CheckScan: A Reference Hashing for Identity Document Quality Detection

Musab Al-Ghadi, Petra Gomez-Krämer, Jean-Christophe Burie

CheckScan: A Reference Hashing for Identity Document Quality Detection

Musab Al-Ghadi[1], Petra Gomez-Krämer[1], Jean-Christophe Burie[1],

[1]L3i, La Rochelle University, La Rochelle, France

ABSTRACT

One of important challenges in the document liveness detection process for identity document verification is quality verification. To tackle this challenge, this paper proposes a reference hashing approach to discriminate between the original template of the identity document image and the scan one, which is called checkScan. Actually, the discrimination process takes place between two aligned identity document images. The proposed approach is made up of two steps: feature selection based on Fast Fourier Transform (FFT) and hash construction. Feature selection based on FFT involves partitioning the identity document image into set of non-overlapping blocks, then the FFT magnitudes for each partitioned block is calculated in order to select a specific number of FFT magnitudes peaks as discriminative features. The hash construction step quantizes the selected peaks into binary codes by applying a new quantization approach that is based on the coordinates of the selected peaks. These two steps are combined together in this work to achieve good discriminate (well anti-collision) capability for distinct identity document image. Experiments were conducted in order to analyze and identify the most proper parameter to achieve higher discrimination performance. The experimental results were performed on the Mobile Identity Document Video dataset (MIDV-2020), and the results show that the proposed approach builds binary codes quite discriminative for distinct identity document images.

Keywords: identity document, liveness detection, quality verification, anti-collision, hashing

1. INTRODUCTION

Nowadays, digitization and distributed enrollment (via mobile) envelops all sectors of the world and is gaining popularity, especially with the Covid-19 pandemic, which has accelerated their use. Personal identity documents remain the best reference to ensure the authenticity of the user and to whom special treatment is owed. The falsification of these documents is on the rise and is still a problem as long as there is no automatic and rapid deep identity document verification system to detect all possibilities of forgery [1]. To tackle this in an optimized way and to ensure digital trust, an intelligent and precise verification solution is required. Indeed, the identity document verification can be designed at several levels including facial verification [2,3], structural conformity [4], perceptual/visual verification [5–7], content coherence [8] and also, quality verification. The last one is considered as one of important challenges in the document liveness detection process for identity document verification, which involves to check automatically for the identity documents by comparing scanned version with the existing original template. Solving this challenge helps in two ways: (i) detecting the quality of the identity document and the source used to capture the identity document. (ii) verifying the generality of any proposed identity document verification system. However, designing such a system requires to find a solution for some other relevant tasks related to identity document localization and information extraction in different real-life scenarios [9]. Hence, several parameters can take part in these tasks such as varying backgrounds, angles, light effects, perspective and camera qualities [4]. The solution is therefore based on a so-called detector approach which manages to read the entire scanned identity documents in order to be able to crop the image of identity document, then to recognize the features of the component characters (photo, text, holograms, guilloches, etc.) in order to be able to link between them in order to validate the authenticity or to reject it [10].

This paper proposes a reference hashing approach and set up baselines for identity document image scan detection. The problem tackles in this paper is one challenge in the document liveness detection process for identity document verification. The proposed approach represents a solution for identity document verification from the side of quality verification. Hence, the proposed approach is made up of two steps: feature extraction and hash construction. Feature extraction step involves selecting n magnitude peaks of the FFT components as discriminative features. To this end, this step starts by partitioning the identity document image into set of non-overlapping blocks, then the FFT magnitudes for each partitioned block is calculated in order to select a specific number of FFT magnitudes peaks. The hash construction step quantizes the selected

peaks into binary codes by applying a new quantization approach using the coordinates of the selected peaks. These two steps are combined together in this approach to achieve well discriminate capability for distinct identity document image. The remainder of this paper is organized as follows: the related work is presented in section 2. Section 3 presents the proposed approach. Section 4 is dedicated to the experimental results. This paper ends with a conclusion in section 5.

2. RELATED WORK

This section explores the most relevant state-of-the-art papers in the domain of identity document verification.

An identity document authentication approach based on face verification and information recognizing is proposed in [2]. This approach is achieved in two steps: the first step involves identity document extraction based on Morphology Transformed Feature Mapping (MTFM) model, while the second step involves extracting the significant features based on Inception-ResNet model. The achieved verification accuracy of the identity documents is ranged 96.0-97.5%. In [4] the authors proposed a machine learning based approach for identity document acquisition and verification. This approach processed the identity document images in two steps: the first step involves pre-processing module for identity document localization. While, the second step involves designing a classifier module for verifying the identity document type and its legitimacy using visual pattern features. The classifier model based on extracting global and local features from the identity document image, and then these features are inserted into the Support Vector Machine (SVM) and Random Forest (RF) classifiers to test the document classification. The accuracy rate of this approach reached 97.5% and the F1-score exceeded 0.97. An approach for periodic pattern detection on passport document images is proposed in [5]. The logo that is often presented in the passport page is considered as security periodic pattern and used in control for fraud detection. This approach is designed in two steps: the first step involves applying the FFT of the processed periodic element and then taking into account its FFT magnitudes to select k peaks candidate locations. By calculating the average information in the 8 neighbours of each peak; a threshold is defined and used for fraud detection. The accuracy rate of this approach ranged 98-99%. On other level of identity document verification, a perceptual based verification approach for identity document images is proposed in [6]. This approach based on extracting set of visual features from the identity document image and using them to discriminate whether it is a genuine or a fraudulent image. The obtained average accuracy of this approach reached 90%.

The Convolutional Neural Networks (CNNs) were also used to offer set of solutions in the challenge of identity document verification. A deep learning approach for identity document verification called CheckSim is proposed in [7]. This approach adapted two CNN based models called Siamese and triplet in order to extract significant features and then to calculate the distance between the feature vector of the reference identity document and the identity document in control. The similarity between two identity documents is expressed through the distance between the two extracted feature vectors. The verification accuracy for Siamese network model reached 98.7% and it reached 99.2% for the triplet network model. In [8] the authors proposed a deep learning based approach for detecting counterfeit banknote documents based on a recurrent comparison for two textured background blocks; one from the genuine banknote document and the other from the counterfeit banknote document. The proposed approach learning the difference between the two processed blocks iteratively with attention model into specific zones in the background of the banknote document. The achieved accuracy for this approach ranged 90-98%. Another identity document verification approach based on CNN is proposed in [9]. This approach works in two tasks: the first task involves localizing the security objects like seal, signature and stamp in the processed identify document using the oriented fast and Rotated Brief (ORB) method. While, the second task based on the Optical Character Recognition (OCR) and Linear Binary Pattern (LBP) to extract the significant features from the processed identity document. The authenticity of the processed document is evaluated by matching the LBP as sliding window operations. The forgery detection accuracy of this approach ranged 76-97%. The mentioned state-of-the-art prompt the motivation to propose a new approach for identity document image verification on the level of quality verification.

3. SYSTEM MODEL

The proposed hashing approach for efficient discrimination between template identity document image and scanned one is made up of two steps: feature extraction based on FFT magnitudes and hash construction. These two steps are integrated together in this approach in order to define a unified threshold (λ) and unified significant blocks (zones) for each of the processed identity document category. These two outputs lead to well discriminate the original template identity document image and the other we want to verify. The following subsections detail the steps of the proposed approach.

3.1 Feature Selection

FFT is an excellent region-based feature extractor and is used efficiently in down-scaling tasks to extract features relevant enough to represent the images and distinguish them correctly [11]. The proposed approach starts by partitioning the identity document image into a set of non-overlapping blocks and then, for each block, for a given identity document, the FFT is applied to extract discriminative features. Basically, the FFT transforms the processed block into low frequencies and high frequencies. The low frequency components represent the most of the information in the processed block and specifically the slowly varying components in the block. While, the high frequency components, represent the details information of the proposed block, specifically the fast varying components in the block like the edges and the noises in the background of the block. Indeed, the proposed approach is more interested with the edges details of the processed block. That is because these details are the most elements in the identity document, which are negatively affected by the scan operation from the quality level point of view. Hence, considering these elements as significant features to compare or discriminate between the original template of identity document image and the scan one could has a sound. To this end, the magnitudes of the high frequency FFT components for each processed block are calculated and are then sorted in descending order to select *n* of highest values as discriminative features. These features called the peaks. Algorithm 1 presents the pseudo-code for selecting FFT peaks for a given identity document image.

Algorithm 1 *FFT* peaks selection algorithm

INPUT: Aligned identity document image in size L×L, *n*: number of selected peaks
procedure FFT PEAKS SELECTION(identity document image, *n*)
    partition the identity document into m×n blocks
    for each $block_i$ in blocks **do**
        apply *FFT* technique.
        shift the zero-frequency of *FFT* components to the center of the spectrum.
        calculate the magnitude of the shifted *FFT* components.
        set 10×10 block around center of spectrum to zero
        subtract the upper/lower diagonal parts of *FFT* magnitudes to detect the diagonal portion that holds the peaks.
        select *n* peaks in the diagonal parts of the *FFT* magnitude to be as discriminative features.
    end for
    return *n FFT* peaks
end procedure

Algorithm 1 works with well aligned identity document in order to select *n* of FFT peaks. It starts by partitioning the given identity document into set of m×n blocks and then for each block_i, the FFT technique is applied and the zero-frequencies of FFT components are shifted to the center of the spectrum in order to distinguish between the low and high frequencies. 10×10 regions around the center of the FFT spectrum is set to zero in order to avoid select *n* components from the low frequencies. After that, the upper/lower diagonal parts of FFT magnitudes are subtracting in order to detect the diagonal portion of the high frequencies which holds the FFT peaks. And finally, *n* of FFT peaks from the diagonal parts of the FFT magnitude are selected as discriminative features for the processed block. Visual representation for all steps in algorithm 1 is presented in figure 1. The selected *n* peaks of each processed block are passed into hash construction step; specifically algorithm 2, in order to map them into binary codes. Hence, algorithm 2 aims to generate a unique signature for each block in the processed identity document image.

3.2 Hash Construction

This step works to map the selected *n* peaks into binary codes via a new quantization algorithm 2. To do so, the coordinates of the corners *A* and *B* should be defined, where corner *A* represents bit value 1 and corner *B* represents bit value 0. The coordinates of corner *A* is defined by crossing point between the minimum row and maximum column values within the coordinates of the selected peaks. While, the coordinates of corner *B* is defined by crossing point between the maximum row and minimum column values within the coordinates of the selected peaks. Then, each peak_i from the set of *n* peaks is quantized either to 0 or to 1 according to the minimum Euclidean distance to corner *A* or corner *B* as illustrated in figure 2. One example is introduced in figure 2. Here, six peaks are selected as interesting peaks where peaks ={P_1, P_2,P_3,P_4,P_5,P_6}, and the coordinates of corners *A* and *B* are defined and the Euclidean distance between each peak_i and

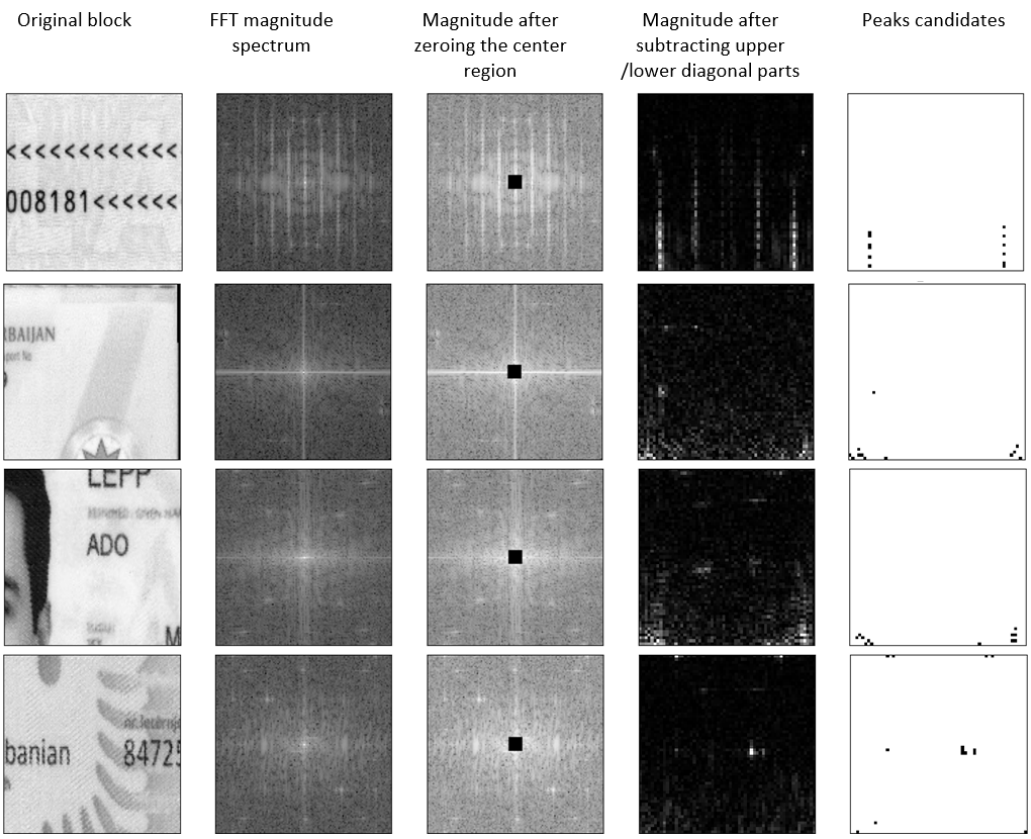

Figure 1. Visual representation for all steps in algorithm 1.

the corners A and B is calculated. Then, peak_i is quantized either to 0 or to 1 according to the minimum Euclidean distance to corner A or corner B. Hence, the signature for these peaks will be as signature = $\{0,0,0,1,1,1\}$. Algorithm 2 presents the pseudo-code for quantizing the n selected FFT peaks into a binary code.

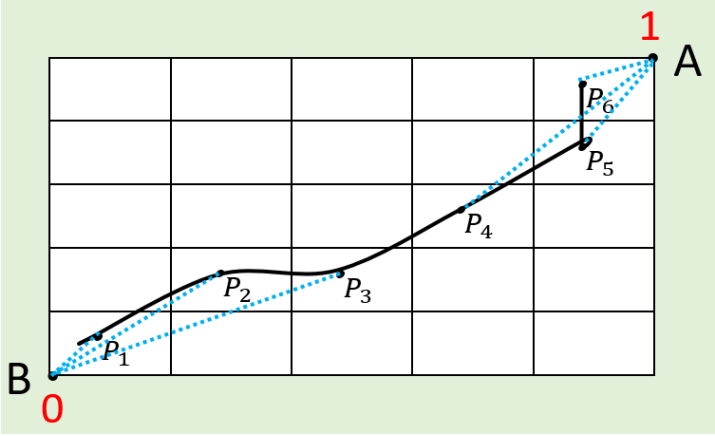

Figure 2. Quantization scheme.

4. EXPERIMENTAL RESULTS

This section details the information about the dataset and the performance metrics used. As well as, the experimental results of the proposed approach.

Algorithm 2 Quantization algorithm

INPUT: *n* FFT peaks
procedure BUILD A SIGNATURE(*n* FFT peaks)
    for each $peak_i$ in peaks: i={0,1,...,n}) **do**
        find the Euclidean distance between $peak_i$ and the corners *A* and *B*
        if Euclidean distance ($peak_i$,A) < Euclidean distance ($peak_i$,B) **then**
            $hash_i$ = 1
        else
            $hash_i$ = 0
        end if
    end for
    return *signature* = $\{hash_i, hash_1, ..., hash_n\}$
end procedure

4.1 Identity Document Dataset and Performance Metrics

To evaluate the performance of the proposed hashing approach, the MIDV-2020 dataset [12] is used. The MIDV-2020 dataset consists of 1000 video clips, 2000 scanned images, and 1000 photos of 1000 unique dummy identity documents with their annotations to read the ground-truths; these samples comprise identity document and passport for different 10 countries. Specifically, the 1000 template identity document images and the 1000 upper right scanned identity documents are selected to check the performance of the proposed approach. It is worth noting that the 1000 template identity document images were created from Wikimedia Commons as template samples, while the 1000 upper right scanned identity documents were created by scanning the template samples using Canon LiDE 220 and Canon LiDE 300 scanners with a resolution 2480×3507. 2/3 of the total identity document samples (1000 template + 1000 scanned) are selected randomly to use as a training set and the rest portion, which is 1/3 of the total size of the identity document samples is used as a testing set. The implementation of the proposed approach has been carried out in Python and running on HP laptop, Intel(R), Core(i7).

The accuracy rate and the error rate metrics are used to evaluate the discrimination performance of the proposed approach. Where the accuracy rate and the error rate are calculated according to the equation 1, as reported in [13].

$$accuracy\_rate(\lambda) = \frac{x_1(hd < \lambda)}{X_1} \quad , \quad error\_rate(\lambda) = \frac{x_2(hd < \lambda)}{X_2} \tag{1}$$

where x_1 is the number of similar pairs of identity document image blocks classified as similar blocks, x_2 is the number of distinct pairs of identity document image blocks classified into similar blocks, X_1 and X_2 correspond to the total number of similar and distinct pairs of identity document image blocks, respectively. λ is the discriminative threshold to consider pairs of identity document image blocks as similar or distinct blocks. *hd* is the Hamming distance between the hash codes of the two processed blocks. For more clarifying, similar pairs mean that the first block and the second block are belong either to the original template images group or to the scanned images group, while distinct pairs mean that the first block belongs to the original template images group and the second block belongs to the scan images group, or vice versa. The receiver operating characteristics (ROC) curve is used to evaluate the discrimination performance with different thresholds. The True Positive Rate (*TPR*) and False Positive Rate (*FPR*) of the ROC curve indicate the discriminative capability of the proposed approach. The *TPR* and *FPR* ratios representing the accuracy rate and error rate in equation 1, respectively.

4.2 Parameter Determination

The discrimination threshold (λ) has a direct influence on the performance of the proposed approach and it needs to be determined. To do so, the *hd* between the hash codes of all blocks of the original template of identity documents are globally calculated. This is achieved by stacking all partitioned blocks into *k* categories, and for each block category a global Hamming distances between the hash codes of all blocks are calculated. As example, figure 3 presents the *hd* distributions for 16 blocks of the original template identity document images in case of selecting 8 FFT peaks. Hence, *hd* = {0,1,...,8}. For example, suppose that we have 100 original identity document images for a specific country, and each of these images is resized as 512×512 and after is partitioned into 16 blocks each in size 128×128. Then, totally will

have 16 *hd* distributions, and the maximum frequency for a specific *hd* is equal 10000. From figure 3 we can inference two significant results: (i) defining the reference (most significant) blocks to use in the discrimination task. (ii) λ as a discriminative threshold. Statistically, the most interesting blocks for discriminative use in figure 3 are the blocks 6–10, that's because the *hd* distribution is dense in *hd* = 0, which indicates the most similar block. Actually, the highest *hd* distribution as 0 makes it easy to discriminate the original template identity document images against the other scanned blocks or could be forged blocks. Then, we can initially define $\lambda = 0$ as unique discriminative threshold and the blocks $=\{block_6, block_7, block_8, block_9, block_{10}\}$ are reference blocks for a given country. Visually, the information hold in each of these reference blocks are much varying in its background.

To verify the previous mentioned inferences, the Hamming distances between the hash codes of all blocks of the original template of the identity document images and the hash codes of all blocks of the scanned identity document images are also calculated. Figure 4 presents the *hd* distributions between 16 blocks of the original template identity document images and the scanned identity document in case of selecting 8 FFT peaks. The *hd* distribution in figure 4 confirms the assumed inferences from 3, where the *hd* = 0 in the interesting blocks 6–10 present the lowest frequency and the frequencies values are almost equal 0. This demonstration leads to define $\lambda = 0$ as a discriminative threshold between the original template identity document images and scanned identity document images. According to all of the above, each country will have a reference (global) threshold and set of reference blocks for discrimination use. The discriminative thresholds and the set of reference blocks for all processed countries are stored in a database for using after in the discrimination process.

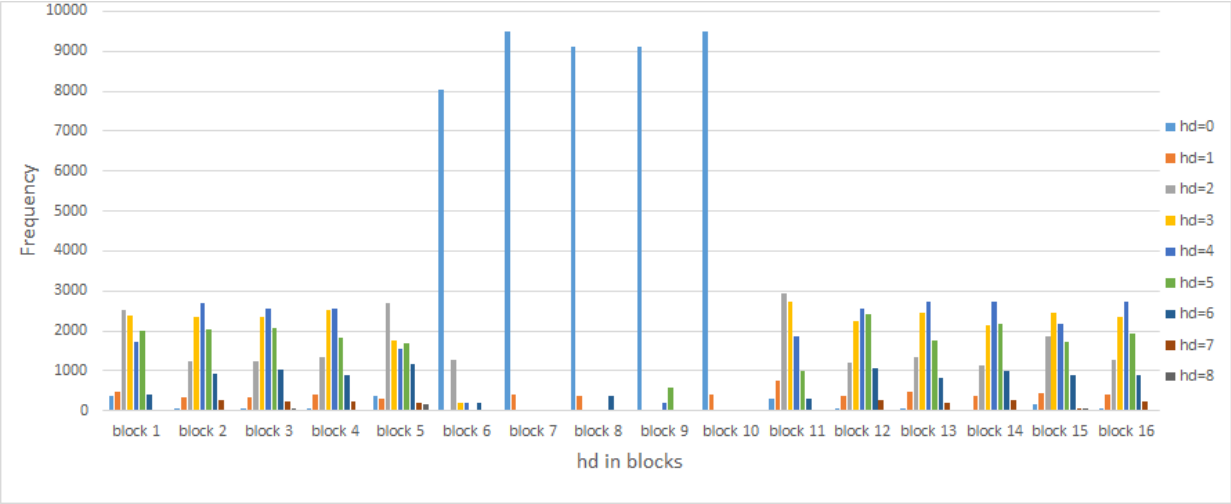

Figure 3. *hd* between the set of blocks of the original template identity document images.

4.3 Performance Test

Table 1 shows the accuracy and the error rates performances of the proposed approach for the 10 mentioned countries in MIDV-2020 with different lengths of hash codes (different lengths of selected FFT peaks).

The results in table 1 are very interesting where the accuracy rate exceeds 95% with hash length 8 bits and exceeds 99% with hash code length 16 bits, and reaches 100% with hash code length 32 bits. With hash code length 4 bits, the error rate is a little bit high for most identity document countries. This high error rate with hash code length equal 4 bits can be explained due to the tight space (small bits length) for generating more discriminative hash codes for the distinct blocks. So, increasing the length of the hash code allows more space to generate more discriminative hash codes and hence enhancing the accuracy rate.

Moreover, figure 5 shows the discrimination performance in *TPR-FPR* curves @4bits, @8bits, @16bits and @32bits with varying discriminative thresholds λ)s. The obtained *TPR* and *FPR* ratios in figure 5 demonstrate the discriminative capability of the generated hash codes using the proposed approach, and we can see that the *TPR* of the ROC curve that is generated using 32 bits is outperform the *TPR* ratios of the ROC curves that is generated using 4 and 8 and 16 bits by large margins.

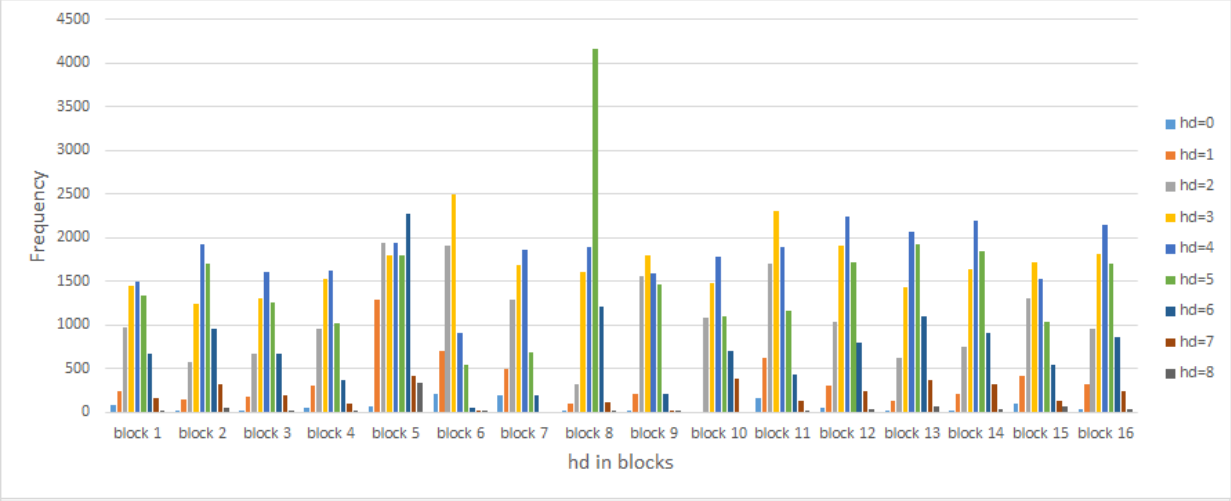

Figure 4. *hd* between the set of blocks of the original template and the scanned identity document images.

Table 1. Accuracy and error rates of the proposed approach, with 4, 8, 16, and 32-bits hash codes.

Country\ID	4 bits		8 bits		16 bits		32 bits	
	accuracy_rate	error_rate	accuracy_rate	error_rate	accuracy_rate	error_rate	accuracy_rate	error_rate
alb_id	0.93	0.7	0.99	0.01	0.99	0.01	1.0	0.0
aze_passport	0.71	0.29	1.0	0.0	1.0	0.0	1.0	0.0
esp_id	0.96	0.04	0.99	0.01	1.0	0.0	1.0	0.0
est_id	0.93	0.07	0.99	0.01	1.0	0.0	1.0	0.0
fin_id	0.89	0.11	0.99	0.01	1.0	0.0	1.0	0.0
grc_passport	0.86	0.14	1.0	0.0	1.0	0.0	1.0	0.0
lva_passport	0.89	0.11	0.96	0.04	1.0	0.0	1.0	0.0
rus_passport	0.93	0.07	0.99	0.01	1.0	0.0	1.0	0.0
srb_passport	0.91	0.09	0.98	0.02	1.0	0.0	1.0	0.0
svk_id	0.97	0.03	1.0	0.0	1.0	0.0	1.0	0.0

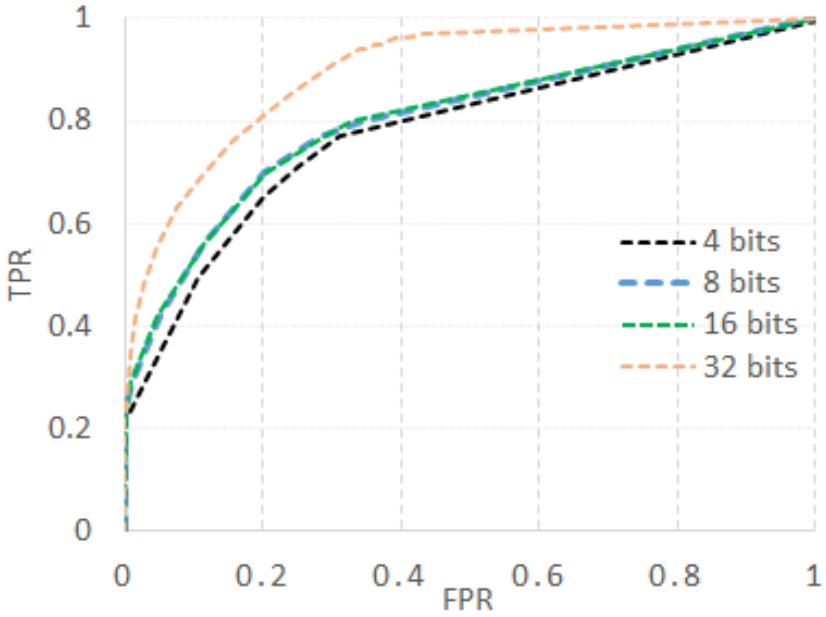

Figure 5. ROC curve of the proposed approach with 4, 8, 16 and 32-bits hash codes.

5. CONCLUSION

A reference hashing approach for quality verification of identity document images is proposed in this paper. The quality verification was studied in means of original template images and scanned versions. The magnitude of the FFT components are exploited to select and identify the most discriminative features to build a unique hash code for each block in the processed identity document image. To this end, an adaptive quantization scheme is proposed to generate the unique hash code for each processed block in the identity document image. The generated hash codes have good discriminative capability for distinct identity document images. The discrimination performance was evaluated on one public identity documents called MIDV-2020 dataset. The proposed approach provides an overall good performance. A future line of research is exploring the proposed approach to achieve quality verification in means of scanning-printing operation.

ACKNOWLEDGMENTS

This work is financed by the FUI IDECYS+ project.

REFERENCES

[1] A. Berenguel, O. R. Terrades, J. Llados, and C. Canero, "E-counterfeit: A mobile-server platform for document counterfeit detection," in 2019 International Conference on Document Analysis and Recognition (ICDAR), 15–20 (2019).

[2] X. Wu, J. Xu, J. Wang, Y. Li, W. Li, and Y. Guo, "Identity authentication on mobile devices using face verification and ID image recognition," Procedia Computer Scienc **162**, 932—939 (2019).

[3] A. Chinapas, Polpinit, N. Intiruk, and K. R. Saikaew, "Personal verification system using ID card and face photo," International Journal of Machine Learning and Computing, **9**, 407–412 (2019).

[4] A. Castelblanco, J. Solano, C. Lopez, E. Rivera, L. Tengana, and M. Ochoa, "Machine learning techniques for identity document verification in uncontrolled environments: A case study," in Mexican Conference on Pattern Recognition, 271–281, Springer (2020).

[5] T. Chernov, V. M. Kliatskine, and D. Nikolaev, "A method of periodic pattern detection on document images," in Digital Library of the European Council for Modelling and Simulation, (2015).

[6] N. Ghanmi and A. Awal, "A new descriptor for pattern matching: Application to identity document verification," in 13^{th} IAPR International Workshop on Document Analysis Systems (DAS), 375–380, IEEE (2018).

[7] N. Ghanmi, C. Nabli, and A. Awal, "Checksim: A reference-based identity document verification by image similarity measure," in Barney Smith E.H., Pal U. (eds) Document Analysis and Recognition – ICDAR 2021 Workshops, 422–436, Springer (2021).

[8] A. B. Centeno, O. R. Terrades, J. L. Canet, and C. C. Morales, "Recurrent comparator with attention models to detect counterfeit documents," in 2019 International Conference on Document Analysis and Recognition (ICDAR), 1332–1337 (2019).

[9] M. S. Yoosuf and R. Anitha, "Forgery document detection in information management system using cognitive techniques," Journal of Intelligent and Fuzzy Systems , 8057–8068 (2020).

[10] S. Usilin, D. Nikolaev, and D. Sholomov, "Guilloche elements recognition applied to passport page processing," in 8^{th} Open German - Russian Workshop Pattern Recognition and Image Understanding, 1–5 (2011).

[11] B. Yanikoglu and A. Kholmatov, "Online signature verification using fourier descriptors," EURASIP Journal on Advances in Signal Processing **2009**, 1–17 (2009).

[12] K. Bulatov, E. Emelianova, D. Tropin, and et al., "MIDV-2020: A comprehensive benchmark dataset for identity document analysis," Computer Vision and Pattern Recognition, arXiv:2107.00396 , 1–17 (2021).

[13] J. Ouyang, G. Coatrieux, and H. Shu, "Robust hashing for image authentication using quaternion discrete fourier transform and log-polar transform," Digital Signal Processing **41**, 98–109 (2015).

AUTHORS' BACKGROUND

Name	Title	Research Field	Personal website
Musab Al-Ghadi	Postdoc	image processing, data authentication, fraud detection	https://www.researchgate.net/profile/Musab_Al-Ghadi
Petra Gomez-Krämer	Associate Professor	image, video and document processing, fraud detection	https://pageperso.univ-lr.fr/petra.gomez/
Jean-Christophe Burie	Full Professor	image processing , pattern recognition, fraud detection and identity theft, object tracking.	https://l3i.univ-larochelle.fr/Burie-Jean-Christophe-MCF-HDR