

# Weighted Multi-scale Local Binary Pattern Histograms for Face Recognition

Olegs Nikisins

Institute of Electronics and Computer Science

14 Dzerbenes Str., Riga, LV1006, Latvia

Email: Olegs.Nikisins@edi.lv

**Abstract**—This paper proposes a novel face recognition methodology, which is based on the combination of the texture operator, namely Multi-scale Local Binary Pattern (MSLBP), face image filtering and discriminative feature weighting algorithms. Presented MSLBP principle enhances the discriminative power of the original Local Binary Pattern operator. The combination of the MSLBP and low-pass filtering improves the stability of the feature vector for different scales of the input face. Proposed mini-batch discriminative feature weighting methodologies are applied to the feature space in the feature and block levels in order to enhance the components more relevant to the recognition process. The observed PCA based data compression algorithm significantly reduced the dimensionality of the feature vector with a minimal loss in the precision of the recognition process. The identification precision exceeded the threshold of 99% for the frontal subsets of a color FERET database.

**Index Terms**—Local Binary Pattern operator, face recognition, discriminative feature weighting, multi-class classification.

## I. INTRODUCTION

The main blocks of the biometric recognition systems are the feature extraction module and the classification module [1]. Both of these aspects are the fields of extensive research and are addressed in this paper. One of popular approaches for the design of the feature extractor in the field of face recognition is based on the texture classifier, namely Local Binary Patterns (LBP). The LBP operator was introduced in [2] and later utilized for the face recognition problem by many researchers. A lot of extensions of the LBP paradigm were developed since it was introduced, for example Multi-scale Local Binary Patterns (MSLBP) [3] and Multi-scale Block Local Binary Patterns (MB-LBP) [4]. Both of these techniques partially degrades the dependence of LBP-based descriptor from the scale of the input image, which is a positive moment in the task of multi scale face recognition. The combination of the MSLBP and MB-LBP is proposed in this paper in order to develop robust scale invariant face recognition algorithm. This combination is based on the MSLBP operator itself and low-pass filtering of the input image.

Another challenging aspect in the field of face recognition is the design of the classification module. The usual identification approaches are based on various Nearest Neighbor Classifiers (NNC) [5], [6]. The use of complicated classifiers, such as Artificial Neural Networks or Support Vector Machine, is often inconvenient or even impossible due to insufficient intra-class information and significant amount of classes /

individuals in the database. An algorithm for the weighting of discriminative features (DFW) is developed in [1] in order to compensate the statistical incompleteness of the NNC by utilizing the information from all classes. An extension of the DFW principle is proposed in this paper. The extended version of the algorithm is more stable, predictable and provides better recognition results. The improvement is based on the special procedure of the learning data selection, which allows to maximize the distance between feature vectors, which are most likely to cause the misclassification. The information obtained in the process of weights learning is incorporated in the recognition process by the use of weighted nearest neighbor classifier (WNNC). The DFW principles are utilized in two levels: block-level and feature-level weighting.

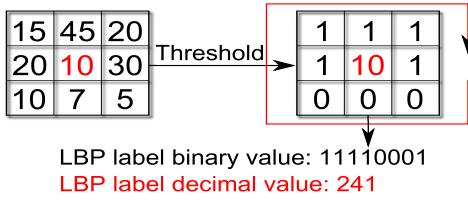
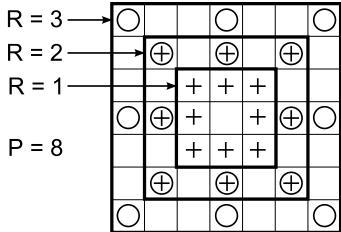
The problem of the effective memory usage is also addressed in this paper. The proposed data compression methodology is based on the PCA principles and the usage of  $L_1$  distance in the recognition module. This combination results in the significant reduction of the dimensionality of the feature space with a minimal loss in the identification precision.

The paper is organized as follows: preliminaries about the LBP and MSLBP operators are provided in section II; section III introduces the details about the MSLBP based face recognition algorithm; section IV formulates the ideas for the enhancement of the DFW methodology; the PCA-based data compression algorithm is described in section V; the experimental results and conclusive remarks are discussed in sections VI and VII.

The parameters of the proposed algorithms are evaluated on the frontal subsets of a **color** FERET database [7].

## II. MULTI-SCALE LOCAL BINARY PATTERN OPERATOR

The original LBP operator was introduced in [2] as a texture descriptor. In the input image each pixel is labelled by thresholding its  $3 \times 3$  - neighborhood with the center value and representing the result as a binary number. An example of the labeling procedure for  $3 \times 3$  region of an input image is illustrated in Figure 1. The histogram of labels is used as the descriptor of the image. Later extensions of LBP operator [8] use neighborhoods of different sizes. The notation  $(P, R)$  is usually used for the neighborhoods description [8], where  $P$  is the number of sampling points on a circle of radius  $R$ . A histogram of the labelled image  $f_l(x, y)$  can be calculated:


 Fig. 1. Labeling example of  $3 \times 3$  neighborhood, ( $P = 8, R = 1$ )

 Fig. 2. Principle of MSLBP labeling process of the  $7 \times 7$  neighborhood, ( $P = 8, R = \{1, 2, 3\}$ )

$$H_i = \sum_{x,y} I\{f_l(x,y) = i\}, i = 0, \dots, n-1, \quad (1)$$

where  $n = 2^P$  is the number of different labels and

$$I\{A\} = \begin{cases} 1, & A \text{ is true} \\ 0, & A \text{ is false.} \end{cases} \quad (2)$$

In order to save spatial information about the object the division of LBP transformed image into small regions  $R_0, R_1, \dots, R_{m-1}$  is required. A spatially enhanced histogram is calculated by the substitution of region histograms into a single feature histogram:

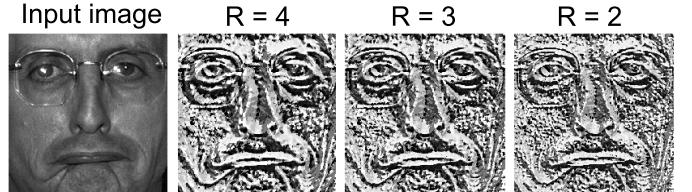
$$H_{i,j} = \sum_{x,y} I\{f_l(x,y) = i\} I\{(x,y) \in R_j\} \quad (3)$$

$$, i = 0, \dots, n-1, j = 0, \dots, m-1.$$

The idea of Multi-scale Local Binary Pattern (MSLBP) is introduced in the [3] and is based on a very simple principle of varying the radius  $R$  of the LBP label and combining the resulting histograms. The neighborhood of the MSLBP operator is described with the following parameters ( $P, R = \{R_1, R_2, \dots, R_{n_R}\}$ ), where  $n_R$  is the number of radii utilized in the process of MSLBP calculation, Figure 2. Each pixel in MSLBP image is described with  $n_R$  values.

A spatially enhanced LBP histograms for different values of  $R = \{R_1, R_2, \dots, R_{n_R}\}$  could be determined according to the Equation 3:  $H^{(1)}, H^{(2)}, \dots, H^{(n_R)}$ . The MSLBP histogram is calculated by summing these vectors:

$$H^{MSLBP} = \sum_{i=1}^{n_R} H^{(i)}. \quad (4)$$


 Fig. 3. An example of LBP transformation results for different  $R$  values

### III. MULTI-SCALE LOCAL BINARY PATTERN BASED FACE RECOGNITION

The MSLBP histogram from the Equation 4 is an effective representation of the regional and global face features. The main idea behind the MSLBP is a better stability of the descriptor for different scales of the object / face in our case. This fact is based on the observation, that the texture of the material varies for different magnification factors. LBP operator was originally introduced, as a texture descriptor and the result of the LBP transformation clearly depends on the scale of the object. Introduced MSLBP principle partially resolves this challenging aspect. This issue is also addressed in [4], where another extension of the LBP, called Multi-scale Block Local Binary Patterns (MB-LBP), is introduced. In MB-LBP, the comparison between single pixels in LBP is simply replaced with the comparison between average gray-values of the square sub-regions containing neighborhood pixels. This idea makes the descriptor more robust and reduces the noise.

Inspired by the idea of MB-LBP the **mean filtering** of the input face image is performed before further processing. This preprocessing step reduces the negative influence of the texture scale on the stability of the MSLBP histograms. Mean filtering provides slightly better results than average filtering introduced in MB-LBP.

The second preprocessing step is the **upright rotation** of the face image if the deviation of the eye line from the horizontal exceeds the determined threshold:  $\alpha_{eyeline}$ .

The MSLBP transformation of the preprocessed face image is performed next. In fact MSLBP is the same as LBP transformation with various values of the parameter  $R$ , see Figure 3 for details. The transformed images (Figure 3) are divided into  $K \times K$  regions in order to save the spatial information about the object. The LBP histogram is calculated for each region, Equation (1). Obtained regional histograms are stacked sequentially into the corresponding feature histograms:  $H^{(1)}, H^{(2)}, \dots, H^{(n_R)}$ . The resulting MSLBP histogram is calculated next according to the Equation (4). The process of spatially enhanced LBP histogram calculation is displayed in Figure 4, in case of MSLBP this process is performed  $n_R$  times.

In real life systems the scales of captured face images are different. The normalization of spatially enhanced histograms  $H^{(1)}, H^{(2)}, \dots, H^{(n_R)}$  is needed before the calculation of  $H^{MSLBP}$  in order to get a coherent description:

$$H^{(i)} \equiv H^{(i)} / ((f_x - 2 \cdot R_i) \cdot (f_y - 2 \cdot R_i)), i = 1, \dots, n_R, \quad (5)$$

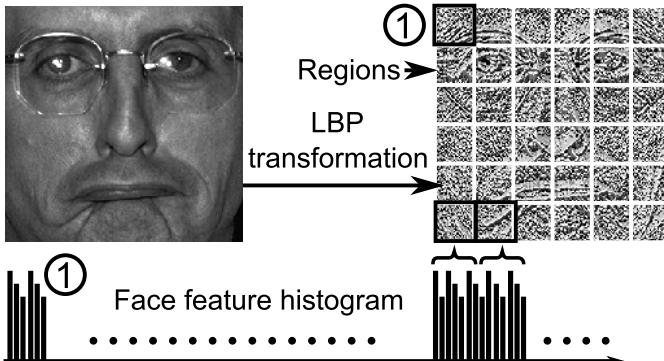


Fig. 4. The process of LBP face feature histogram forming with the regioning of LBP image  $6 \times 6$

where  $f_x$  and  $f_y$  stands for the horizontal and vertical sizes in pixels of the input face image.

A nearest neighbor classifier is usually utilized in the task of face recognition [5], [6]. Among the most popular approaches for similarity / distance measures are Euclidean Distance, Histogram intersection, Chi-square statistics.

A general problem in the biometrics is having a plethora of classes and only a small, sometimes even one, training example per class / person. This aspect makes the usability of complicated classifiers, such as Artificial Neural Network or Support Vector Machines, inconvenient or even impossible. For this reason a mini-batch discriminative feature weighting algorithm (MBDFW) is developed in [1]. MBDFW compensates the incompleteness of statistical data in nearest neighbor classifier (NNC) by utilizing the information from all classes during the learning process. This information is then implemented into the recognition process by using the weighted nearest neighbor classifier (WNNC).

The intuitive idea of histogram blocks or features weighting is based on the observation, that not all features are equally important in the recognition process. The principles of feature and block weighting are displayed in Figure 5. Enhancing of the components more relevant for the recognition improves the classification system.

Weighted NNC is used for measuring of the similarity between two MSLBP histograms, after the process of block / features weighting is performed:

- histogram intersection with feature-level weighting:

$$D(A, B) = \sum_i \min(w_i A_i, w_i B_i), \quad (6)$$

where  $i = 1, \dots, N$  and  $w_i$  is the weight for feature  $i$ ;

- histogram intersection with block-level weighting:

$$D(A, B) = \sum_j (w_j \sum_i (A_{i,j}, B_{i,j})), \quad (7)$$

where  $i = 1, \dots, n, j = 1, \dots, k^2$  and  $w_j$  is the weight for histogram block  $j$ .

Others NNC methodologies for the comparison of the feature vectors can also be extended to the weighted form.

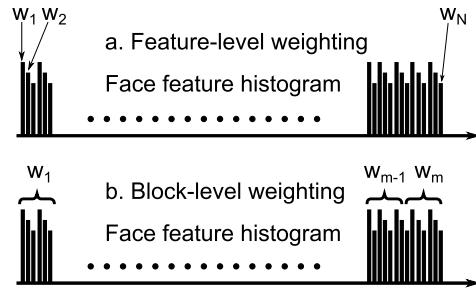


Fig. 5. Approaches for weighting in the feature (a) and block (b) levels

#### IV. DISCRIMINATIVE FEATURE WEIGHTING ALGORITHM IN THE BLOCK-LEVEL

Not all components are equally important in the recognition process. An obvious idea is to enhance the features according to their discriminative importance. This idea in context with LBP is originally introduced in [5], where authors propose an empirical methodology for the adjustment of the block weights in the LBP histogram, Figure 5 (b). The weighting principle is later enhanced in [1], where an iterative methodology of weights adjustment is introduced and extended into two levels: feature-level and block-level weighting, Figure 5. In [1] is demonstrated, that the combination of iterative block-level weighting and empirical feature-level weighting provides the best recognition performance.

Suppose, that  $d_i^{intra}$  is the value of Squared Euclidean distance between two weighted intra-class (same person) histograms  $\tilde{x}_i^{(1)}$  and  $\tilde{x}_i^{(2)}$ ,  $i = 1, \dots, M$  - is the number of the class / person and  $d_{i,j}^{inter}$  - the value of Squared Euclidean distance between two inter-class (different persons) histograms  $\tilde{x}_i^{(1)}$  and  $\tilde{x}_j^{(2)}$  with constraint  $i \neq j$ .

The principle introduced in [1] is based on the iterative re-estimation of the block weights  $\mathbf{w} = \{w_1, w_2, \dots, w_m\}$  in order to minimize  $d_i^{intra}$  and maximize  $d_{i,j}^{inter}$  for all classes  $i = 1, \dots, M$ . For this purpose the corresponding pairs  $(d_i^{intra}, d_{i,j}^{inter})$  for all classes  $i$  are utilized at each iteration of the re-estimation process, where  $j$  is selected **randomly**. The random selection of  $j$  implements a **mini-batch** principle in the learning process with intent to accelerate it, but also makes it noisy and slightly unstable.

Inspired by the ideas of neocortex simulation introduced in [9] a novel methodology of the learning data selection is introduced in this paper. The introduced principle is based on the assumption, that the sequence of the inflow of the learning data is important. The **random** selection of the data in [1] is replaced with the special selection procedure, which is performed after each  $\delta_{iter}$  iterations of the learning algorithm:  $\mathbf{n}_{iter} = \{1, i \cdot \delta_{iter}\}, i = 1, \dots, N_{iter}/\delta_{iter}$ , where  $\mathbf{n}_{iter}$  - vector with the iteration values for which the data selection procedure is performed,  $N_{iter}$  - total number of the iterations. For the iteration form the  $\mathbf{n}_{iter}$  set the **closest** (not random [1]) histogram  $\tilde{x}_j^{(2)}$  is determined for each  $\tilde{x}_i^{(1)}$  with constraint  $i \neq j$ . These histograms are utilized for the calculation of the  $(d_i^{intra}, d_{i,j}^{inter})$  pairs and are used in the learning

process for the next  $\delta_{iter}$  iterations. When the next value of the iteration from the  $\mathbf{n}_{iter}$  set is achieved the estimation of closest histograms is performed again. The Euclidean distance is used for similarity measure. An intuitive explanation of the proposed idea is simple: to maximize the distance between the histograms, which are **most likely** to cause the misclassification (face recognition errors). The proposed principle of learning data selection makes the learning process more stable and predictable, while a **mini-batch** principle and the computational efficiency of the learning algorithm are still represented.

## V. PCA FOR DATA COMPRESSION

One of important aspects in many applications is the reduction of the feature vector length. Data compression is important both for the reduction of memory required to store the data and for the computational efficiency of the recognition stage, which is actual in the case of big databases. The length of the face feature vector in the case of MSLPB is equal to  $N = K^2 \cdot 2^P$ , if  $K = 8, P = 8$  then  $N = 16384$ . Such parameters would require more than 130MB of memory space to store 1000 templates with double precision. One of possible solutions is to use the Principal Component Analysis (PCA) for the reduction of the feature space dimensionality [10].

One of popular methodologies for the evaluation of face recognition algorithms utilizes two subsets of the face database: probe and gallery sets [11]. These subsets are used for the calculation of Cumulative Match Characteristics and each subset contain 1 face image per person. Suppose  $\tilde{X}^{(1)}$  and  $\tilde{X}^{(2)}$  are the subsets of face database, where rows  $\tilde{\mathbf{x}}^{(1)}$  and  $\tilde{\mathbf{x}}^{(2)}$  stands for the **weighted** MSLBP face histograms. The dimensions of the subsets:  $\tilde{X}^{(1)} \in \mathbb{R}^{M \times N}$  and  $\tilde{X}^{(2)} \in \mathbb{R}^{M \times N}$ . The proposed PCA-based compression methodology is consequentially performed for each block of the histogram (Figure 4). The indexes of the features that belong to the block with the sequential number  $k$ :

$$\mathbf{v} = \{(k-1) \cdot N/K + 1, (k-1) \cdot N/K + 2, \dots, k \cdot N/K\}.$$

Features of the block  $k$  from the subsets  $\tilde{X}^{(1)}$  and  $\tilde{X}^{(2)}$  are then concatenated into a single array:

$$X = \{\tilde{X}_{:,v}^{(1)}, \tilde{X}_{:,v}^{(2)}\}, X \in \mathbb{R}^{2 \cdot M \times (N/K)}. \quad (8)$$

The first step of the algorithm is the preprocessing of the input data called **mean normalization**. The mean value  $\mu_j$  of the features  $X_{:,j}$  is subtracted from it:

$$X_{:,j} \equiv X_{:,j} - \mu_j, \text{ where } \mu_j = \frac{1}{2 \cdot M} \sum_{i=1}^{2 \cdot M} X_{i,j}. \quad (9)$$

The covariance matrix of normalized  $X$  is:

$$\Sigma = \frac{1}{M} \cdot X' X. \quad (10)$$

The singular value decomposition is utilized to compute the eigenvectors and eigenvalues of the covariance matrix  $\Sigma$ :

$$[U, S, V] = \text{svd}(\Sigma), \quad (11)$$

where the columns of matrix  $U$  are the eigenvectors of  $\Sigma$  and the diagonal elements of  $S$  are the corresponding eigenvalues. In order to reduce the dimensionality of the input data the  $n_e$  number of vectors of  $U$  are selected:

$$U_{reduce} = U_{:, \{1:n_e\}}. \quad (12)$$

The projected data is now calculated as follows:

$$Z = X \cdot U_{reduce}, \quad (13)$$

where the rows of  $Z$  are the new feature vectors of the block  $k$  with reduced dimensionality:  $Z \in \mathbb{R}^{2 \cdot M \times n_e}$ .

The parameter  $n_e$  is determined for each block  $k$  in order to keep the constant variances **Var** of the data for each block. For this purpose the following equation is used:

$$\text{Var} = \frac{\sum_{i=1}^{n_e} s_{i,i}}{\sum_{i=1}^{N/K} s_{i,i}}, \quad (14)$$

, where  $n_e$  is the parameter to be determined and  $s_{i,i}$  are the diagonal elements of the matrix  $S$ .

## VI. SIMULATION RESULTS

Evaluation of the proposed face recognition methodology is performed on a **color FERET** database [7]. The standard subsets **fa** and **fb** are selected from the color FERET database. The subset **fa** is used as a gallery set and **fb** as a probe set [11]. The total number of classes / persons in the database is  $M = 993$  and individuals were asked for a different facial expressions in **fa** and **fb** sets.

A **closed universe model** is selected for the evaluation of the algorithm performance [11]. In a closed universe identification every probe image has a corresponding matching template in the database. This assumption allows to determine the ability of the algorithm to identify a probe image. The results are usually represented in the form of **Cumulative Match Characteristics** (CMC), where the horizontal axis is **rank** and the vertical is the **probability of correct identification**  $P_I$ . The probability at rank one  $P_I(r=1)$  is the parameter, which is often used in the literature to compare the performance of different algorithms.

### A. Evaluation of the parameters for MSLBP-based face recognition algorithm

Proposed MSLBP-based face recognition algorithm has a lot of variables to be optimized. The MSLBP operator is described with the following parameters:  $(P, R = \{R_1, R_2, \dots, R_{n_R}\})$ , Section II. In order to reduce the considered space of the parameters the following variables are set to constant values:  $P = 8, n_R = 3$ . Only  $n_R = 3$  radii are utilized in the process of MSLBP calculation, what is a good choice both for the computational and recognition performances. The set of the radii is selected as follows:

$$R = \left\{ \frac{L-1}{2} - 2, \frac{L-1}{2} - 1, \frac{L-1}{2} \right\},$$

where  $L$  is the size of the MSLBP neighborhood (Example, Figure 2:  $L = 7$ ).

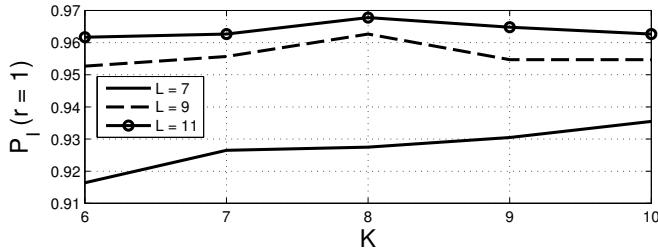


Fig. 6. Probabilities of correct identification at rank one for different  $L$  and  $K$  values (**fa** and **fb** subsets of a color FERET).

Another parameter of the algorithm to be optimized is the regioning factor  $K$ . Regioning implements the spatial information about the object into the feature vector, but very detailed partitioning of the object results in the statistical insufficiency of the region data.

The probabilities of correct identification at rank one are calculated for the subsets **fa** and **fb** of a color FERET database with different parameters  $L = \{7, 9, 11\}$ ,  $K = \{6, 7, 8, 9, 10\}$  in order to select the optimal values. For the classification of the patterns the histogram intersection methodology is selected. The upright rotation of the face images is used if the deviation of the eye line from the horizontal exceeds the value  $\alpha_{eyeline} = 3^\circ$ . The size of the face is selected empirically:  $f_x = 2,8 \cdot d_{eye}$ ,  $f_y = 3.3 \cdot d_{eye}$ , where  $d_{eye}$  stands for the interocular distance. The resulting curves are displayed in Figure 6. An explicit maximum  $P_I(r = 1) = 96.8\%$  is observed for the parameters  $L = 11$  and  $K = 8$ , which are utilized in the subsequent optimizations.

Once the  $L$  and  $K$  parameters are selected the **mean filtering** of the size  $3 \times 3$  is added to the preprocessing of the input image. This stage reduces the negative influence of the texture scale on the recognition result. The value of  $P_I$  with  $L = 11$ ,  $K = 8$  and mean filtering:  $P_I = 97.8\%$ .

#### B. Discriminative feature weighting in the block and feature levels

Not all component in the MSLBP histograms are equally important in the recognition process. Proposed weighting methodologies in the feature and block levels (Figure 5) partially resolves this issue by incorporating the weights into the NCC.

For learning purposes the data sets  $X^{(1)}$  and  $X^{(2)}$  are formed from the subsets **fa** and **fb**. The dimensionality of  $X^{(1)}$  and  $X^{(2)}$  is  $\mathbb{R}^{993 \times N}$ . According to the results in [1] the combination of **empirical** feature-level weighting with **iterative** block-level weighting provides the best recognition performance.

The selection of the **empirical** feature-level weights  $w = \{w_1, w_2, \dots, w_N\}$  is based on the following equations [1]:

$$w_j = 1 - \frac{(M-1) \sum_{i=1}^M (x_{i,j}^{(1)} - x_{i,j}^{(2)})^2}{M \sum_{i=1}^{M-1} (x_{i,j}^{(1)} - x_{i+1,j}^{(1)})^2}$$

$$w_j = \begin{cases} w_j, & \text{if } w_j \geq 0, \\ 0, & \text{if } w_j < 0, \end{cases} \quad (15)$$

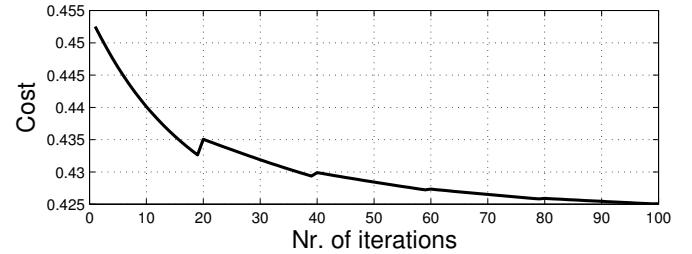


Fig. 7. An example of the cost function dependence from the number of iterations of block-level weighting algorithm ( $\eta = 10$ ,  $\alpha = 0.5$ ,  $\delta_{iter} = 20$ , **fa** and **fb** subsets of a color FERET; feature-level weighting is performed).

where the upper indexes in the notations  $x_{i,j}^{(1)}$  and  $x_{i,j}^{(2)}$  stand for the number of training example, which belongs to the same person and the lower indexes stand for the class/person number  $i$  and the number of the feature  $j$ . An explanation of the Equation 15 is the following - set high initial weights for the features, which are stable in the intra-class level and have high inter-class variance. The value of  $P_I(r = 1)$  for the subsets **fa** and **fb** of a color FERET with feature-level weighting according to the Equation 15 is  $P_I = 98.1\%$ . The weighted histogram intersection is utilized in this case.

The proposed **iterative** block-level weighting is performed next. The weights to be adjusted are the elements of a vector  $\mathbf{w} = \{w_1, w_2, \dots, w_{(K \cdot K)}\}$ . The weighting methodology is similar to the one described in [1], but enhanced with the proposed technique of learning data selection, Section IV. All block weights are initially set to one:  $w_{i,iter=0} = 1$ , where  $iter$  is the number of the iterations of the re-estimation algorithm. The algorithm described in [1] requires to set the learning rate  $\eta$  and the degree of the correct classification influence on the magnitude of the cost function  $\alpha$ . The values  $\eta = 10$  and  $\alpha = 0.5$  are selected.

An example of the cost function for the block-level weighting algorithm is displayed in Figure 7. The MSLBP histograms for the subsets **fa** and **fb** of a color FERET database are already weighted in the feature-level according to the Equation 15 in this case. The small steps are visible in the cost function after each  $\delta_{iter} = 20$  iterations, this fact could be explained with the learning data selection procedure, which is performed after each  $\delta_{iter}$  iterations.

The value of the probability of correct identification at rank one for the subsets **fa** and **fb** after 500 iterations of the block-level weighting algorithm is  $P_I = 99.2\%$ . The corresponding CMC is displayed in Figure 8.

#### C. Reduction of the feature vector dimensionality with PCA

The length of the feature vector with previously determined parameters is  $N(K = 8, P = 8) = K^2 \cdot 2^P = 16384$ . The PCA-based methodology, Section V, is utilized next to reduce the dimensionality of the feature space. The variance **Var** of the data in the blocks of the MSLBP histogram varies in the range  $[0.850, 0.995]$  with the step 0.005. The dependence of  $P_I(r = 1)$  from the value of the variance retained (**fa** and **fb** subsets, color FERET) is displayed in Figure 9. The  $L_1$

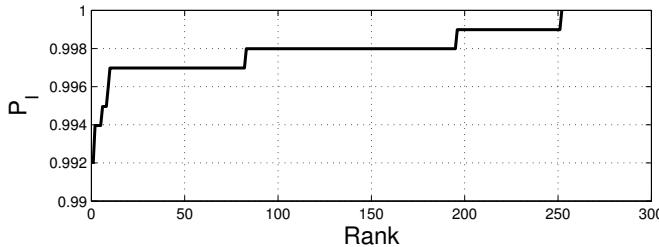


Fig. 8. Cumulative Match Characteristics after the combination of MSLBP, mean filtering, bar and block level weighting(**fa** and **fb** subsets of a color FERET).

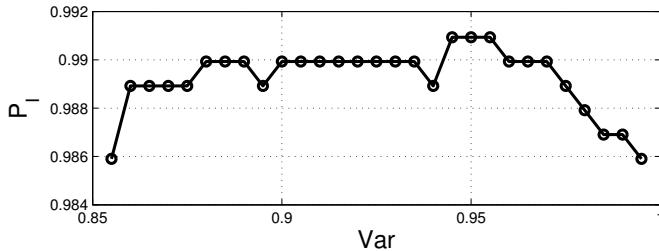


Fig. 9. The dependence of  $P_I(r=1)$  from the value of the data variance  $\text{Var}$  retained in the blocks of the MSLBP histogram(**fa** and **fb** subsets of a color FERET).

distance is utilized to compare the feature vectors after the PCA-based compression.

The highest identification probability  $P_I(r=1) = 99.1\%$  is achieved for the value of the variance  $\text{Var} = 0.945$ . The length of the feature vector in this case is  $N_{\text{reduce}} = 731$  and the data compression ratio:

$$\text{CR} = N_{\text{reduce}}/N = 731/16384 = 0.045.$$

The influence of the proposed processing steps onto the value of  $P_I(r=1)$  for the **fa** and **fb** subsets of a color FERET is summarized here:

- MSLBP ( $L = 11, K = 8, P = 8, n_R = 3$ ): 96.8%,
- MSLBP + mean filter (MF): 97.8%,
- MSLBP + MF + Feature Weighting (FW): 98.1%,
- MSLBP + MF + FW + Block Weighting (BW): **99.2%**,
- MSLBP + MF + FW + BW + PCA ( $N = 731$ ): 99.1%.

## VII. CONCLUSION

A novel face recognition methodology is introduced in this paper. Presented algorithms are based on the MSLBP texture operator and mean filtering in the feature extraction module and discriminative feature weighting and weighted nearest neighbor classifiers in the recognition module. Introduced combination of the MSLBP and mean filtering provided high accuracy in the task of face recognition, what could be explained with the improved stability of the proposed descriptor for different scales of the input face images.

The proposed extension of the mini-batch discriminative feature weighting algorithm makes the learning process more robust and clear than in [1]. It improves the performance of the face recognition system by incorporating the feature and

block weights into the nearest neighbor classifier. The idea behind the introduced learning principle is simple: maximize the distance between the MSLBP face histograms, which are **most likely** to cause the recognition errors (not between random histograms [1]).

The PCA based data compression methodology significantly reduces the memory requirements. The length of the feature vector is reduced more than 20 times. This aspect results both in memory and computational savings, which are very important in real life applications and in the embedded realizations. The loss in the recognition performance is minimal: 0.1% only.

The total of seven algorithms are evaluated in [1], [6], [12] and [13], but the performance for the **fa** and **fb** subsets of a color FERET database does not exceed 97%. The proposed top result of the probability of correct identification at rank one is **99.2%**. The recognition performance after the PCA-based compression of the feature vectors is 99.1%.

## ACKNOWLEDGMENT

This work has been supported by a funding from the European Regional Development Fund within the grant No.2010/0285/2DP/2.1.1.1.0/10/APIA/VIAA/098.

## REFERENCES

- [1] O. Nikisins and M. Greitans, "A mini-batch discriminative feature weighting algorithm for lbp - based face recognition," *Peoceedings of International Conference on Imaging Systems and Techniques (IST 2012)*, pp. 170–175, July 2012.
- [2] T. Ojala, M. Pietikainen, and D. Harwood, "A comparative study of texture measures with classification based on feature distributions," *Pattern Recognition* 29, pp. 51–59, 1996.
- [3] C. H. Chan, "Multi-scale local binary pattern histogram for face recognition," Ph.D. dissertation, Centre for Vision, Speech and Signal Processing School of Electronics and Physical Sciences University of Surrey, 2008.
- [4] S. Liao, X. Zhu, Z. Lei, L. Zhang, and S. Z. Li, "Learning multi-scale block local binary patterns for face recognition," in *ICB*, 2007, pp. 828–837.
- [5] T. Ahonen, A. Hadid, and M. Pietikainen, "Face recognition with local binary patterns," *Computer Vision, ECCV 2004 Proceedings, Lecture Notes in Computer Science 3021, Springer*, pp. 469–481, 2004.
- [6] O. Nikisins and M. Greitans, "Reduced complexity automatic face recognition algorithm based on local binary patterns," *Peoceedings of 19th International Conference on Systems, Signals and Image Processing (IWSSIP 2012)*, pp. 447–450, April 2012.
- [7] [Http://face.nist.gov/colorferet/](http://face.nist.gov/colorferet/).
- [8] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24, p. 971987, 2002.
- [9] J. Hawkins and S. Blakeslee, *On Intelligence*. St. Martin's Griffin; First Edition, 2005.
- [10] D. Kumar, C. S. Rai, and S. Kumar, "Principal component analysis for data compression and face recognition," Institute of Science and Technology, Klawad, Tech. Rep., 2008.
- [11] 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.
- [12] Y. M. R. J. Young Choi, K. N. Plataniotis, "Using colour local binary pattern features for face recognition." *Proceedings of 2010 IEEE 17th International Conference on Image Processing, ICIP 2010*, pp. 4541–4544, 2010.
- [13] S. G. K. Delac, M. Grgic, "Independent comparative study of pca, ica, and lda on the feret data set," *International Journal of Imaging Systems and Technology*, vol. 15, no. 5, p. 252260, 2006.