

Influence of Image Enhancement Techniques on Effectiveness of Unconstrained Face Detection and Identification

Igor Vukovic¹, Petar Cisar², Kristijan Kuk², Milos Bandjur³, Brankica Popovic^{2,*}

¹Ministry of the Interior of the Republic of Serbia,
Bulevar Mihajla Pupina St. 2a, RS-11000 Belgrade, Serbia

²University of Criminal Investigation and Police Studies,
Cara Dusana St. 196, RS-11080 Zemun, Serbia

³Faculty of Technical Sciences, University of Pristina in Kosovska Mitrovica,
Kneza Milosa St. 7, RS-38220 Kosovska Mitrovica, Serbia
brankica.popovic@kpu.edu.rs

Abstract—In a criminal investigation, along with processing forensic evidence, different investigative techniques are used to identify the perpetrator of the crime. It includes collecting and analyzing unconstrained face images, mostly with low resolution and various qualities, making identification difficult. Since police organizations have limited resources, in this paper, we propose a novel method that utilizes off-the-shelf solutions (Dlib library Histogram of Oriented Gradients-HOG face detectors and the ResNet faces feature vector extractor) to provide practical assistance in unconstrained face identification. Our experiment aimed to establish which one (if any) of the basic image enhancement techniques should be applied to increase the effectiveness. Results obtained from three publicly available databases and one created for this research (simulating police investigators' database) showed that resizing the image (especially with a resolution lower than 150 pixels) should always precede enhancement to improve face detection accuracy. The best results in determining whether they are the same or different persons in images were obtained by applying sharpening with a high-pass filter, whereas normalization gives the highest classification scores when a single weight value is applied to data from all four databases.

Index Terms—Face detection; Face recognition; Histogram of oriented gradients; Image processing.

I. INTRODUCTION

As one of the main biometric traits, the application of face recognition is of increasing importance in the areas of video surveillance, access control, human-computer interaction, border control surveillance, and crime investigation [1]–[3]. The application in a criminal investigation to identify the perpetrators of crimes, which is of particular interest for this paper, stands out primarily because of the emphasis on security in modern society [4]. There are two decisive reasons for the importance of the application of face recognition in criminal investigations. First, the human face carries different information about identity, age, gender, race, emotions, and mental states [1].

Second, the face recognition does not require the cooperation of a person and can be performed unobtrusively. Both reasons give the application of this biometric method an advantage in analysing video and photo surveillance in the search for known perpetrators, then for analytical and forensic purposes in the detection and prosecution of defendants and discovering missing persons, primarily children [5]. However, the negative consequences of misinformation obtained from the face recognition system can directly impact citizens' lives and human rights. Therefore, the criminal investigator has a decisive role in the process of identification of persons. Information systems and software tools that process a large amount of data can only give suggestions. Criminal investigators expect such information systems and tools to provide them with a comprehensive analysis of the evidence gathered with the essential technical requirements, as accurate as possible suggestions. In addition, investigators expect the systems and tools to help them make the final decision providing enhanced images on which suggestion is made (more suitable for visual inspection by the investigator). Information systems and tools base the process of face recognition on recognizing its general patterns achieving the best results with photos of faces placed frontally. This process presents a challenge in unconstrained conditions where the recognition rate could drop dramatically using standard technologies [6]. The problems are in the following factors that occur in such situations [1], [2], [7], [8]:

- Factors that can cause variation in the appearance can be intrinsic (depending on the face's physical structure, such as expression, age, and occlusion) and extrinsic (depending on illumination, scale/resolution, pose, noise, and blur);
- Factors influencing the similarity in face appearance between different individuals, such as kinship and face manipulation;
- Factors related to the technical characteristics of the sensors, such as low resolution.

Unconstrained conditions are the main characteristics of video and photo material that criminal investigators usually process to identify the perpetrators of criminal acts. Sources of images as digital evidence today can be the product of covert police surveillance, digital forensic reports of confiscated mobile phones and computers, and the material produced in open-source intelligence (OSINT) work. In this research, we experimentally determined the impact of the visual improvement of images on the process of face recognition, i.e., identification. Face identification is a process of comparing faces with all feature templates stored in the database (computation a one-to-many similarity) to find the subject identity among several possibilities that match the most [2], [9]. We perform an experiment using the Dlib version of the Histograms of oriented gradients (HOG) descriptor for face detection and ResNet to create a 128-dimensional face features vector. Dlib is a C++ toolkit that includes machine learning algorithms for face detection and a feature point algorithm [10]. It uses ResNet, a deep residual network first proposed by Kaiming He, for 128-dimensional face features vector extraction. The particular experimental setup offers face identification with minimal requirements for resources in data processing and storage. The main reason lies in the practical applications of such systems and tools in a criminal investigator's environment commonly faced with limited resources. The features vector of 128 elements can be stored in a database and used for comparison during identification. It is important to emphasize that amount of comparisons is further proportional to the size of the database.

Many papers deal with the preprocessing of images in face recognition, not necessarily in the way images are visually enhanced. The following are examples that represent groups of such research. One example is when the preprocessing is intended for the appropriate technologies applied, so the conversion of color photography to black and white (shades of gray) is performed due to the use of Gabor filters to extract features that represent the face [4]. Many comparative studies also measure the impact of different illumination preprocessing techniques on face recognition performance, concluding that such techniques give almost perfect results in controlled lighting variations and that better visual preprocessing results do not guarantee better recognition accuracy [11]. Another way to approach preprocessing is to replace missing image information with existing ones, e.g., creating an approximately symmetric virtual face image using the least degraded half of the original face image (for example, half of a non-shaded face makes an entire face) [7]. In this way, preprocessing reduces the negative effects of heterogeneous lighting. On the other hand, it creates a set of samples for face recognition instead of using original photographs, which solves insufficient samples of appropriate quality for matching. Similar to the previous example, using existing information and trained intelligent and expert systems, methods are proposed to normalize variations in poses [12]. In some research, image preprocessing is suggested, yet improvement in recognition rate is achieved later in the process, for example, combining two descriptors that merge at the classifier level [9], [13].

This paper's contribution is reflected in determining

whether the face detection and identification effectiveness will increase regarding the application of specific enhancement techniques on unconstrained images (emphasizing the ones obtained in police investigation work). To simulate the material created due to forensic reports and covert police surveillance, a particular set of data was made for this paper's needs as an additional contribution. When comparing the effects of various image improvements during the face identification, we considered unbalanced sets of positive and negative comparison results. Unbalanced sets of matching results correspond to the actual conditions of potential application within criminal information systems and tools. In this way, the results have a broader use-value in applying HOG descriptors, face features vectors and other methods of recognizing faces outside the domain of criminal investigations.

The rest of the paper is organized as follows. In the next section, the materials and methods we use in the experiment are described. Then experimental setup is explained, followed by a discussion of obtained results and the conclusion we have made.

II. MATERIALS AND METHODS

A. Databases with Labelled Face Images

The experiment was conducted on four labeled image databases, three of which are publicly available.

The fourth was created for this research only (internally called SSIM) to simulate the material obtained by digital forensics units or covert surveillance police operations. It contains 6,027 face images of 328 different persons. For most of them, the photos were taken in a period longer than 15 years, resulting in significant physical appearance changes due to aging, as shown in Fig. 1.



Fig. 1. Representative examples from the SSIM database.

Face images vary in resolution from 31×31 pixels to 2385×2385 pixels. The quality of images is different, which, in addition to the various resolutions, is another essential feature of pictures collected for criminal purposes. It is necessary to point out that apart from a few professional ones, all other images are created with amateur or consumer equipment. The lighting is mainly artificial and placed frontally. Due to the lower intensity of the light source, the faces are in the shadow. Image sharpness is shown to be generally satisfactory for visual recognition, but images that are more or less out of focus predominate. Although front-facing is dominant, there are different poses, with many photos with semi-profiles and profiles.

The IMDB-WIKI database [14] contains 523,051 photos of celebrities collected by crawling from IMDb and

Wikipedia websites. The database is interesting for our research because it contains photographs with significant differences in the age of the same persons (from childhood to adulthood). Because faces belong primarily to actors, the transformations in appearance due to trends or movie roles often make recognition a challenge. Images contain scenes from films that can be said to simulate unconstrained conditions. In addition, the pictures vary in resolution. The authors of this database downloaded the information about the name, date of birth, gender, and all images related to a person from IMDb site celebrities' profiles. In the caption of the site pictures, there is a date when the picture was taken. It was assumed that when there is one photo on a specific person's profile, it is probably a photo of the person whose profile it is. In contrast, when it comes to pictures with more than one person, they only use images where the second strongest face detection is below a threshold. The set's analysis revealed problems with the wrong faces and assigned data on the years when the images were created, which required correction by hand. To deal with the most accurate data from IMDB-WIKI, we took only photographs that we could confirm to be the correct person.

The Extended Yale Face Database B is a set that initially contained 16,128 photos of the faces of 28 different people under nine poses and 64 illumination conditions. The authors created a database for the systematic testing of face recognition methods under extensive variations in illumination and poses, without the influence of different facial expressions, aging, and occlusion [15]. All face photos in the database have a resolution higher than 150 pixels.

The Labeled Faces in the Wild (LFW) database provide a large number of face images taken under various conditions, starting with the newspaper photographs depicting people in different poses and the facial expressions under different lighting [16], [17]. The photos resolution is mainly less than 150 pixels, which corresponds to images police investigators collect by OSINT. The database contains 13,233 images belonging to 5,749 different people, of which 1,680 people have two or more pictures.

B. Image Enhancement

For the experiment, to meet the operational needs of the police work, we used the basic enhancement operations: contrast-stretching (normalization), sharpening, high-pass filtering, reducing the speckles within a photo, and resizing, as well as a combination of these operations. Enhancement is automated using the *convert* application belonging to the ImageMagick 7.0 set of tools. It is a free software that over a command-line interface creates, edits, composes, and converts raster graphics in many different formats, using multiple threads for a calculation to increase the performance [18]. Every enhancement starts with the unprocessed images (UNP) of the four databases.

As previously stated, one of the main characteristics of the used databases is different image resolution, which distribution over the databases is shown in Fig. 2. It can be noticed that the number of images with a resolution lower than 150 pixels is uneven within the databases, but in total such images participate to a sufficient extent that the

influence of resizing can be significant. We used the *convert* application and *-resize* option to increase all face images' size smaller than 150 pixels. The data processed in this way we marked with R, or if it is a combination of different preprocessing methods in question, R is on the beginning. The normalization (N) represents the increase in contrast in an image by stretching the range of intensity values. It is necessary to determine the lower and upper pixel value limits. Based on the previous setting of the operator of the *convert* command, the lowest pixel intensity value will be 2 % of the black-point values and the highest value will be 99 % of the white-point values. The operator uses histogram bins to determine the range of color values that needs to be stretched. Figure 3 and Figure 4 show the raw image and the image after resizing, and normalization is applied. Operator *-unsharp* (S) of the application *convert* sharpens an image by performing convolution with a Gaussian operator of the given radius and standard deviation (*sigma*) [18]. The selected value of the standard deviation is 5, and the bigger number produces a sharper image. The radius is 0, which means that a suitable radius for the corresponding *sigma* value is chosen by the *convert* application [18].

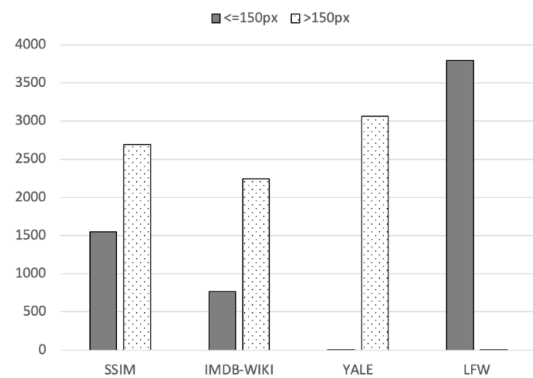


Fig. 2. Resolution distribution in the databases used in the experiment.

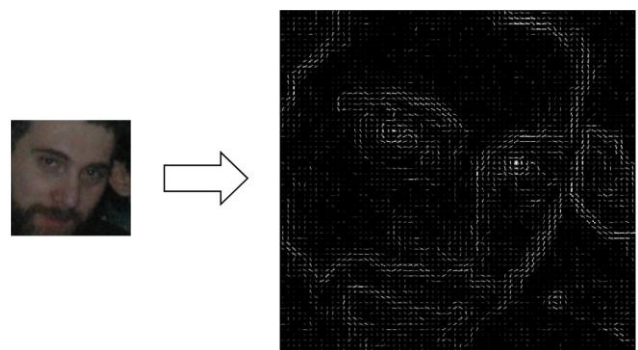


Fig. 3. HOG descriptor of unprocessed image.

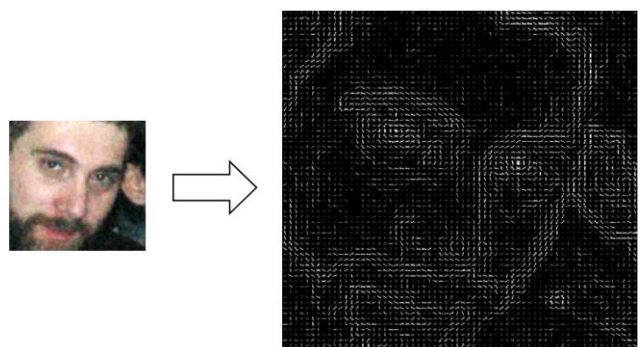


Fig. 4. HOG descriptor of enhanced image.

A high-pass filter (HP) and resize and despeckle (RD) combines several different operators within the *convert* application to sharpen images and remove noise by blurring, respectively.

C. Face Detection and Feature Vector Extraction

The face recognition system generally includes: Face Detection and Extraction, Feature Extraction and Representation, and Feature Matching [2], [19]. One of the popular methods for face detection is Histograms of oriented gradients (HOG), proposed by Dalal and Triggs [20].

The HOG is a feature descriptor that is invariant to 2D rotations and scaling [2]. The idea behind the HOG descriptor is that the distribution of intensity gradients can characterize the shape of objects in photographs. It is sensitive to the slightest changes in shape unless the whole object is consistent [21].

To obtaining a HOG descriptor, the image needs to be divided into blocks of pixels, i.e., cells. A histogram of the edges' orientation, i.e., gradient directions, for each cell is calculated. The normalization is performed using a cell block pattern, which gives a HOG descriptor for each cell and then the whole image's feature vector. Further, the feature vector is used for the face detection and recognition [21]. The first step is to calculate the amplitude of the gradients of each pixel of the image $I(x, y)$ in the horizontal and vertical directions, based on which the magnitude of the gradients $|G|$ and the orientation angle γ are obtained [13], [22], [23].

The following formulas are used to calculate the magnitude and orientation:

$$dx = I(x+1, y) - I(x-1, y), \quad (1)$$

$$dy = I(x, y+1) - I(x, y-1), \quad (2)$$

$$|G| = \sqrt{dx^2 + dy^2}, \quad (3)$$

$$\gamma(x, y) = \tan^{-1}\left(\frac{dx}{dy}\right), \gamma \in [-\pi, \pi]. \quad (4)$$

Split of the orientation range into k bins, computation of the histogram within the cell (HC), and then integration into the block by combining $b_1 \times b_2$ cells, we can obtain the histogram of the block HB [22]:

$$HC(k)_i = HC(k)_i + |G|, \quad (5)$$

$$if \ I(x, y) \in cell_i \ and \ \gamma \in bin(k), \quad (5)$$

$$HB_j = \{HC_1, HC_2, \dots, HC_{b_1 \times b_2}\}. \quad (6)$$

The orientation histogram of each cell is constructed by accumulating the gradient magnitudes related to the corresponding class interval concerning the corresponding orientation angle [13]. In addition, the blocks overlap so that each cell contributes more than one block [24].

Due to variations in the image, such as brightness or contrast between foreground and background, the gradient values also vary significantly, which is why L2-norm block normalization is applied [22]–[25]

$$[h]NHB_j = \frac{HB_j}{\sqrt{\|HB_j\|_2^2 + e^2}}. \quad (7)$$

The NHB_j is a normalized block, whereas e is a constant with the purpose of avoiding division by zero. The integration of the normalized blocks builds a HOG descriptor. In this paper, we used a Dlib version of a HOG face detector and ResNet. Although not currently state of the art, the Dlib HOG face detector's implementation is an off-the-shelf solution that achieves reasonable precision and processing speed [26], [27]. Figure 3 and Figure 4 also show a visual representation of the HOG, created using the skimage Python library [28], before and after enhancement. There are face contours with more information after normalization is applied.

The descriptor extraction transforms multidimensional arrays of discrete spatial units (pixels) into lower dimensions while retaining enough data to represent the information. The face is represented by the vector, which contains the unique features of the face [2].

III. EXPERIMENT

In everyday work, the police investigators routinely use different image processing software trying to obtain the enhanced images more suitable for visual inspection and decision-making regarding identity resolution. It is a tedious work requiring a lot of effort and time that they often do not have. The state-of-the-art solutions based on machine learning algorithms trained on huge databases to provide high face recognition effectiveness are expensive and not available for utilization within most police organizations.

Therefore, the basic idea behind our work was to design a novel system for collecting, storing, and analysing unconstrained face images in real-time with limited resources (regarding processing power and storage capacity) that will assist in police investigation work. To simulate limited resources, we used a consumer laptop with an Intel i3-5005U dual-core 2.0 GHz base frequency processor and 4 GB of RAM in our experiment.

The first step we took was to conduct an experiment to establish which one (if any) of the basic image enhancement techniques should be applied prior to the unconstrained face detection and identification to increase the effectiveness.

At the beginning of the experiment, we create a copy with a sharpened image, normalized, resized, etc., for each photo from the databases. Next, the HOG-based face detection is performed on the unprocessed images and all made copies followed by the feature vectors extraction. The last step of the recognition process consists of creating a working set representing the intersection of the sets of all obtained feature vectors, which will be considered in the matching process.

Only the faces detected by applying all the enhancement techniques were taken into account and further analysed during the experiment. In addition, the databases with significantly more images were reduced to the SSIM database size for a more uniform comparison. Based on that

the SSIM database was reduced to 4,243 images, the IMDB-WIKI database was reduced to 3,011 images (330 different persons), the Yale B database was reduced to 3,069 images, and the LFW database was reduced to 3,803 photographs (showing 119 people).

Figure 5 shows the course of the experiment. Images from each database are firstly processed with different enhancement methods previously described. Subsequently, the HOG and ResNet functions from the Dlib library were used to perform the face detection and face-feature vectors' extraction. The working set V_{WS} is determined by

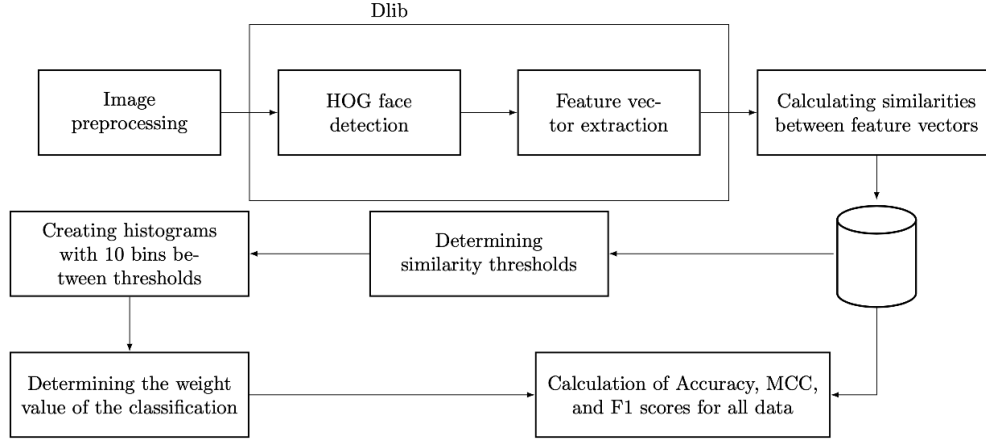


Fig. 5. The procedure of conducting the experiment.

Algorithm 1. Automation of image enhancement and creation of a working set.

```

Input: Database with unprocessed face images for recognition phase
 $I_{UNP} \leftarrow \{I_1, I_2, \dots, I_n\}$ 
Output: A working set to be used for face matching composed of a face
feature vectors  $V_{WS}$ 
 $V_S \leftarrow \{ \}$  //S - sharpened images
 $V_N \leftarrow \{ \}$  //N - normalized images
 $V_{NS} \leftarrow \{ \}$  //NS - normalized and sharpened images
 $V_{RD} \leftarrow \{ \}$  //RD - resizing and despeckle combination
...
 $V_{RHP} \leftarrow \{ \}$  //RHP - resizing and high-pass filter
for  $i \leq n$  do
   $I_{Si} \leftarrow (\text{convert } I_i \text{ -unsharp } 0x5)$ 
   $\vec{V}_i \leftarrow \text{dlib.face\_recognition\_model.v1().compute\_face\_descriptor}(I_{Si})$ 
   $V_S \cup \{ \vec{V}_i \in \mathbb{R}^{128} \}$ 
   $I_{Ni} \leftarrow (\text{convert } I_i \text{ -normalize})$ 
   $\vec{V}_i \leftarrow \text{dlib.face\_recognition\_model.v1().compute\_face\_descriptor}(I_{Ni})$ 
   $V_N \cup \{ \vec{V}_i \in \mathbb{R}^{128} \}$ 
  ...
   $I_{RHP} \leftarrow (\text{convert } I_i \text{ -resize } 150x \setminus < -hp^*)$ 
   $\vec{V}_i \leftarrow \text{dlib.face\_recognition\_model.v1().compute\_face\_descriptor}(I_{RHP})$ 
   $V_{RHP} \cup \{ \vec{V}_i \in \mathbb{R}^{128} \}$ 
end for
 $V_{WS} \leftarrow V_S \cap V_N \cap \dots \cap V_{RHP}$ 

```

Note: * - the specified option in real application replaces a series of commands.

The process of the face matching and cosine angle calculation is described by (8)–(10). The identification process consists of determining the similarity value (\mathcal{G}) between the feature vectors representing the detected faces and belonging to the working set by a combination without repetition $\binom{|V_{WS}|}{2}$:

$$\vec{V} \in \mathbb{R}^{128}, \quad (8)$$

$$V_{WS} = \{ \vec{V}_1, \vec{V}_2, \vec{V}_3, \dots, \vec{V}_i, \vec{V}_j, \dots, \vec{V}_n \}, \quad (9)$$

comparing the vectors from the sets of images obtained after different enhancements, and it is equal to the values of the reduced databases. Further in the experiment, we use only the vectors obtained after all the performed enhancements, whereas the others are excluded. This part of the experiment is presented in more detail in Algorithm 1. The face matching is made using cosine similarity (distance), which is a measure of similarity based on each vector's component composition. We choose this similarity measure between two non-zero vectors because it is a naturally normalized distance where the outcome is bounded in $[0, 1]$.

$$\{ (\vec{V}_i, \vec{V}_j) \mid \vec{V}_i, \vec{V}_j \in V_{WS}, i \neq j \}, \quad (10)$$

$$\mathcal{G} = \cos \Theta = \frac{\vec{V}_i \times \vec{V}_j}{\|\vec{V}_i\| \|\vec{V}_j\|}. \quad (11)$$

In the experiment, we used the confusion matrix to determine the relationship between the actual values and the prediction obtained by selecting the appropriate threshold [29].

Figure 6 shows the confusion matrix for a two-class classification.

| | | Predicted Class | |
|------------|----------|-----------------|----------|
| | | Positive | Negative |
| True Class | Positive | TP | FN |
| | Negative | FP | TN |

Fig. 6. The confusion matrix for a two-class classification.

TP (True positives) values represent the correct prediction of a positive outcome, meaning the result of comparing faces within two different photos is the same, and it corresponds to reality. TN (True negatives) values represent the correct prediction of a negative outcome, i.e., predicting that there are different persons on images, which also corresponds to the actual situation. FP (False positives) presents false predictions claiming to be the same person on the matching images. In contrast, FN (False negatives) represents wrong predictions about different people when it

is the same person. Two additional values are considered: the $TNrate$ or $Specificity$ illustrating the percentage of negative samples correctly classified, whereas $TPrate$ or $Recall$ ($Sensitivity$) represents the percentage of positive samples correctly classified (examples shown in Fig. 7 and Fig. 8, respectively) [30]:

$$TPrate(Recall) = \frac{TP}{TP + FN}, \quad (12)$$

$$TNrate(Specificity) = \frac{TN}{FP + TN}. \quad (13)$$

Before selecting the threshold to measure the impact of enhancement on the effectiveness of face identification on all databases and after storing the similarity values in the database, we determined the similarity thresholds representing the uncertainty interval's boundaries. Within the uncertainty interval, $TPrate$ and $TNrate$ are below 100%. In other words, with the \mathcal{G} value, we cannot indicate with complete certainty whether it is the same or different persons. The upper limit of the interval for each of the subsets of the enhancement images is the maximum similarity value where the result of matching is TN (\mathcal{G}_{max}^-). The lower limit of the interval is the minimum similarity value where the face matching result is TP (\mathcal{G}_{min}^-).

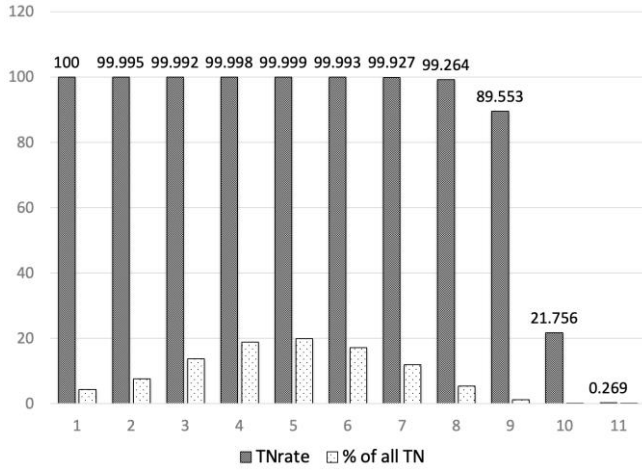


Fig. 7. Example of histogram of the distribution of true negatives (TN).

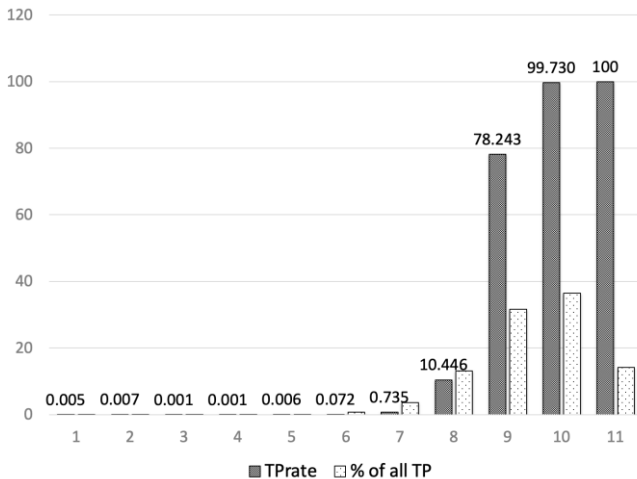


Fig. 8. Example of histogram of the distribution of true positives (TP).

We used histograms, such as the examples shown in Fig. 7 and Fig. 8, during the analysis of the experiment's data. The bin width h of the histogram represents the difference between the values of the previously defined maximum and minimum similarity limits divided by k , the chosen number of bins

$$h = (\mathcal{G}_{max}^- - \mathcal{G}_{min}^+) / k. \quad (14)$$

In Fig. 7, the uncertainty interval is shown by bins from 2 to 11, whereas in bin 1, $TNrate$ is 100% ($\mathcal{G} < \mathcal{G}_{min}^+$). The uncertainty interval in Fig. 8, where the focus is on the true positives, is covered by bins from 1 to 10, whereas in bin 11, $TPrate$ is 100% ($\mathcal{G} > \mathcal{G}_{max}^-$).

The unique threshold or weight value (w) for the set of similarities related to enhancement applied over all four databases were calculated using the mean value ($\bar{\mathcal{G}}$) and standard deviation (s):

$$\bar{\mathcal{G}} = \frac{1}{4} \sum_{i=1}^4 \mathcal{G}_i, TPrate > 0.9, \quad (15)$$

$$s = \sqrt{\frac{1}{3} \sum_{i=1}^4 (\bar{\mathcal{G}} - \mathcal{G}_i)^2}, \quad (16)$$

$$w = \bar{\mathcal{G}} - s. \quad (17)$$

We used the weight value w for the classification according to the rule by which all similarity values $\mathcal{G} > w$ indicate the same person. In contrast, in the case of $\mathcal{G} \leq w$, faces in the pictures belong to different persons. Based on the confusion matrix elements, we determined accuracy (Acc) as the number of accurate predictions based on the determined weight value w and the total number of predictions, i.e., matching [29]

$$Acc = \frac{TP + TN}{TP + FP + TN + FN}. \quad (18)$$

In the unbalanced sets, where the number of samples of one class is significantly higher than of another, accuracy can no longer be considered a sufficiently reliable measure. The accuracy too optimistically estimates the classifier performance for the majority class [31]. On a large scale of persons who need to be identified during the criminal investigation in actual conditions, a disproportionate number of different and the same persons are expected to favor different ones. Therefore, other classification quality measures have been applied in the paper, namely the $F1$ score and Matthews correlation coefficient (MCC). A standard measure used in unbalanced sets is $F1$, representing the harmonic mean of precision and recall, and it has the following form [31]:

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall}, \quad (19)$$

$$Precision = \frac{TP}{TP + FP}, \quad (20)$$

$$F1 = \frac{2 \times TP}{2 \times TP + FP + FN}. \quad (21)$$

The minimum value of the measure, i.e., $F1 = 0$, is obtained when the number of correctly positively classified samples is equal to zero, $TP = 0$. In contrast, the perfect classification, $F1 = 1$, is considered when the number of

incorrect negative and positive classifications is equal to zero, $FN = FP = 0$. What is immediately noticeable is that the $F1$ measure does not consider accurate negatively classified predictions TN .

Additional verification of the quality of the classification was performed by the MCC , which considers all confusion matrix elements

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP) \times (TP + FN) \times (TN + FP) \times (TN + FN)}}. \quad (22)$$

Suppose $MCC = 0$; the classification result is equal to a random selection. A positive value indicates the correct classification (a value of +1 coefficient represents a perfect classification). In contrast, the negative values indicate worse classification than a random selection, where the value of a coefficient of -1 represents a wide discrepancy between the prediction and the concrete classes. The MCC is the only measure of the binary classification that generates a high score only when a particular model performs a successful classification regardless of which class is dominant [31].

IV. RESULTS AND DISCUSSION

The experimental results suggest that the application of different types of image enhancement impacts the total number of detected faces.

By counting the feature vectors from each set of enhanced images, we obtained the data shown in Table I and Fig. 9, representing the impact of enhancement on face detection effectiveness. The up arrow (\uparrow) in Table I shows an increase in the number of detected faces by applying some of the visual enhancement methods compared to the unprocessed images. The down arrow (\downarrow) indicates that the number of detected faces decreased after preprocessing, whereas the equals sign ($=$) indicates that the number of detected images did not change. Based on the markings in Table I, it can be concluded that the enhancement has a significant positive effect on the number of detected faces. The best results we achieved by applying sharpening after enlarging images with a resolution of less than 150 pixels.

Figure 9 allows a clearer view of the overall impact on face detection's effectiveness concerning the samples from all databases that we used in the paper. The chart on Fig. 9 shows the mean value of the number of detected persons for all databases after min-max normalization. The photo sharpening gives the best results both independently (S, RS) and in combination with normalization (NS). Comparing the results of applying the same enhancement method with and without resizing in Fig. 9 shows that resizing images with a resolution of less than 150 pixels improves face detection.

The differences in the number of detected faces in Table I emphasize the need to measure the impact of image enhancement on the identification process's effectiveness based on a previously created working set.

In the simulation of the face identification applied in the experiment, we performed binary classification by comparing the faces from one image with all other faces from the working set. The working set is built on a specific

image enhancement method within the database. Then the procedure is repeated for each remaining subsequent image. Figure 10 and Figure 11 show the result of the first step of the described face identification process relating to the same face image, before and after enhancement, respectively.

TABLE I. THE NUMBER OF DETECTED FACES BY DATABASES AND APPLIED IMAGE ENHANCEMENT METHODS.

| | SSIM (6,027) | IMDB (10,044) | YALE (5,992) | LFW (4,163) |
|-----------|------------------|-------------------|-------------------|------------------|
| I_{HP} | 5514 \uparrow | 9618 \downarrow | 3358 \uparrow | 3956 \uparrow |
| I_N | 5465 \uparrow | 9611 \downarrow | 3340= \uparrow | 3949 \uparrow |
| I_S | 5542 \uparrow | 9701 \uparrow | 3539 \uparrow | 3984 \uparrow |
| I_{NS} | 5556 \uparrow | 9696 \uparrow | 3542 \uparrow | 3975 \uparrow |
| I_{UNP} | 5419= \uparrow | 9623= \uparrow | 3340= \uparrow | 3940= \uparrow |
| I_R | 5488 \uparrow | 9587 \downarrow | 3340= \uparrow | 3941 \uparrow |
| I_{RHP} | 5552 \uparrow | 9607 \downarrow | 3358 \uparrow | 3959 \uparrow |
| I_{RN} | 5534 \uparrow | 9582 \downarrow | 3340= \uparrow | 3948 \uparrow |
| I_{RS} | 5615 \uparrow | 9701 \uparrow | 3539 \uparrow | 4005 \uparrow |
| I_{RD} | 5421 \uparrow | 9623= \uparrow | 3129 \downarrow | 3972 \uparrow |

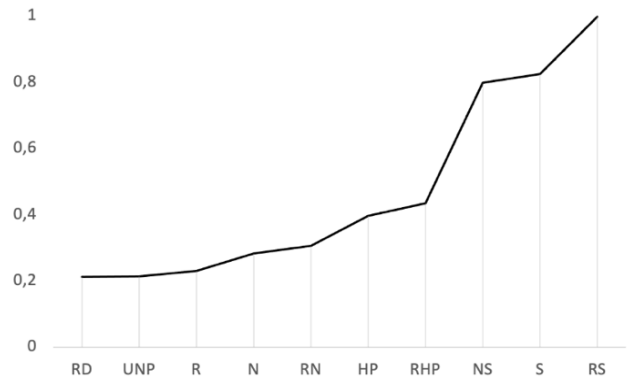


Fig. 9. Impact of enhancement on the total number of detected persons.

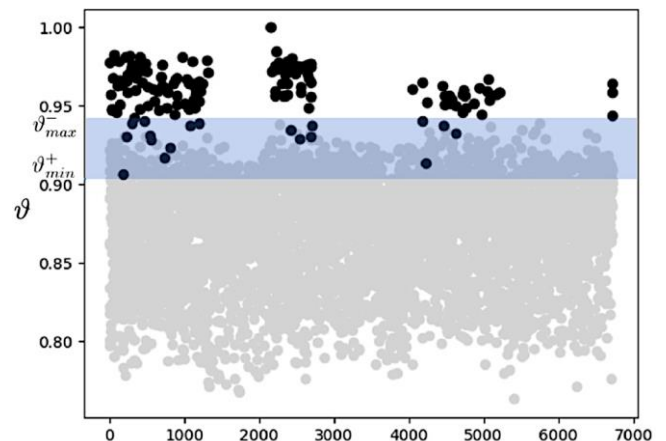


Fig. 10. Similarity values of unprocessed images obtained by matching one face with all others in the SSIM database.

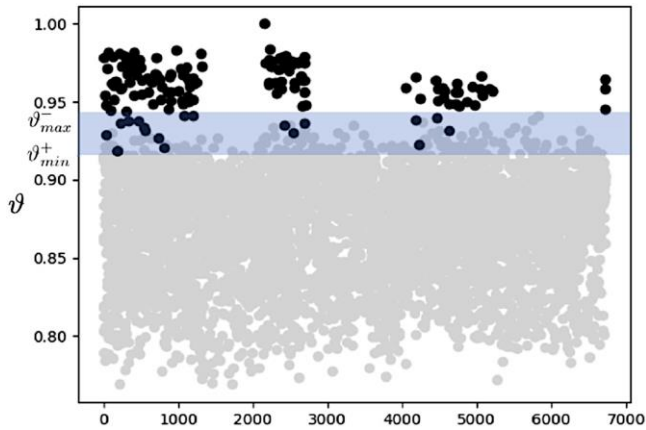


Fig. 11. Similarity values of normalized images obtained by matching one face with all others in the SSIM database.

The similarity value indicating the same person is represented by dark dots in Fig. 10 and Fig. 11, whereas light dots represent different people's faces.

To determine the weight value w based on which the binary classification would be performed, the problem of determining the most optimal similarity ρ is encountered. The problem is visually presented as a shaded area in Fig. 10 and 11, which we called the "uncertainty intervals". The uncertainty interval is bordered by the threshold values above which all comparisons indicate the same face (all dark spots) and below which we can claim that all faces are different (all bright spots). In addition, the comparison of Fig. 10 and Fig. 11 shows how the uncertainty interval changes after image enhancement. In this particular case, the advantage of normalization to unprocessed images is noticeable. The continuing identification with other faces from the images produces new different intervals within the same set. The goal is to make the overall interval as narrow as possible and with fewer similarity values within the uncertainty interval.

By measuring the results based on face matching from all sets by different image enhancement methods and then normalizing the obtained values for a unified representation on the chart, we obtained Fig. 12. The chart shows the uncertainty interval's size ratio and the number of similarities, whose values can indicate both the same and different persons, i.e., they are within the interval. The length of the interval affects the number of similarities within the interval. The best results are given by applying high-pass filters with and without resizing (RHP and HP, respectively). Comparing all enhancement methods combined with resizing of face images does not affect the results to the extent that the processing with which it is combined has impacted.

The face identification in the actual application related to a criminal investigation is characterized by a disproportionate number of comparisons that give positives and negatives. We analysed the results from both perspectives separately in the experiment since sometimes it is crucial to be able to exclude possible suspects (TN) from further investigation, and in that manner, free occupied resources. The examples of the histograms shown in Fig. 7 and Fig. 8 were used for this purpose. For example, in Fig. 7, data from the bin in which $TNrate$ is equal to 100 % were

used, i.e., comparisons give correct results that they are different persons. In addition, the data from the first eight bins give a high degree of certainty that there are different persons on the matched images, over 99 %.

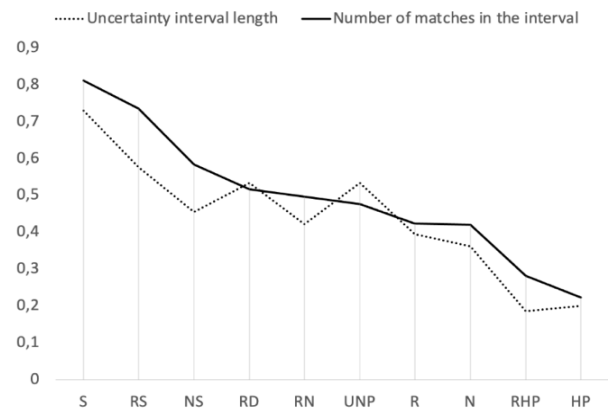


Fig. 12. The relationship between the length of the uncertainty interval and the number of similarities within the interval after applying different image enhancement methods.

We normalized data within the same database, and then the mean value was calculated for all four databases for each set of applied image enhancements. Figure 13 shows a chart of the impact of enhancement on effectiveness in the identification process where it can be argued that these are different persons.

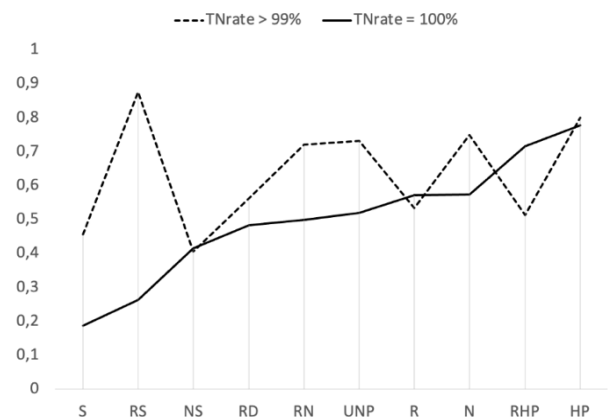


Fig. 13. The total ratio of the number of correctly classified negatives through all databases and applied the image enhancement method with a high $TNrate$.

The best results in claiming that these are different images with 100 % certainty are obtained by sharpening using a high-pass filter (HP, RHP). The influence of different image enhancement methods at $TNrate$ of 100 % fully corresponds to the comparison of the length of the uncertainty interval. When it comes to the certainty greater than 99 % that these are different persons, which also covers the most of the number of matching, the best results we still obtained by sharpening (RS, HP) and normalization.

Similarly, we compared the impact of different image enhancement methods on effectiveness in determining that it is most likely the same person (shown in Fig. 14).

In the classification where it can be claimed with complete certainty that the result of matching is the same person's face, the best results we obtained by applying sharpening with resizing face images of less than 150 pixels

and then by normalization. The large number of comparisons in which it can be claimed with a certainty of over 90 % that they are the same persons, which includes the most correctly performed classifications with positives, is obtained by applying normalization (RN, N).

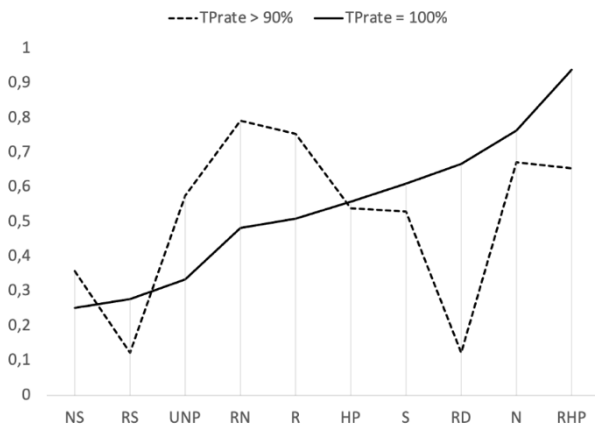


Fig. 14. The total ratio of the number of correctly classified positives through all databases and applied the image enhancement method with a high $TPrate$.

Finally, we performed the data classification from all databases using the appropriate unique weight value w for a particular image enhancement method. The scores for all three measures of effectiveness (Acc , $F1$, and MCC) are shown in Table II.

TABLE II. SCORES OF THE CLASSIFICATION PERFORMED USING THE WEIGHT VALUE.

| | \bar{g} | s | Acc | $F1$ | MCC |
|------------|-----------|--------|---------|---------|---------|
| HP | 0.9468 | 0.0046 | 0.9934↓ | 0.8156↓ | 0.8248↓ |
| N | 0.9457 | 0.0060 | 0.9943↑ | 0.8464↑ | 0.8513↑ |
| NS | 0.9467 | 0.0058 | 0.9930↓ | 0.8035↓ | 0.8129↓ |
| UNP | 0.9460 | 0.0042 | 0.9940= | 0.8336= | 0.8406= |
| RD | 0.9465 | 0.0031 | 0.9924↓ | 0.7784↓ | 0.8927↓ |
| R | 0.9468 | 0.0050 | 0.9940= | 0.8344↑ | 0.8413↑ |
| RHP | 0.9452 | 0.0047 | 0.9939↓ | 0.8319↓ | 0.8385↓ |
| RN | 0.9456 | 0.0051 | 0.9942↑ | 0.8397↑ | 0.8457↑ |
| RS | 0.9448 | 0.0040 | 0.9933↓ | 0.8138↓ | 0.8222↓ |
| S | 0.9441 | 0.0056 | 0.9937↓ | 0.8267↓ | 0.8322↓ |

The relationship of the effect made by applied image enhancement methods with the performed normalization of the results data is shown in Fig. 15.

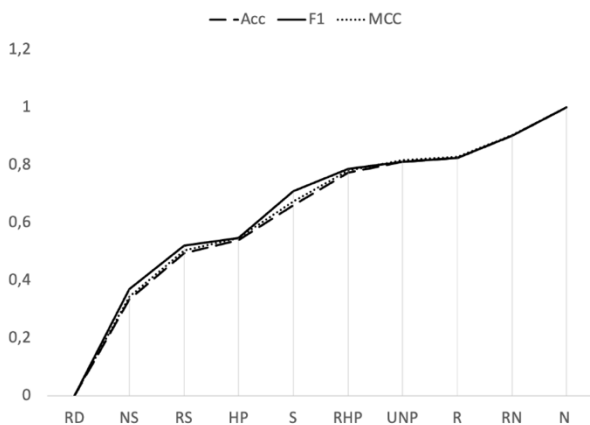


Fig. 15. The ratio of scores for classification by applying weight values.

The best score we obtained by applying normalization, first without and then with resizing. In contrast, although the

unprocessed images receive a high score at the applied weight value of w , resizing gives better results than purely unprocessed images. Regardless of the applied evaluation method, the scoring ratio is uniform.

V. CONCLUSIONS

In the criminal investigation, it is imperative to identify the perpetrator, but it is equally important (if not even more) to be able to exclude (with high certainty) the possible suspects. The consequences of a wrongly accused citizen can possibly decrease support to police work and create some kind of insecurity in society. Having that in mind, our research was conducted to establish whether it is possible to develop systems for assistance in a criminal investigation regarding identity resolution in unconstrained conditions.

This paper has experimentally demonstrated the influence of image enhancement on unconstrained face detection and identification. The databases we used are labeled, which made it possible to identify persons and measure the effects of image enhancement. The selection of databases simulates the actual conditions for analysing unconstrained face images by criminal investigators.

The main conclusions that can be drawn from the presented work are:

- A novel method that utilizes off-the-shelf solutions (the Dlib library HOG face detectors and the ResNet faces feature vector extractor), along with different image enhancement techniques, is presented to provide practical assistance to police investigators in unconstrained face identification;
- This method improves the effectiveness of the used Dlib HOG functions;
- The weight value w (classification threshold value) was experimentally determined to give optimal results, which provide that the length of the uncertainty interval decreases with the application of normalization ($F1$ coefficient increased by 1.5 percent, MCC increased by 1.3 percent);
- Experimenting on all four databases (images of different quality, size, and resolution) provides us a conclusion that resizing the image (especially with a resolution lower than 150 pixels) and sharpening should always precede the enhancement to improve face detection accuracy (the increase in accuracy is from 0.8 to 5.9 percent);
- Classification accuracy greater than 99 % is comparable to the state-of-the-art solutions [32]. The advantage of the proposed method is that it can be applied (and further improved) relying on the existing (limited) hardware and software resources within the police organization;
- In addition, the proposed enhancements provide images of an improved quality that makes it easier for criminal investigators to perform a visual inspection and make a decision based on the system's suggestion (prediction).

Our research has shown, that even if the police cannot afford state-of-the-art solutions, the utilization of off-the-shelf solutions followed by experimental research regarding the application of various image processing techniques can provide a real system that will greatly assist the police investigators.

In future work, we will analyse possible effectiveness improvements by choosing the enhancement method depending on the image's specific state. The results are expected to reduce further the uncertainty interval.

ACKNOWLEDGMENT

We thank our colleagues from the European Union's Horizon 2020 SPIRIT project consortium (Grant No. 786993) who provided insight and expertise that greatly assisted the research.

CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest.

REFERENCES

- [1] M. Taskiran, N. Kahraman, and C. E. Erdem, "Face recognition: Past, present and future (a review)", *Digital Signal Processing*, vol. 106, art. 102809, 2020. DOI: 10.1016/j.dsp.2020.102809.
- [2] U. Jayaraman, P. Gupta, S. Gupta, G. Arora, and K. Tiwari, "Recent development in face recognition", *Neurocomputing*, vol. 408, pp. 231–245, 2020. DOI: 10.1016/j.neucom.2019.08.110.
- [3] M. Pang, Y.-m. Cheung, B. Wang, and R. Liu, "Robust heterogeneous discriminative analysis for face recognition with single sample per person", *Pattern Recognition*, vol. 89, pp. 91–107, 2019. DOI: 10.1016/j.patcog.2019.01.005.
- [4] A. AB. Saabia, T. El-Hafeez, and A. M. Zaki, "Face recognition based on grey wolf optimization for feature selection, in *Proc. of the International Conference on Advanced Intelligent Systems and Informatics 2018. AISI 2018. Advances in Intelligent Systems and Computing*, vol. 845. Springer, Cham., 2019. DOI: 10.1007/978-3-319-99010-1_25.
- [5] N. H. Barnouti, "Improve face recognition rate using different image pre-processing techniques", *American Journal of Engineering Research*, vol. 5, no. 4, pp. 46–53, 2016.
- [6] L. An, M. Kafai, and B. Bhanu, "Dynamic Bayesian network for unconstrained face recognition in surveillance camera networks", *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 3, no. 2, pp. 155–164, 2013. DOI: 10.1109/JETCAS.2013.2256752.
- [7] Y. Xu, Z. Zhang, G. Lu, and J. Yang, "Approximately symmetrical face images for image preprocessing in Face recognition and sparse representation based classification", *Pattern Recognition*, vol. 54, pp. 68–82, 2016. DOI: 10.1016/j.patcog.2015.12.017.
- [8] L. Nanni, A. Lumini, and S. Brahnam, "Ensemble of texture descriptors for face recognition obtained by varying feature transforms and preprocessing approaches", *Applied Soft Computing*, vol. 61, pp. 8–16, 2017. DOI: 10.1016/j.asoc.2017.07.057.
- [9] N. Crosswhite, J. Byrne, C. Stauffer, O. Parkhi, Q. Cao, and A. Zisserman, "Template adaptation for face verification and identification", *Image and Vision Computing*, vol. 79, pp. 35–48, 2018. DOI: 10.1016/j.imavis.2018.09.002.
- [10] Q. Chen and L. Sang, "Face-mask recognition for fraud prevention using Gaussian mixture model", *J. Vis. Commun. Image R.*, vol. 55, pp. 795–801, 2018. DOI: 10.1016/j.jvcir.2018.08.016.
- [11] H. Han, S. Shan, X. Chen, and W. Gao, "A comparative study on illumination preprocessing in face recognition", *Pattern Recognition*, vol. 46, pp. 1691–1699, 2013. DOI: 10.1016/j.patcog.2012.11.022.
- [12] M. Haghghi, M. Abdel-Mottaleb, and W. Alhalabi, "Fully automatic face normalization and single sample face recognition in unconstrained environments", *Expert Systems with Applications*, vol. 47, pp. 23–34, 2016. DOI: 10.1016/j.eswa.2015.10.047.
- [13] H. T. Minh Nhat and V. T. Hoang, "Feature fusion by using LBP, HOG, GIST descriptors and canonical correlation analysis for face recognition", in *Proc. of 26th International Conference on Telecommunications (ICT)*, Hanoi, Vietnam, 2019, pp. 371–375. DOI: 10.1109/ICT.2019.8798816.
- [14] R. Rothe, R. Timofte, and L. Van Gool, "Deep expectation of real and apparent age from a single image without facial landmarks", *International Journal of Computer Vision*, vol. 126, no. 24, pp. 144–157, 2018. DOI: 10.1007/s11263-0160940-3.
- [15] A. S. Georghiadis, P. N. Belhumeur, and D. J. Kriegman, "From few to many: Illumination cone models for face recognition under variable lighting and pose", *IEEE Trans. Pattern Anal. Mach. Intelligence*, vol. 23, no. 6, pp. 643–660, 2001. DOI: 10.1109/34.927464.
- [16] M. Kawulok, M. E. Celebi, and B. Smolka (Eds.), *Advances in Face Detection and Facial Image Analysis*. Springer International Publishing, 2016, pp. 189–248. DOI: 10.1007/978-3-319-25958-1.
- [17] G. B. Huang, M. A. Mattar, T. L. Berg, and E. Learned-Miller, "Labeled faces in the wild: A database for studying face recognition in unconstrained environments", in *Proc. of the Workshop on Faces in 'Real-Life' Images: Detection, Alignment, and Recognition*, Erik Learned-Miller and Andras Ferencz and Frederic Jurie, Marseille, France, 2008.
- [18] ImageMagick. [Online]. Available: <https://imagemagick.org/script/index.php>
- [19] X. Dong, S. Kim, Z. Jin, J. Y. Hwang, S. Cho, and A. B. J. Teoh, "Open-set face identification with index-of-max hashing by learning", *Pattern Recognition*, vol. 103, 2020. DOI: 10.1016/j.patcog.2020.107277.
- [20] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection", in *Proc. of 2005 IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2005, pp. 886–893. DOI: 10.1109/CVPR.2005.177.
- [21] P. Kumar, S. L. Happy, and A. Routray, "A real-time robust facial expression recognition system using HOG features", in *Proc. of 2016 International Conference on Computing, Analytics and Security Trends (CAST)*, 2016, pp. 289–293. DOI: 10.1109/CAST.2016.7914982.
- [22] D. Manju and V. Radha, "A novel approach for pose invariant face recognition in a surveillance videos", *Procedia Computer Science*, vol. 167, pp. 890–899, 2020. DOI: 10.1016/j.procs.2020.03.428.
- [23] B. Li and G. Huo, "Face recognition using locality sensitive histograms of oriented gradients", *Optik*, vol. 127, no. 6, pp. 3489–3494, 2016. DOI: 10.1016/j.ijleo.2015.12.032.
- [24] A. Laucka, V. Adaskeviciute, D. Andriukaitis, "Research of the Equipment Self-Calibration Methods for Different Shape Fertilizers Particles Distribution by Size Using Image Processing Measurement Method," *Symmetry*, vol. 11, no. 7, p. 838, Jun. 2019. DOI: 10.3390/sym11070838.
- [25] R. Kapoor, R. Gupta, L. H. Son, S. Jha, and R. Kumar, "Detection of power quality event using histogram of oriented gradients and support vector machine", *Measurement*, vol. 120, pp. 52–75, 2018. DOI: 10.1016/j.measurement.2018.02.008.
- [26] B. Johnston and P. de Chazal, "A review of image-based automatic facial landmark identification techniques", *EURASIP J. Image Video Proc.*, art. no. 86, 2018. DOI: 10.1186/s13640-018-0324-4.
- [27] D. Gradolewski, D. Maslowski, D. Dziak, B. Jachimczyk, S. T. Mundlamuri, C. G. Prakash, and W. J. Kulesza, "A distributed computing real-time safety system of collaborative robot", *Elektronika ir Elektrotechnika*, vol. 26, no. 2, pp. 4–14, 2020. DOI: 10.5755/j01.eie.26.2.25757.
- [28] S. van der Walt *et al.*, "scikit-image: Image processing in Python", *PeerJ*, p. 2:e453, 2014. DOI: 10.7717/peerj.453.
- [29] C.-C. Wu *et al.*, "Prediction of fatty liver disease using machine learning algorithms", *Computer Methods and Programs in Biomedicine*, vol. 170, pp. 23–29, 2019. DOI: 10.1016/j.cmpb.2018.12.032.
- [30] Q. Zou, S. Xie, Z. Lin, M. Wu, and Y. Ju, "Finding the best classification threshold in imbalanced classification", *Big Data Research*, vol. 5, pp. 2–8, 2016. DOI: 10.1016/j.bdr.2015.12.001.
- [31] D. Chicco and G. Jurman, "The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation", *BMC Genomics*, vol. 21, art. no. 6, pp. 1–13, 2020. DOI: 10.1186/s12864-019-6413-7.
- [32] T. Gwyn, K. Roy, and M. Atay, "Face recognition using popular deep net architectures: A brief comparative study", *Future Internet*, vol. 13, no. 7, p. 164, 2021. DOI: 10.3390/fi13070164.



This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution 4.0 (CC BY 4.0) license (<http://creativecommons.org/licenses/by/4.0/>).