Adv Syst Sci Appl 2024; **01**; 142-162
Published online at https://ijassa.ipu.ru.

A New Face Swap Detection Technique for Digital Images

Rasha Thabit [1,2*], Heba Mohammed Fadhil [3], Hassan Falah Fakhruldeen [4,5], Akram Hatem Shather[6], Mohanad A. Al-Askari[7]

[1)] *Dijlah University College, Baghdad, Iraq*

[2)] *Al-Iraqia University, Baghdad, Iraq*

[3)] *University of Baghdad, Baghdad, Iraq*

[4)] *Imam Ja'afar Al-sadiq University, Baghdad, Iraq*

[5)] *University of Kufa, Kufa, Iraq*

[6)] *Al Kitab University, Altun Kopru, Kirkuk, Iraq*

[7)] *University of Al-Anbar, Al-Anbar, Iraq*

Abstract: In recent years, the rapid development of deep learning-based face image manipulation algorithms and applications became one of the challenges that are facing information forensics and information security systems. Using these applications, one can easily swap the face in a digital image with another face for different intentions where most of them are malicious intentions. Different face swap detection techniques have been presented in recent years to check the authenticity of the face in a digital image. Most of the available techniques are machine-learning or deep-learning based which makes them vulnerable to false detection results in addition to the time-consuming training process. In this paper, a new technique for face swap detection (FSD) is presented based on the image watermarking process. The proposed technique consists of two main algorithms called embedding and authentication algorithms. Several experiments have been conducted to evaluate the performance of the proposed technique and to prove its efficiency in detecting fake faces. The proposed technique outperforms various deep-learning-based techniques because no training is required and the detection accuracy is 100 %. The performance of the proposed FSD technique is promising therefore it is applicable in different practical applications.

Keywords: Information security, Information forensics, Deep-Fake detection, Face swap detection, Face image manipulation detection.

1. INTRODUCTION

Biotechnology and biometric information have been utilized in different verification systems such as security and financial industries [1]. The biometric information can be generated using extrinsic sources such as iris, fingerprint, and face or using intrinsic sources such as hand-vein, finger-vein, and palm-vein [2,3]. Over the years, the facial recognition systems [4] have been widely used to identify individuals or their feelings using images or videos processing techniques, however, identity theft and delivering fake information have been considered as challenges to these systems [5,6]. Face swap is a type of Deep-Fakes that refers to the process of replacing the face of one person with another in a digital image [7,8]. The face swap has several advantages in the movies and games industry [9], however, it has been considered as one of the most dangerous attacks that are required to be detected effectively especially when it is applied with malicious intentions. The fake face images can be used for fake videos and news to destroy the reputation of people, threaten them, identity theft, and many other malicious intentions [10–13]. Over the years, different face-swap algorithms and applications

* Corresponding author: rashathabit@yahoo.com

have been presented consequently different face-swap detection techniques have been implemented to serve the data security and digital media forensics fields [14–17]. The research community in the last few years witnessed an increased interest in face swap detection because these techniques became requested by many institutions and companies [17].

In [18], a face-swap detection technique based on machine learning has been presented. In this work, 83 landmarks for each face are extracted and used for training different classifiers. Different machine learning techniques have been tested such as support vector machine (SVM) [19], random forest (RF) [20], and multi-layer perceptron's (MLP) [21]. The images are classified into two classes that are innocent class and swapped images class. Two types of classifiers (i.e., linear and non-linear) have been tested to find the best. The results of the non-linear classifier were better than that of the linear classifier, however, both types of classifiers have recorded false detection results and the best accuracy that have been obtained was around 92 % for only a specific images dataset.

In [22], another face-swap detection technique has been presented based on deep learning and the error level analysis (ELA) process. The main principle of this technique is related to the errors that are generated in the manipulated face images. When a face region is cut and another face is pasted in its place, an error level will acquire between the manipulated area and the surrounding area. The Residual Network (ResNet-18) has been used with custom dense layers for the classification process. To reduce the time for training which may reaches days or weeks, the authors suggested the use of transfer learning. The results of this work were promising; however, false detection results have been also recorded and the best accuracy that have been obtained was around 97 % for only a specific images dataset.

In [23], a face-swap detection technique based on a convolution neural network (CNN) has been presented. The method of detection consists of two stages that are preprocessing stage and the classifier stage. In the preprocessing stage, the features of the face are extracted and the alignment process is applied [24, 25]. In the classifier stage, MobileNet-like CNN pre-trained on an ImageNet has been used [26, 27]. This technique obtained better accuracy results compared to previous techniques based on MesoNet and Xception networks [28, 29], however, false detection results have been obtained and the best accuracy that have been obtained was between 98 % to 99 % based on the images dataset. The main problems of this technique are the use of a specific images' dataset and the time complexity for training and testing processes.

As explained before, there are some limitations in the machine-learning and deep-learning-based face swap detection techniques such as the false detection results, the high time complexity for training and testing, the need for high-quality images for training, the need for large datasets, and others. To avoid these limitations, this paper presents a new face-swap detection (FSD) technique based on image watermarking. The main idea of the proposed technique is inspired by the Region-of-Interest (ROI) based medical image authentication techniques [30–32]. In this paper, the face area will be considered as ROI and its authentication data will be extracted and embedded in the area outside ROI. The data generation, embedding, extraction, and authentication processes are all based on the watermarking techniques that have been presented in [33–35]. To detect the face area two face detection algorithms are tested to choose the one with the best performance. The proposed technique can distinguish between innocent and fake faces in digital images without the need for training and with 100 % accuracy thus it outperforms the machine learning and deep-learning-based techniques. The proposed technique can be applied for any image regardless its quality which increased its ability in practical applications.

The rest of the paper is organized as follows: in section 2, the related works are presented; in section 3, the algorithms of the proposed FSD technique are explained in details; in section 4, the experiments and their discussion are presented; and finally, section 5 illustrates the conclusions that have been drawn from this research.

2. RELATED WORKS

The proposed FSD technique starts by detecting the face area in the digital image. To detect the face region two widely used face detection algorithms have been tested to choose the one with the best performance. The face detection algorithms are briefly explained in the following subsection. On the other hand, inspired by the medical image authentication technique presented in [34], the proposed FSD adopted Slantlet transform (SLT)-based watermarking. SLT-based watermarking techniques [33], [35–39] proved their efficiency in various applications, therefore, this type of watermarking has been adopted as part of the implemented algorithms in the proposed FSD technique. The second subsection presents the SLT-based embedding and extraction algorithms for a block of image.

2.1. Face Detection

Two well-known face detection algorithms are tested to choose the best one in order to be adopted as the first part of the proposed FSD technique. The AdaBoost-based algorithm [40] and Multi-Task Cascaded CNN (MTCNN)-based algorithm [41] are tested for different face images. In [40], the detection algorithm starts by calculating the integral image for the input image followed by extracting the features using the Harr-like filters. Then a small number of the generated features is selected using the AdaBoost algorithm. To extract the promising regions of the image, the cascade structure is used for the complex classifiers. After several processing steps, the face regions are selected. In [41], a joint face detection and alignment algorithm is presented in which a shallow CNN algorithm is applied to generate the candidate windows. Then a more complex CNN algorithm is used to reject the non-face windows. Thereafter, another robust CNN algorithm is applied as the final touch-up to refine the result and generate the landmark positions. Samples of the preliminary tests for the abovementioned algorithms are shown in Fig. 2.1. The results proved that the MTCNN-based algorithm performs better in comparison with the AdaBoost-based algorithm where the latter has missed some of the faces. Therefore, the MTCNN-based algorithm has been chosen to be applied as the first stage in the proposed FSD technique.

2.2. SLT-Based Watermarking

The SLT-based watermarking algorithms have been applied in different applications and they proved their efficiency in comparison with different other algorithms in terms of visual quality, robustness, and time complexity [35], [42–44]. Based on the previous studies, we suggested the used of SLT-based watermarking in face swap detection algorithms. In the proposed FSD technique, the SLT-based watermark embedding and extraction algorithms for a single image block have been adapted from [34]. The adopted SLT-based embedding and extraction algorithms are explained in Table 2.1 and Table 2.2, respectively.

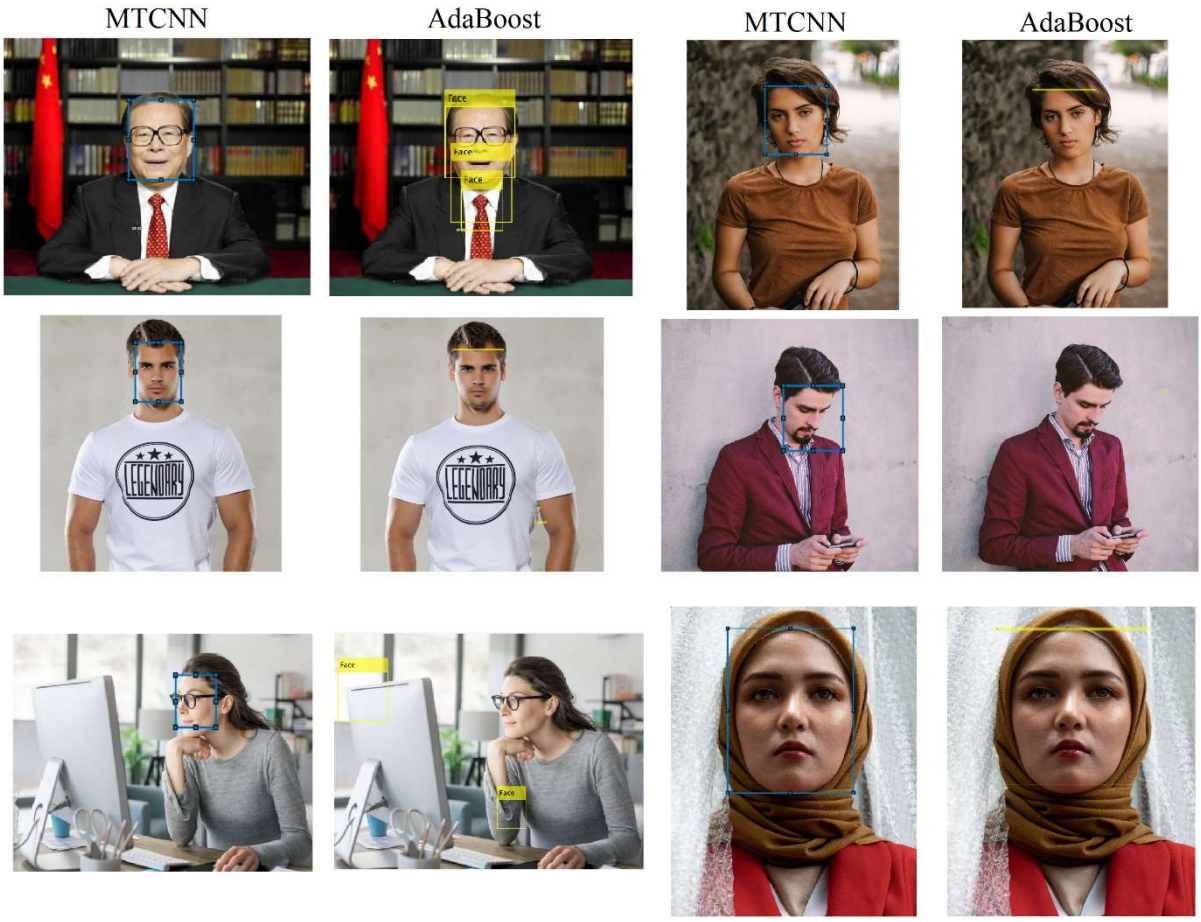

Fig. 2.1. Preliminary test results for MTCNN-based and AdaBoost-based face detection algorithms.

Table 2.1. SLT-based watermark embedding algorithm for a single block [34].

Input: Original image block (size 16×16 pixels) and binary sequence of 64 bits. **Output:** Watermarked image block (size 16×16 pixels).	
Step 1	Read the input image block B and the binary sequence Bin_{seq}.
Step 2	Transform B using SLT matrix as follows: $T_B = [SLT_{16}]\,[B]\,[SLT_{16}^T]$ Where B and T_B are the original and watermarked blocks, SLT_{16} and SLT_{16}^T are the SLT matrix and its transpose both of size (16×16).
Step 3	Divide the coefficients in T_B into four subbands called $(LL, HL, LH, and\ HH)$. $LL = T_B(1:8, 1:8)$ $HL = T_B(1:8, 9:16)$ $LH = T_B(9:16, 1:8)$ $HH = T_B(9:16, 9:16)$ Where $LL, HL, LH, and\ HH$ are Low-Low, High-Low, Low-High, and High-High subbands, respectively.
Step 4	For $x = 1\ to\ 64$ $b = Bin_{seq}(x)$ For $i = 1\ to\ 8$ For $j = 1\ to\ 8$ $D1 = HL(i,j) - LH(i,j)$ If $b = 1$ and $D1 < Threshold$, then increase $HL(i,j)$ and decrease $LH(i,j)$ as follows: $NewHL(i,j) = HL(i,j) + \frac{Threshold - D1}{2}$ $NewLH(i,j) = LH(i,j) - \frac{Threshold - D1}{2}$

	If $b = 1$ and $D1 \geq Threshold$, then do-nothing. $D2 = LH(i,j) - HL(i,j)$ If $b = 0$ and $D2 < Threshold$, then increase $LH(i,j)$ and decrease $HL(i,j)$ as follows: $NewHL(i,j) = HL(i,j) - \frac{Threshold-D}{2}$ $NewLH(i,j) = LH(i,j) + \frac{Threshold-D2}{2}$ If $b = 0$ and $D2 \geq Threshold$, then do-nothing. Note: The $Threshold$ variable is used for controlling the visual quality and the robustness of the embedded watermark. The $Threshold$ value that has been adopted in [34] is 3 because it gives a good compromise between visual quality and robustness.
Step 5	Replace the original HL and LH subbands in T_B with $NewHL$ and $NewLH$ subbands. Save the resultant matrix as $NewT_B$.
Step 6	Apply inverse SLT on $NewT_B$ to obtain the watermarked image block as follows: $W_B = [SLT_{16}^T]\ [NewT_B]\ [SLT_{16}]$ Where W_B is the watermarked image block that carries a binary sequence of 64 bits.

Table 2.2. SLT-based watermark embedding algorithm for a single block [34].

Input: Original image block (size 16×16 pixels) and binary sequence of 64 bits.. **Output:** Watermarked image block (size 16×16 pixels).	
Step 1	Read the input watermarked image block W_B.
Step 2	Transform W_B use the ng SLT matrix as follows: $T_{WB} = [SLT_{16}]\ [W_B]\ [SLT_{16}^T]$ Where T_{WB} is the transformed watermarked image block.
Step 3	Divide the coefficients in T_{WB} into four subbands called $(LL, HL, LH, and\ HH)$. $LL = T_{WB}(1:8, 1:8)$ $HL = T_{WB}(1:8, 9:16)$ $LH = T_{WB}(9:16, 1:8)$ $HH = T_{WB}(9:16, 9:16)$
Step 4	Let $x = 1$ For $i = 1\ to\ 8$ For $j = 1\ to\ 8$ $b(x) = 1\ when\ HL(i,j) \geq LH(i,j)$ $b(x) = 0\ when\ LH(i,j) > HL(i,j)$ $x = x + 1$ The loop continues until extracting a binary sequence of length 64 bits.

3. PROPOSED FSD TECHNIQUE

The proposed FSD technique consists of two main algorithms called embedding and authentication algorithms. The embedding algorithm is applied at the sender side to detect the face area, generate its authentication information and to hide this information in the region outside the face area. The authentication algorithm is applied at the receiver side in which the embedded information is extracted and compared with the information generated from the received face area to ensure the authenticity of the face in the digital image. When the compared information is not matched the face is classified as unauthentic and the manipulated region is localized in the received image. The following subsections illustrate the details of the proposed embedding and authentication algorithms.

3.1. Proposed Embedding Algorithm

The embedding algorithm of the proposed FSD technique can be summarized as illustrated in Fig. 3.1. The algorithm starts by reading the original (Red, Green, and Blue) RGB color face image $I_o(:,:,i)$ where $i = 1,2,3$ which refers to the R, G, and B channels of the input image, respectively. In order to detect the face box, the MTCNN is applied to I_o as explained in section 2.1.

The output of the MTCNN algorithm is an array contains four readings (y,x,w,h) where (x,y) is the top left corner of the detected face's box, w is the width of the box, and h is the height of the box. To define the detected box in terms of pixels' positions, the top left corner and the bottom right corner must be defined as integers. The readings (y,x,w,h) are rounded to their nearest integer numbers that are greater than or equal to the values. Let the resultant array after round is (y_r,x_r,w_r,h_r), the top left corner of the face box is (x_r,y_r) and the bottom right corner of the face box (x_r+h_r,y_r+w_r). Thus, the pixels' positions of the face box can be defined as $(x_r:x_r+h_r,y_r:y_r+w_r)$. A binary mask image I_M is generated according to the obtained positions of pixels that are related to the face box. The following procedure is conducted to generate I_M:

- Read the size of I_o $(M \times N \times C)$, where $M = height\ (I_o)$, $N = width\ (I_o)$, and $C = Number\ of\ channels\ in\ (I_o)$;
- Generate a binary image of zeros with size $(M \times N)$;
- Convert the pixels at the positions $(x_r:x_r+h_r,y_r:y_r+w_r)$ to ones;
- Save the resultant binary image as I_M.

After generating the mask image, each channel from I_o can be watermarked using the following procedure:

- Read the channel image $I_o(:,:,i)$ where i refers to the level of the channel.
- Divide the channel image and I_M into non-overlapping blocks each of size (16×16) pixels.
- Classify the blocks of the channel image as shown in Fig. 3.2 where the mean value of the pixels (μ) in each I_M block is calculated to classify the blocks into two groups (i.e., Face blocks and Non-Face blocks).
- Generate the authentication information from the 'Face blocks' by calculating the mean value of the pixels for each block followed by rounding the result to the nearest integer value. The resultant values are converted to binary sequences and concatenated to generate one binary sequence which must be embedded in the 'Non-Face blocks'.
- To increase the robustness of the embedded sequence, apply BCH (11,15) encoding. To embed binary sequence in 'Non-Face blocks', the sequence must be divided into non-overlapping subsequences each of length 64 bits. In order to prepare the binary sequence bch_{seq} for the embedding, the length of the sequence must be divisible by 64. The following steps are applied to prepare the binary subsequences for embedding:
    - $Rem = length\ (bch_{seq})/64$;
    - If $Rem = 0$ then $Bin_{seq} = bch_{seq}$;
    - Else $Extend = 64 - Rem$, $Extend_{seq} = zeros(1:Extend)$;
    - $Bin_{seq} = [bch_{seq}, Extend_{seq}]$.
- Apply SLT watermarking algorithm (explained in section 2.2) to embed the binary subsequence of Bin_{seq} in 'Non-Face blocks'. The resultant watermarked 'Non-Face blocks' and the original 'Face-blocks' are used to construct the watermarked channel image.

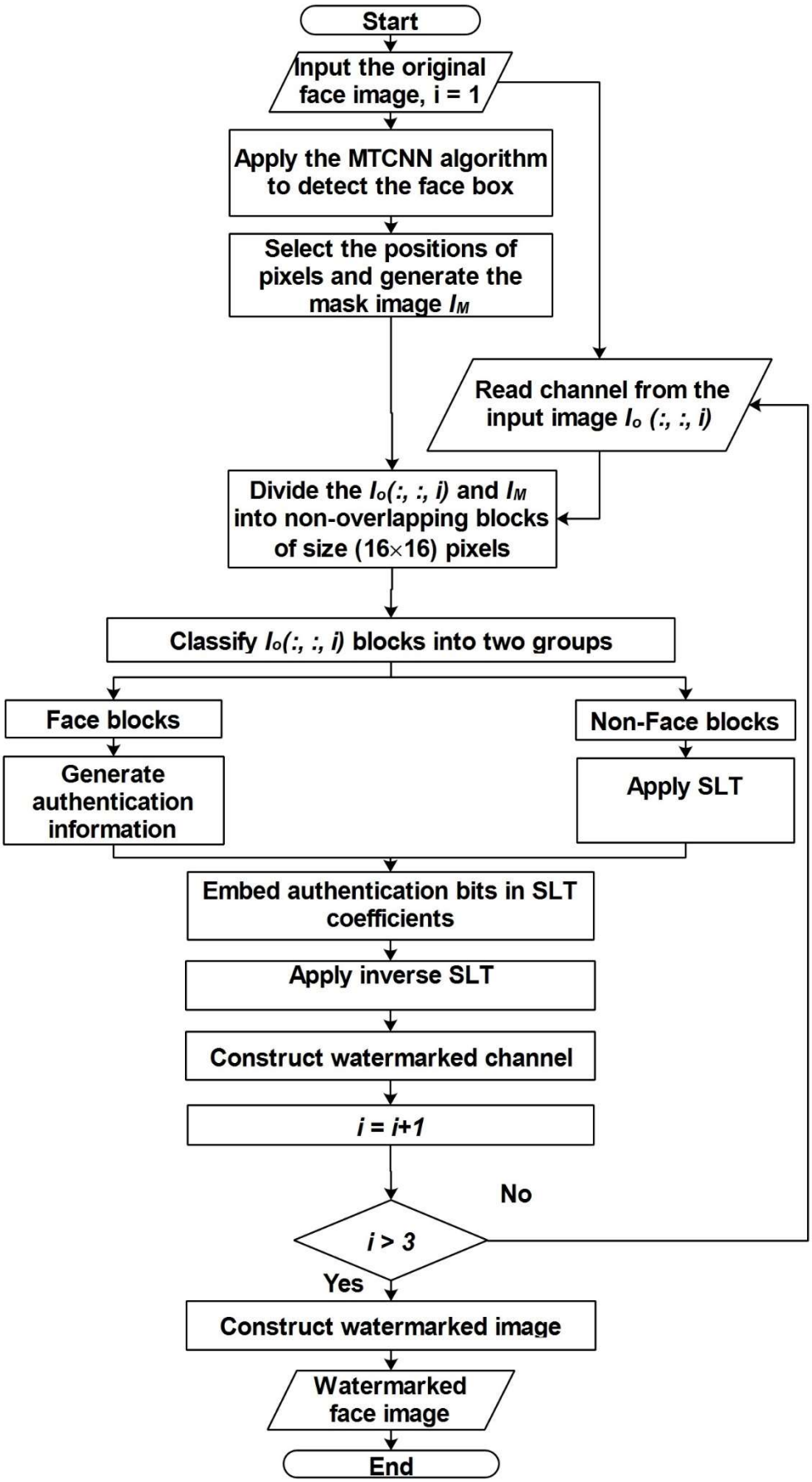

Fig. 3.1. The embedding algorithm of the proposed FSD technique.

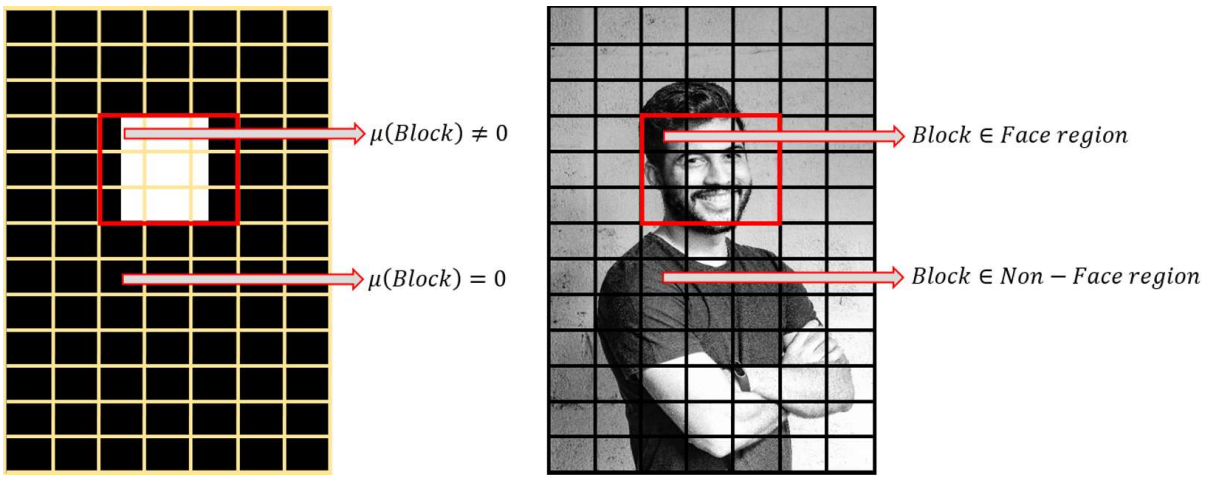

Fig. 3.2. Classification of channel image blocks into two groups.

The procedure of watermarking one channel image is repeated to generate the watermarked channel images which are used to construct the resultant watermarked face image I_W. The watermarked image is sent to the receiver side at which the authentication algorithm must be applied to ensure the safety of the face image and to reveal any manipulation in the face region if exists. The proposed FSD technique doesn't need the original image or the authentication information at the receiver side which makes the technique completely blind and it can reveal manipulations using the received watermarked image only.

3.2. Proposed authentication algorithm

The authentication algorithm of the proposed FSD technique can be summarized as illustrated in Fig. 3.3. The algorithm starts by reading the watermarked face image I_W and applying the MTCNN algorithm to detect the face box as explained in the embedding procedure. The pixels specification, mask image I_M generation, and blocks classification procedure are the same as those which have been explained at the embedding side.

The following procedure is repeated to extract the embedded authentication data and to calculate the authentication data from the 'face blocks' in the received I_W:

- Read channel image $I_W(:,:,i)$ from the received I_W and read the generated I_M.
- Divide $I_W(:,:,i)$ and I_M into non-overlapping blocks of size (16×16).
- Classify the blocks into two groups 'Face blocks' and 'Non-Face blocks' as explained in the embedding procedure.
- Calculate the new authentication data from the received 'Face blocks'.
- Extract the embedded authentication data using SLT extraction algorithm (explained in section 2.2) from 'Non-Face blocks'.

As shown in Fig. 3.3, the extracted authentication data and the calculated authentication data are compared to check the authenticity of the 'Face blocks'. The block is considered authentic when the compared data are identical. If the compared data are not identical, then the 'Face block' is considered not authentic and a border is drawn on the block to localize it in the face image. The face image is considered authentic only when all the blocks in the face region are authentic.

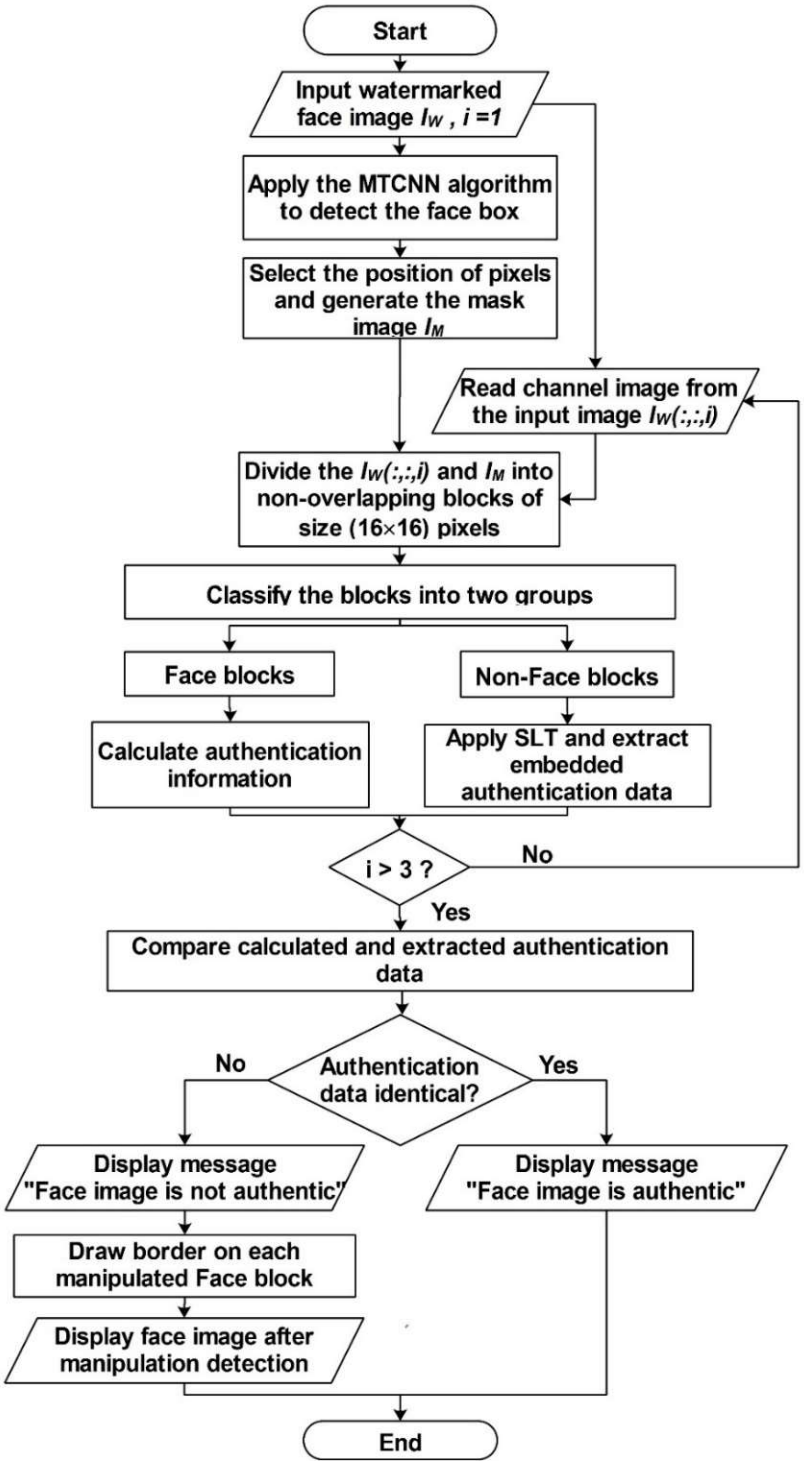

Fig. 3.3. Authentication algorithm of the proposed FSD technique.

4. RESULTS AND DISCUSSION

To test the performance of the proposed FSD technique, the experiments have been conducted for color face images with different sizes which have been collected using Google Image Search from different websites such as [45–48]. Samples of the test images are shown in Fig. 4.1.

Fig. 4.1. Sample test images.

The first experiment has been conducted to ensure the accuracy of the proposed FSD technique in detecting the face's pixels and generating the mask image. To ensure the safety of the face images after watermarking, the visual quality of the watermarked face images has been tested. The ability of the proposed FSD technique in detecting fake faces has been evaluated. Experiments have been conducted to test the capacity and its relationship with the number of 'Non-Face blocks'. The experiments also include test of payload and its relationship with the number of 'Face blocks'.

The following subsections present the experimental results followed by a general comparison between the proposed FSD technique and the previous deep-learning based face swap detection techniques.

4.1. Accuracy of Mask Image Generation

The mask image generation process depends on the accuracy of selecting the pixels that are related to the face region as illustrated in section 3. The accuracy of generating the mask image has been tested for different test images before proceeding to other tests. The experimental results proved that the proposed technique can accurately select the positions of face's pixels and generate the mask image without errors. Samples of the MTCNN results and their related mask images are shown in Fig. 4.2.

Fig. 4.2. Sample results of mask image generation test.

4.2. Visual Quality Test

The visual quality of the resultant watermarked images after watermark embedding using the proposed FSD technique must be tested to ensure that there are no visual artifacts in the image, no errors in rearranging the blocks, and no errors in constructing the watermarked images. The first test is conducted by displaying the original face image and its resultant watermarked image side by side to check differences. Samples of this test are shown in Fig. 4.3. The results proved that the watermarked images are unscathed and there are no visual artifacts in the image. The second test has been conducted to calculate the Peak Signal-to-Noise Ratio (PSNR) in dB and the Mean Squared Error (MSE) between the original image and the watermarked image. The experiments have been conducted for the test images shown in Fig. 4.1 and the results are shown in Table 4.1. The results proved that the visual quality of the watermarked image depends on the ratio of the face's blocks to the non-face's blocks and depends also on the contents of the image where some images require less changes to carry the authentication bits while others require more changes in the image contents. The PSNR results shown in Table 4.1, illustrate the efficiency of the proposed FSD technique in generating watermarked images with high visual quality.

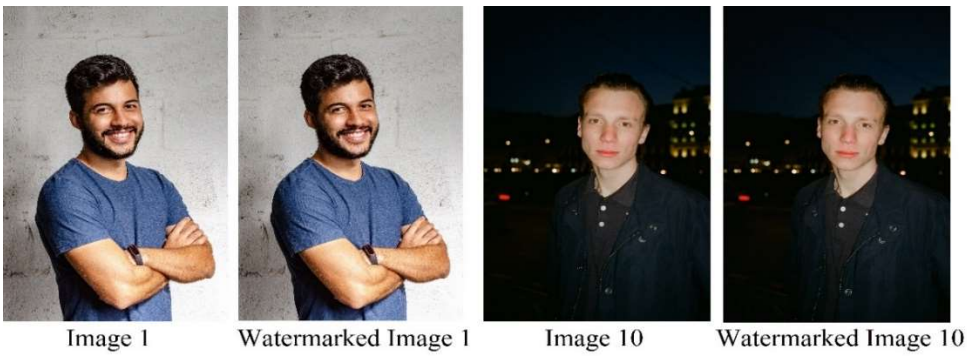

Fig. 4.3. Sample watermarked images.

Table 4.1. Visual quality test results.

Image name	Image size	Size of face area	MSE	PSNR (dB)
Image 1	3000×1987×3	668×533×3	0.313	+53.17
Image 2	7952×5304×3	1452×1214×3	0.0181	+65.56
Image 3	5472×3648×3	1206×944×3	0.0679	+59.81

Image 4	3888×2592×3	1206×944×3	0.0496	+61.17
Image 5	3456×5184×3	2029×1677×3	0.042	+61.90
Image 6	5075×5760×3	2453×1955×3	0.0761	+59.32
Image 7	3008×2008×3	409×338×3	0.003	+73.42
Image 8	6240×4160×3	1037×885×3	0.0097	+68.28
Image 9	2395×2395×3	1096×913×3	0.0763	+59.31
Image 10	3027×2007×3	779×625×3	1.129	+47.60
Image 11	5599×3733×3	1461×1277×3	0.4056	+52.05
Image 12	3442×2295×3	1277×998×3	0.0426	+61.84
Image 13	4090×7360×3	1291×985×3	0.1367	+56.77
Image 14	6000×4000×3	1203×1024×3	0.0144	+66.56
Image 15	2620×2096×3	1383×1076×3	0.4011	+52.10

4.3. Fake Faces Detection Test

To test the ability of the proposed FSD technique in detecting fake faces in digital images, face swap attack has been imposed on the watermarked face images using Ps Adobe Photoshop version (21.2.1). The results of this experiment proved the efficiency of the proposed FSD technique in revealing the fake face in the digital image and localizing the manipulated part in the face region. Samples of the results are shown in Fig. 4.4– Fig. 4.9. The proposed FSD technique detects fake faces for all test images without error which makes the accuracy 100 % regardless the quality of the test images.

Fig. 4.4. Fake face detection using the proposed FSD technique for ‘Image 2’.

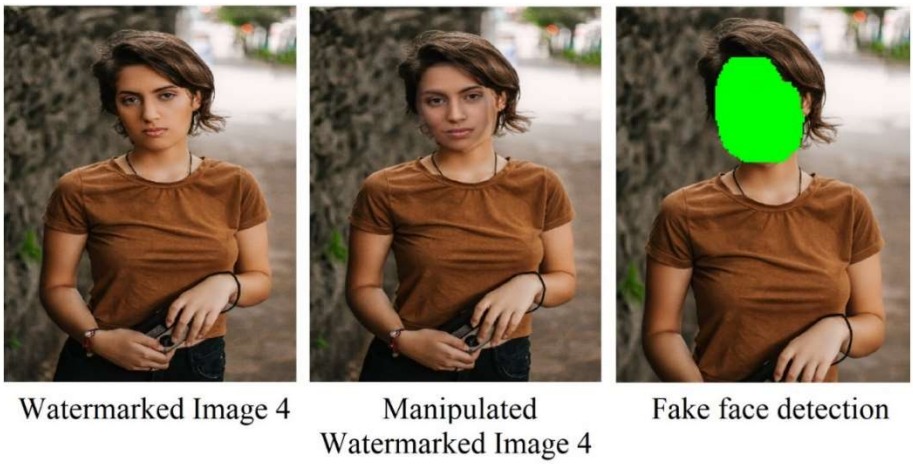

Fig. 4.5. Fake face detection using the proposed FSD technique for 'Image 4'.

Fig. 4.6. Fake face detection using the proposed FSD technique for 'Image 6'.

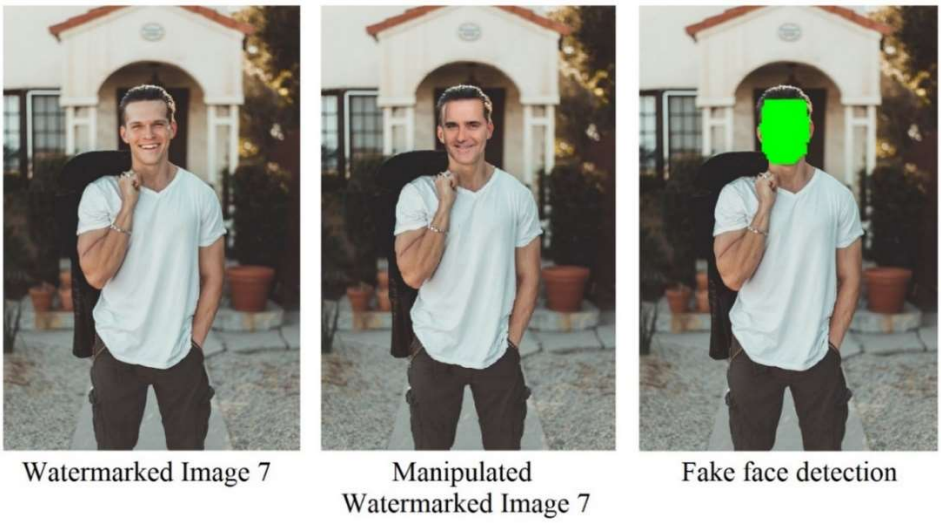

Fig. 4.7. Fake face detection using the proposed FSD technique for 'Image 7'.

Fig. 4.8. Fake face detection using the proposed FSD technique for 'Image 13'.

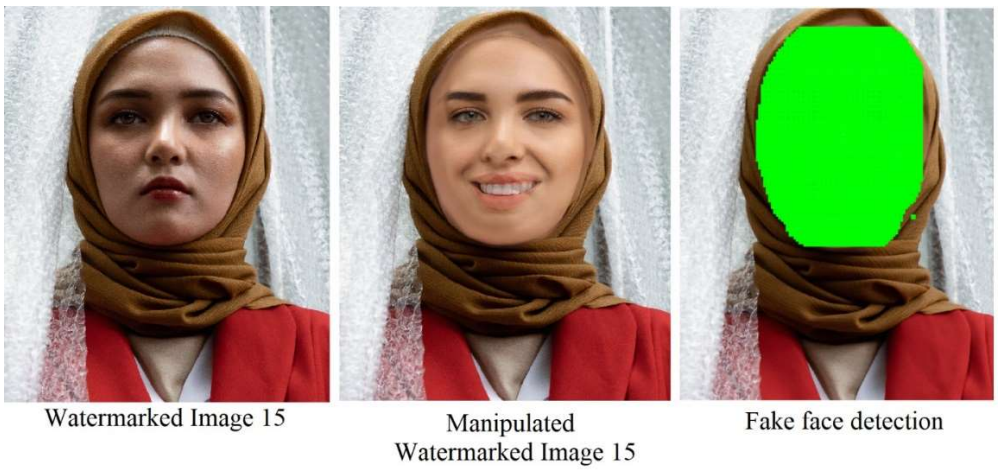

Fig. 4.9. Fake face detection using the proposed FSD technique for 'Image 15'.

4.4. Embedding Capacity Test

The embedding capacity of the proposed FSD technique depends on the size of the image and the size of the face region. As mentioned in the embedding procedure, each (16×16) block from the 'Non-Face blocks' can carry 64 bits which is based on the adopted SLT watermarking technique (explained in section 2). The number of the 'Non-Face blocks' in one channel is multiplied by 3 to calculate total number of the 'Non-Face blocks' in the image. Then the total number of the 'Non-Face blocks' is multiplied by 64 bits to calculate the total embedding capacity of the image. The results of this test are shown in Table 4.2 and the relationship between the total embedding capacity and number of 'Non-Face blocks' is illustrated in Fig. 4.10. The results proved that the larger the number of 'Non-Face blocks', the higher embedding capacity and vice versa.

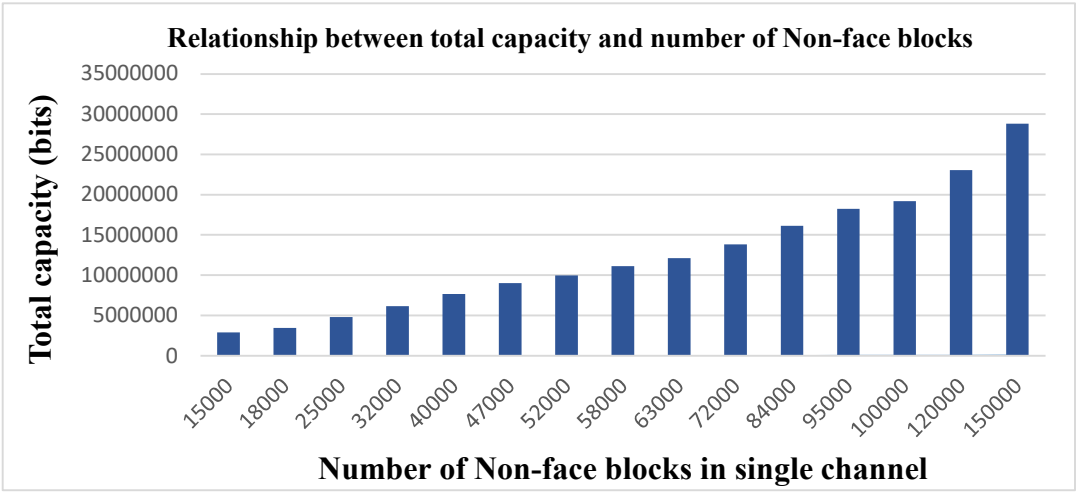

Fig. 4.10. Relationship between total capacity and number of 'Non-face blocks'.

Table 4.2. Embedding capacity test results.

Image name	Image size	Size of face area	No. of NFB [1] in single channel	No. NFB [1] in three channels	Total capacity (bits)
Image 1	3000×1987×3	668×533×3	21683	65049	4163136
Image 2	7952×5304×3	1452×1214×3	157500	472500	30240000
Image 3	5472×3648×3	1206×944×3	73416	220248	14095872
Image 4	3888×2592×3	1206×944×3	36720	110160	7050240
Image 5	3456×5184×3	2029×1677×3	56416	169248	10831872
Image 6	5075×5760×3	2453×1955×3	95178	285534	18274176

Image 7	3008×2008×3	409×338×3	22928	68784	4402176
Image 8	6240×4160×3	1037×885×3	97704	293112	18759168
Image 9	2395×2395×3	1096×913×3	18141	54423	3483072
Image 10	3027×2007×3	779×625×3	21625	64875	4152000
Image 11	5599×3733×3	1461×1277×3	73957	221871	14199744
Image 12	3442×2295×3	1277×998×3	25705	77115	4935360
Image 13	4090×7360×3	1291×985×3	112278	336834	21557376
Image 14	6000×4000×3	1203×1024×3	88810	266430	17051520
Image 15	2620×2096×3	1383×1076×3	15437	46311	2963904

[1] No. of NFB = Number of 'Non-Face blocks'.

4.5. Payload Test

The payload in each channel from the input image refers to the total number of bits that are generated in Bin_{seq} as explained in (subsection 3.1). The generated payload depends on the size of the face region. As mentioned in the embedding procedure, the authentication information is generated from the 'Face blocks', converted to binary, BCH coded, and prepared for embedding in 'Non-Face blocks'. The results of payload test are shown in Table 4.3 and the relationship between the total payload and number of 'Face blocks' is illustrated in Fig. 4.11. The results proved that the larger the number of 'Face blocks', the higher payload and vice versa.

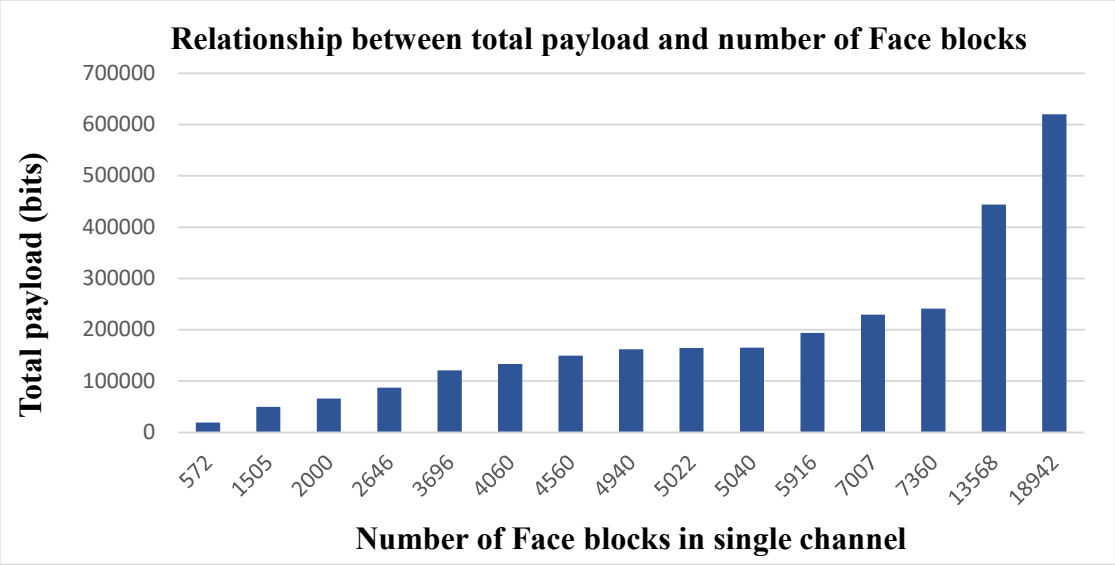

Fig. 4.11. Relationship between total payload and number of 'Face blocks'.

Table 4.3. Payload test results.

Image name	Image size	Size of face area	No. of FB in single channel	Length of Bin_{seq} in single channel	Total payload (bits)
Image 1	3000×1987×3	668×533×3	1505	16448	49344
Image 2	7952×5304×3	1452×1214×3	7007	76480	229440
Image 3	5472×3648×3	1206×944×3	4560	49792	149376
Image 4	3888×2592×3	1206×944×3	2646	28928	86784
Image 5	3456×5184×3	2029×1677×3	13568	148032	444096
Image 6	5075×5760×3	2453×1955×3	18942	206656	619968
Image 7	3008×2008×3	409×338×3	572	6272	18816
Image 8	6240×4160×3	1037×885×3	3696	40384	120960

Image 9	2395×2395×3	1096×913×3	4060	44352	133056
Image 10	3027×2007×3	779×625×3	2000	21888	65664
Image 11	5599×3733×3	1461×1277×3	7360	80320	240960
Image 12	3442×2295×3	1277×998×3	5040	55040	165120
Image 13	4090×7360×3	1291×985×3	5022	54848	164544
Image 14	6000×4000×3	1203×1024×3	4940	53952	161856
Image 15	2620×2096×3	1383×1076×3	5916	64576	193728
[1] No. of FB = Number of 'Face blocks'					

4.6. Time Complexity

The execution time required for the embedding and authentication procedure is very important in practical applications, therefore, the run-time of the proposed FSD algorithms have been calculated for different test images. Since the run-time is affected by the hardware and software used in the experimental tests it is useful to mention that the computer used in the experimental tests has 1.80 GHz Intel® Core TM i7 CPU and 16 GB memory. The software used is MATLAB (R2020a) and the commands used for this experiment are tic and toc commands. The results of this experiment are shown in Table 4.4 which proved the efficiency of the proposed FSD technique in executing embedding and authentication algorithms in seconds. The execution time of the algorithms depends on the size of the face image and the total payload. The higher the payload, the longer the execution time. As shown in the results, the execution time for authentication is lower than that for the embedding which makes the proposed FSD technique suitable for the practical application that require face authentication before proceeding to other processing steps. The test for small size images gives much lower execution time, for instance an image of size (164×307×3) and payload (1024) required (0.31235) seconds for embedding and (0.2701) seconds for authentication.

4.7. Comparison with Previous Techniques

As mentioned in the introduction section, the false detection, high time complexity for training and testing, the need for high-quality images for training, and the need for large datasets are some of the limitations in the face swap detection methods that are based on machine learning and deep learning algorithms. Since the proposed FSD techniques adopted digital watermarking, there is no need for training and it can be applied for any face image regardless its quality. The accuracy of the proposed technique is 100 % and there are no false detection results thus it outperforms the techniques in [18,22,23] which have recorded accuracies around [92 %, 97 %, 98 % to 99 %] for only specific datasets. The results of the techniques in [18,22,23] can be further degraded when the test images are different from the datasets used in the training process while the results of the proposed FSD technique are always accurate.

Table 4.4. Time complexity test results.

Image name	Image size	Size of face area	Total payload (bits)	Embedding time (sec.)	Extraction time (sec.)
Image 1	3000×1987×3	668×533×3	49344	3.3191	1.8827
Image 2	7952×5304×3	1452×1214×3	229440	24.6058	14.1453
Image 3	5472×3648×3	1206×944×3	149376	12.1752	7.3814
Image 4	3888×2592×3	1206×944×3	86784	6.4060	3.5111
Image 5	3456×5184×3	2029×1677×3	444096	12.1816	8.7077
Image 6	5075×5760×3	2453×1955×3	619968	15.9596	12.1469
Image 7	3008×2008×3	409×338×3	18816	2.9516	1.0532
Image 8	6240×4160×3	1037×885×3	120960	14.4640	7.1083

Image 9	2395×2395×3	1096×913×3	133056	5.4332	3.2164
Image 10	3027×2007×3	779×625×3	65664	3.8354	2.4048
Image 11	5599×3733×3	1461×1277×3	240960	12.1705	11.3076
Image 12	3442×2295×3	1277×998×3	165120	5.4606	4.0822
Image 13	4090×7360×3	1291×985×3	164544	17.5745	10.5852
Image 14	6000×4000×3	1203×1024×3	161856	14.5385	7.9509
Image 15	2620×2096×3	1383×1076×3	193728	3.6695	2.3317

5. CONCLUSIONS

The rapid development of new technology and the spread of easy-to-use applications can be considered as double-edged sword where these applications can be used for innocent or malicious intentions. Recently, several applications have been introduced to swap the faces in digital images which can be adopted for identity theft, threatening, destroying the reputation, spreading fake news, and many others. To authenticate the face image, the research community presented different face swap detection techniques based on machine-learning and deep-learning. There are some limitations in these techniques such as the false detections results, the long time required for training and testing, the need for large datasets for training, the need for high quality images to obtain better detection results, etc. In this paper, a new face swap detection (FSD) technique is presented based on images watermarking technology.

The proposed FSD technique have two main algorithms each of them starts by the proposed steps to detect the face region in the image. The embedding algorithm is applied to generate and hide the authentication information while the authentication algorithm is applied to extract and authenticate the face information in the received image. Experiments have been conducted to evaluate the performance of the proposed FSD technique for different test images. The results proved that capacity and payload depend on the size of the image and the size of the face region. The larger the size of the face region, the higher the payload. The embedding capacity increases with the increment in the ratio of the number of blocks outside face region to the number of blocks belong to face region. The subjective and objective evaluation of the visual quality proved the efficiency of the proposed FSD technique where high quality watermarked images are generated without any visible distortions. The proposed FSD technique can effectively detect fake face in the image and the accuracy is 100 %. The execution time of the algorithms is low even for large size images which makes the proposed technique suitable for practical applications. For the future work, different watermarking techniques can be applied and compared with the proposed FSD technique in the aim of further improving the performance.

FUNDING

This research received no external funding.

ACKNOWLEDGEMENTS

The authors would like to thank their institutions for encouraging and supporting their scientific researches.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

REFERENCES

1. Zhao, Y., Wu, M., Zhang, L., Wang, J. & Wei, D. (2018). An Effective Feature Segmentation Algorithm for a Hyper-Spectral Facial Image, *Information*, **9**(10), 261. https://doi.org/10.3390/info9100261

2. Qin, H. & Wang, P. (2019). A Template Generation and Improvement Approach for Finger-Vein Recognition. *Information*, **10**(4), 145. https://doi.org/10.3390/info10040145

3. Ammar, S., Bouwmans, T. & Neji, M. (2022). Face Identification Using Data Augmentation Based on the Combination of DCGANs and Basic Manipulations, *Information*, **13**(8), 370, https://doi.org/10.3390/info13080370

4. Zhi, J., Song, T., Yu, K., Yuan, F., Wang, H., et al. (2022). Multi-Attention Module for Dynamic Facial Emotion Recognition, *Information*, **13**(5), 207, https://doi.org/10.3390/info13050207

5. Semenkov, A., Bragin, D., Usoltsev, Y., Konev, A. & Kostuchenko, E. (2021). Generation of an EDS Key Based on a Graphic Image of a Subject's Face Using the RC4 Algorithm, *Information*, **12**(1), 19, https://doi.org/10.3390/info12010019

6. Figueroa, A., Peralta, B. & Nicolis, O. (2021). Coming to Grips with Age Prediction on Imbalanced Multimodal Community Question Answering Data, *Information*, **12**(2), 48, https://doi.org/10.3390/info12020048

7. Bitouk, D., Kumar, N., Dhillon, S., Belhumeur, P. & Nayar, S. K. (2008). Face swapping: automatically replacing faces in photographs, *ACM Transactions on Graphics*, **27**(3), 1–8, https://doi.org/10.1145/1360612.1360638

8. Korshunova, I., Shi, W., Dambre, J. & Theis, L. (2017). Fast Face-Swap Using Convolutional Neural Networks, *Proc. of 2017 IEEE International Conference on Computer Vision (ICCV)* (Venice, Italy), 3697–3705, https://doi.org/10.1109/ICCV.2017.397

9. Tolosana, R., Vera-Rodriguez, R., Fierrez, J., Morales, A. & Ortega-Garcia, J. (2022*).* An Introduction to Digital Face Manipulation*.* In: Rathgeb, C., Tolosana, R., Vera-Rodriguez, R. & Busch, C. (Eds.), *Handbook of Digital Face Manipulation and Detection* (pp. 3–26). Berlin, Germany: Springer International Publishing, https://doi.org/10.1007/978-3-030-87664-7_1

10. Jiang, L., Li, R., Wu, W., Qian, C. & Loy, C. C. (2020). Deeperforensics-1.0: A large-scale dataset for real-world face forgery detection, *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (Seattle, WA), 2889–2898, https://doi.org/10.1109/CVPR42600.2020.00296

11. Li, L., Bao, J., Yang, H., Chen, D. & Wen, F. (2020). Advancing high fidelity identity swapping for forgery detection, *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (Seattle, WA), 5074–5083, https://doi.org/10.1109/CVPR42600.2020.00512

12. Liu, K., Perov, I., Gao, D., Chervoniy, N., Zhou, W., et al. (2023). Deepfacelab: Integrated, flexible and extensible face-swapping framework, *Pattern Recognition*, **141**, 109628, https://doi.org/https://doi.org/10.1016/j.patcog.2023.109628

13. Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., et al. (2014). Generative adversarial nets, *Proc. of Advances in Neural Information Processing Systems* (Montreal, Quebec, Canada), https://doi.org/10.48550/arXiv.1406.2661

14. Ke, Z., Sun, J., Li, K., Yan, Q. & Lau, R. W. H. (2022). MODNet: Real-Time Trimap-Free Portrait Matting via Objective Decomposition, *Proceedings of the AAAI Conference on Artificial Intelligence*, **36**(1), 1140–1147, https://doi.org/10.1609/aaai.v36i1.19999

15. Chang, H., Lu, J., Yu, F. & Finkelstein, A. (2018). Pairedcyclegan: Asymmetric style transfer for applying and removing makeup, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Salt Lake City, UT), 40–48, https://doi.ieeecomputersociety.org/10.1109/CVPR.2018.00012

16. Rozsa, A., Günther, M., Rudd, E. M. & Boult, T. E. (2019). Facial attributes: Accuracy and adversarial robustness, *Pattern Recognition Letters*, **124**, 100–108, https://doi.org/10.1016/j.patrec.2017.10.024

17. Hassani, A., Malik, H. & Diedrich, J. (2022). Efficiently Mitigating Face-Swap-Attacks: Compressed-PRNU Verification with Sub-Zones. Technologies, *Information*, **10**(2), 46, https://doi.org/10.3390/technologies10020046

18. Zhang, Y., Zheng, L., & Thing, V. L. L. (2017). Automated face swapping and its detection, *Proc. of 2017 IEEE 2nd International Conference on Signal and Image Processing (ICSIP)* (Singapore), 15–19, https://doi.org/10.1109/SIPROCESS.2017.8124497

19. Chang, C.-C. & Lin, C.-J. (2011). LIBSVM: A Library for Support Vector Machines, *ACM Transactions on Intelligent Systems and Technology*, **2**(3), 1–27, https://doi.org/10.1145/1961189.1961199

20. Kodovsky, J., Fridrich, J. & Holub, V. (2011). Ensemble classifiers for steganalysis of digital media, *IEEE Transactions on Information Forensics and Security*, **7**(2), 432–444, https://doi.org/10.1109/TIFS.2011.2175919

21. Zheng, L., Duffner, S., Idrissi, K., Garcia, C. & Baskurt, A. (2016). Siamese multi-layer perceptrons for dimensionality reduction and face identification, *Multimedia Tools and Applications*, **75**(9), 5055–5073, https://doi.org/10.1007/s11042-015-2847-3

22. Zhang, W. & Zhao, C. (2020). Exposing Face-Swap Images Based on Deep Learning and ELA Detection, *MDPI Proceedings Volumes,* **46**(1), 29, 1–8, https://doi.org/10.3390/ecea-5-06684

23. Volkova, S. S. & Bogdanov, A. S. (2021). A deep learning approach to face swap detection, *International Journal of Open Information Technologies*, **9**(10), 16–20.

24. Zhang, S., Zhu, X., Lei, Z., Shi, H., Wang, X., et al. (2017). S3fd: Single shot scale-invariant face detector, *Proc. of the IEEE International Conference on Computer Vision* (Venice, Italy), 192–201.

25. Tang, X., Du, D. K., He, Z., & Liu, J. (2018). PyramidBox: A Context-Assisted Single Shot Face Detector. In: Ferrari, V., Hebert, M., Sminchisescu, C., & Weiss, Y. (Eds.), *Computer Vision – ECCV 2018* (pp. 812–828). Berlin, Germany: Springer Nature, https://doi.org/10.1007/978-3-030-01240-3_49

26. Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., et al. (2009). ImageNet: A Large-Scale Hierarchical Image Database, *Proc. of the 2009 IEEE Conference on Computer Vision and Pattern Recognition*, (Miami, FL), 248–255, https://doi.org/10.1109/CVPR.2009.5206848

27. Sandler, M., Howard, A., Zhu, M., Zhmoginov, A. & Chen, L. (2018). MobileNetV2: Inverted Residuals and Linear Bottlenecks, *Proc. of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition* (Salt Lake City, UT), https://doi.org/10.48550/arXiv.1801.04381

28. A Rössler, A., Cozzolino, D., Verdoliva, L., Riess, C., Thies, J., et al. (2019). FaceForensics++: Learning to Detect Manipulated Facial Images, *Proc. of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV)* (Seoul, Korea), 1–11, https://doi.org/10.1109/ICCV.2019.00009

29. Afchar, D., Nozick, V., Yamagishi, J. & Echizen, I. (2018). MesoNet: a Compact Facial Video Forgery Detection Network, *Proc. of 2018 IEEE International Workshop on Information Forensics and Security (WIFS)* (Hong Kong, China), 1–7, https://doi.org/10.1109/WIFS.2018.8630761

30. Thabit, R. (2021). Review of medical image authentication techniques and their recent trends, *Multimedia Tools and Applications*, **80**(9), 13439–13473, https://doi.org/10.1007/s11042-020-10421-7

31. Khor, H. L., Liew, S.-C. & Zain, J. M. (2017). Region of Interest-Based Tamper Detection and Lossless Recovery Watermarking Scheme (ROI-DR) on Ultrasound Medical Images, *Journal of Digital Imaging*, **30**(3), 328–349, https://doi.org/10.1007/s10278-016-9930-9

32. Eswaraiah, R. & Reddy, E. S. (2014). ROI-based fragile medical image watermarking technique for tamper detection and recovery using variance, *Proc. of the 2014 Seventh International Conference on Contemporary Computing (IC3)* (Noida, India), 553–558, https://doi.org/10.1109/IC3.2014.6897233

33. Thabit, R., & Khoo, B.E. (2014). A New Robust Reversible Watermarking Method in the Transform Domain. In: Mat Sakim, H., Mustaffa, M. (Eds.), *The 8th International Conference on Robotic, Vision, Signal Processing & Power Applications* (pp. 161–168). Singapore: Springer, https://doi.org/10.1007/978-981-4585-42-2_19

34. Thabit, R. & Khoo, B. E. (2017). Medical image authentication using SLT and IWT schemes, *Multimedia Tools and Applications*, **76**(1), 309–332, https://doi.org/10.1007/s11042-015-3055-x

35. Thabit, R., & Khoo, B. E. (2014). Robust reversible watermarking scheme using Slantlet transform matrix, *Journal of Systems and Software*, **88**, 74–86, https://doi.org/https://doi.org/10.1016/j.jss.2013.09.033

36. Thabit, R. & Khoo, B. E. (2015). A new robust lossless data hiding scheme and its application to color medical images, *Digital Signal Processing*, **38**, 77–94, https://doi.org/10.1016/j.dsp.2014.12.005

37. Thabit, R. (2019). Multi-Biometric Watermarking Scheme Based on Interactive Segmentation Process, *Periodica Polytechnica Electrical Engineering and Computer Science*, **63**(4), 263–273, https://doi.org/10.3311/PPee.14219

38. Liu, X., Lou, J., Fang, H., Chen, Y., Ouyang, P., et al. (2019). A novel robust reversible watermarking scheme for protecting authenticity and integrity of medical images, *IEEE Access*, **7**(1), 76580–76598, https://doi.org/10.1109/ACCESS.2019.2921894

39. Ustubioglu, A. & Ulutas, G. (2017). A New Medical Image Watermarking Technique with Finer Tamper Localization, *Journal of Digital Imaging*, **30**(6), 665–680, https://doi.org/10.1007/s10278-017-9960-y

40. Viola, P. & Jones, M. (2001). Rapid object detection using a boosted cascade of simple features, *Proc. of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2001),* (Kauai, HI), 1–9, https://doi.org/10.1109/CVPR.2001.990517

41. Zhang, K., Zhang, Z., Li, Z. & Qiao, Y. (2016). Joint face detection and alignment using multitask cascaded convolutional networks, *IEEE Signal Processing Letters*, **23**(10), 1499–1503, https://doi.org/10.1109/LSP.2016.2603342

42. Rayachoti, E., Tirumalasetty, S. & Prathipati, S. (2020). SLT based watermarking system for secure telemedicine, *Cluster Computing*, **23**, https://doi.org/10.1007/s10586-020-03078-2

43. Mohammed, R. T., & Khoo, B. E. (2012). Image watermarking using slantlet transform, *Proc. of 2012 IEEE Symposium on Industrial Electronics and Applications* (Bandung, Indonesia), 281–286, https://doi.org/10.1109/ISIEA.2012.6496644

44. Thabit, R. & Khoo, B. E. (2022). Robust Reversible Watermarking Application for Fingerprint Image Security, *Advances in Systems Science and Applications*, **22**(1), 117–129, https://doi.org/10.25728/assa.2022.22.1.1176

45. Creators, T. (2024). *Pexels*. [Online]. Available https://www.pexels.com/photo/portrait-photo-of-man-3185944/

46. BUZZ, P. (2024). *Here's how to age multiple faces with FaceApp's old filter*. [Online]. Available https://www.popbuzz.com/internet/viral/age-multiple-faces-faceapp-old-filter/

47. Website, F. (2024). *freepik*. [Online]. Available https://www.freepik.com/free-photo/family-having-great-weekend_857329.htm

48. Images, i. G. (2024). *Family Pictures, Images and Stock Photos*. [Online]. Available https://www.istockphoto.com/search/2/image?phrase=family

