Meyers and L. Wolf, "Using biologically inspired features for face processing," *International Journal of Computer Vision*, vol. 76, no. 1, pp. 93–104, 2008.
- [24] J. Lu, V. E. Liang, X. Zhou and J. Zhou, "Learning compact binary face descriptor for face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 10, pp. 2041–2056, 2015.
- [25] G. Chen, J. Yang, H. Jin, E. Shechtman, J. Brandt *et al.*, "Selective pooling vector for fine-grained recognition," in *2015 IEEE Winter Conf. on Applications of Computer Vision*, Waikoloa, HI, USA, pp. 860–867, 2015.
- [26] J.-X. Mi and J.-X. Liu, "Face recognition using sparse representation-based classification on k-nearest subspace," *PLOS One*, vol. 8, no. 3, pp. e59430, 2013.
- [27] L. Zhang, M. Yang, X. Feng, Y. Ma and D. Zhang, "Collaborative representation based classification for face recognition," *ArXiv Preprint arXiv: 1204.2358*, 2012.
- [28] H. Wang, J. Hu and W. Deng, "Face feature extraction: A complete review," *IEEE Access*, vol. 6, pp. 6001–6039, 2017.
- [29] X. Ma, F. Zhang, Y. Li and J. Feng, "Robust sparse representation based face recognition in an adaptive weighted spatial pyramid structure," *Science China Information Sciences*, vol. 61, no. 1, pp. 1–13, 2018.
- [30] P. Karmakar, S. W. Teng, G. Lu and D. Zhang, "An enhancement to the spatial pyramid matching for image classification and retrieval," *IEEE Access*, vol. 8, pp. 22463–22472, 2020.

- [31] J. Y. Choi, "Spatial pyramid face feature representation and weighted dissimilarity matching for improved face recognition," *The Visual Computer*, vol. 34, no. 11, pp. 1535–1549, 2018.
- [32] S. J. Elias, S. M. Hatim, N. A. Hassan, L. M. Abd Latif, R. B. Ahmad *et al.*, "Face recognition attendance system using local binary pattern (LBP)," *Bulletin of Electrical Engineering and Informatics*, vol. 8, no. 1, pp. 239–245, 2019.
- [33] D. Giveki, M. A. Soltanshahi and G. A. Montazer, "A new image feature descriptor for content based image retrieval using scale invariant feature transform and local derivative pattern," *Optik*, vol. 131, pp. 242–254, 2017.
- [34] M. Agarwal, A. Singhal and B. Lall, "Multi-channel local ternary pattern for content-based image retrieval," *Pattern Analysis and Applications*, vol. 22, no. 4, pp. 1585–1596, 2019.
- [35] K. Grauman and T. Darrell, "The pyramid match kernel: Discriminative classification with sets of image features," in *Tenth IEEE Int. Conf. on Computer Vision (ICCV'05) Volume 1*, Beijing, China, vol. 2, pp. 1458–1465, 2005.
- [36] A. M. Martinez, "The AR face database," CVC Technical Report24, 1998.
- [37] F. S. Samaria and A. C. Harter, "Parameterisation of a stochastic model for human face identification," in *Proc. of 1994 IEEE Workshop on Applications of Computer Vision*, Sarasota, FL, USA, pp. 138–142, 1994.
- [38] G. B. Huang, M. Mattar, T. Berg and E. Learned-Miller, "Labeled faces in the wild: A database for studying face recognition in unconstrained environments," in *Workshop on Faces in 'Real-Life' Images: Detection, Alignment, and Recognition*, Marseille, France, 2008.
- [39] P. J. Phillips, H. Moon, S. A. Rizvi and P. J. Rauss, "The FERET evaluation methodology for face-recognition algorithms," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 10, pp. 1090–1104, 2000.
- [40] J.-Y. Wang, H.-K. Wu and Y.-S. Hsu, "Using mobile applications for learning: Effects of simulation design, visual-motor integration, and spatial ability on high school students' conceptual understanding," *Computers in Human Behavior*, vol. 66, pp. 103–113, 2017.
- [41] I. Naseem, R. Togneri and M. Bennamoun, "Linear regression for face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 11, pp. 2106–2112, 2010.
- [42] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry and Y. Ma, "Robust face recognition via sparse representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 2, pp. 210–227, 2008.
- [43] W. Deng, J. Hu and J. Guo, "Extended SRC: Undersampled face recognition via intraclass variant dictionary," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 9, pp. 1864–1870, 2012.
- [44] Y. Xu, D. Zhang, J. Yang and J.-Y. Yang, "A two-phase test sample sparse representation method for use with face recognition," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 21, no. 9, pp. 1255–1262, 2011.
- [45] G.-F. Lu, J. Zou and Y. Wang, "Incremental complete LDA for face recognition," *Pattern Recognition*, vol. 45, no. 7, pp. 2510–2521, 2012.
- [46] X. He, S. Yan, Y. Hu, P. Niyogi and H.-J. Zhang, "Face recognition using laplacianfaces," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 3, pp. 328–340, 2005.
- [47] C. Shao, X. Song, Z.-H. Feng, X.-J. Wu and Y. Zheng, "Dynamic dictionary optimization for sparse-representation-based face classification using local difference images," *Information Sciences*, vol. 393, pp. 1–14, 2017.
- [48] M. Yang, L. Zhang, J. Yang and D. Zhang, "Regularized robust coding for face recognition," *IEEE Transactions on Image Processing*, vol. 22, no. 5, pp. 1753–1766, 2012.
- [49] M. Yang, L. Zhang, D. Zhang and S. Wang, "Relaxed collaborative representation for pattern classification," in *2012 IEEE Conf. on Computer Vision and Pattern Recognition*, Providence, RI, USA, pp. 2224–2231, 2012.

- [50] A. Y. Yang, S. S. Sastry, A. Ganesh and Y. Ma, "Fast  $\ell_1$ -minimization algorithms and an application in robust face recognition: A review," in *2010 IEEE Int. Conf. on Image Processing*, Hong Kong, China, pp. 1849–1852, 2010.
- [51] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM Journal on Imaging Sciences*, vol. 2, no. 1, pp. 183–202, 2009.