A New Descriptor for Pattern Matching: Application to Identity Document Verification

Nabil Ghanmi and Ahmad-Montaser Awal
AriadNEXT - Pôle R&D document
Rennes, France
Email: {nabil.ghanmi,montaser.awal}@ariadnext.com

***Abstract*—Identity document verification consists on checking its conformity to one or eventually a set of authentic documents. This verification is usually performed through visible patterns matching. In this paper, we propose a new efficient visual descriptor for pattern comparison. As most of existing descriptors incorporate either color or spatial information; the proposed descriptor, called *Grid-3CD*, includes both information. This descriptor is based on color connected components (CC) extracted from a quantified image. It consists of a set of 6-tuples computed on a grid of pixels sampled from the color-quantified image. The 6-tuple of a given pixel describes the density, the mass center, the bounding box and the color of the CC that contains this pixel. The efficiency of this descriptor for identity document verification is shown using two strategies of pattern comparison. The first one is unsupervised and based on a distance measure whereas the second is supervised and based on one-class Support Vector Machine (SVM). The experimentation of the new descriptor on four datasets of identity documents totaling 3250 documents shows an average accuracy of about 90%, outperforming state-of-the-art descriptors.**

***Keywords*-Pattern matching; Identity document verification; fraudulent / authentic document; Grid-3CD; Similarity measures**

I. Introduction

As the generation of fraudulent documents became nowadays easy and frequent due to sophisticated printing tools and powerful image softwares, the automatic verification of document authenticity has become a necessity. The work presented in this paper is a part of an industrial research project IDFRAud[1] proposing a platform for identity document verification and analysis. Document verification can be held at several levels, namely, content consistency, structural conformity, and visual verification. We focus on the visual verification in the current work. It aims at checking whether a given document is visually similar to a reference (an authentic document), more specifically, verifying if security patterns are present and conform to the authentic ones. The main idea of such process relies on image processing and discrimination principles by which visual features are firstly extracted from the document image and then used in order to distinguish between authentic and fraudulent documents. In fact, security patterns might be altered or even missed in a fraudulent document.

The efficiency of this verification depends strongly on the features quality in terms of representativity and discrimination ability. In this work, we study the use of some existing features for identity document verification and we propose a new feature which performs well on the studied document corpus.

The rest of this paper is organized as follows. In section 2, we review several state-of-the-art approaches for document matching and verification and we focus on the most used descriptors. We discuss the use of some existing features for identity document verification and we propose a simple and efficient feature which is adapted to the treated documents in section 3. We discuss the experimentation results in section 4. We conclude and we present some perspectives of this work in section 5.

II. State of the Art

The identity document verification can be seen as a classification or a matching task by which a given document image is classified as authentic or fraudulent. Therefore, we are based on the literature of the CBIR[2] field which deals with the same issues, namely feature selection and extraction as well as similarity measurement.

It should be pointed that CNN-based features have shown impressive performance on a wide range of computer vision applications, notably image classification [15]. Nevertheless, these systems require a huge amount of training data (some tens of thousands of documents). As we deal with identity documents which are very difficult to collect, we do not explore this kind of systems. The other most used features in CBIR systems such as color, texture and shape will be investigated.

In [10], the authors present a comparison of three color features in RGB model for image retrieval using two similarity measurements. The used features are color moments (a 3-dimensionnal vector composed of the mean, the variance and the standard deviation of the image), color histogram which describes the distribution of the image pixels and the color coherence vector (CCV) which is a histogram that takes into account the spatial information by partitioning each bin into two types: coherent and incoherent. A pixel is considered coherent if it belongs to a big CC, otherwise it is considered as incoherent. In order to reduce the features space, the images are quantized from 2^{24} colors to 256. The similarity between two images is measured using two distance types: sum of

[1] This work was achieved in the context of the IDFRAud project ANR-14-CE28-0012, co-financed by the french DGA: http://idfraud.fr/

[2] Content Based Image Retrieval

978-1-5386-3346-5/18 $31.00 © 2018 IEEE
DOI 10.1109/DAS.2018.74

squared differences (SSD) and sum of absolute differences (SAD). The experiments are performed on a corpus of 14500 images considering the features individually and also combined. The authors report that the best performance is obtained using a combination of all the mentioned features and SAD distance. The above features are also used by Shahbahrami et al. in [13]. The authors compared these features with two textural descriptors, namely GLCM and discrete wavelet transform (DWT). A GLCM describes the frequencies of occurrence of grey level combinations among pairs of pixels [11]. Different directions such as horizontal, vertical, diagonal and anti-diagonal are considered. Four statistical features are then extracted: energy, entropy, contrast and homogeneity. The DWT-based features consist of the mean and the standard deviation of 10 sub-bands obtained by applying three times a 2-D DWT on the image. As a similarity measurement, the authors adopted the SSD distance between image feature vectors. The experiments are carried out on a set of 1000 images corresponding to 10 classes. The obtained results showed That CCV and GLCM gave better accuracy.

To make the matching more accurate, Chan et al. [1] used a color run-length (color RL) feature which integrates the information of color and geometrical shapes of the objects in an image. As the computation of this feature is time-consuming, the authors presented a faster representation that uses partial RLs. To compute this feature, the colors of an image are firstly quantized using a limited color palette. For each color value, the partial RL is computed in four directions: 45°, 90°, 135°, 180°. The image comparison is performed using the SAD disatnce between the feature vectors. This approach is tested on a set of 400 images and an accuracy of about 96% is obtained.

Shrivastava et al. [14] used the HOG template descriptor [3] for matching images that are visually similar based on discriminative learning. This technique uses the feature vector and computes a set of weights that express a "uniqueness" factor of each component. The authors employ a linear SVM framework to learn these weights which are used later to compute the visual similarity between two images. The proposed approach was applied on some matching-based applications such as scene completion and scene exploration.

As edges are important features that characterize the image content, the EHD is considered as one of the most robust descriptors for image comparison. It is used in [12] for an image retrieval purpose. Differently from MPEG-7 standard where only the local edge distribution is considered, the authors also used global and semi-global histograms. These two histograms are respectively obtained by totally and partially grouping the bins in different image blocks. The similarity between two images is computed using a weighted SAD distance between their correspondent edge histograms. The experiments are carried out on a set of 11639 images of the MPEG-7 core experiment database. An other efficient feature widely used for texture description is based on the LBP operator. This latter is non-parametric descriptor whose describes each pixel by the relative gray levels of its neighbouring pixels [7]. It is adopted in [6] for matching interest regions between a pair of images and for object category classification using respectively the SSD distance and the SVM classifier. The authors report that this feature performs better than SIFT [4] descriptor.

TABLE I
DOCUMENT VERIFICATION ACCURACY USING STATE-OF-THE-ART FEATURES EXTRACTED WITH DIFFERENT PARTITIONS OF THE INPUT IMAGE

	Accuracy			
Feature	Full Image	3×3	5×5	10×10
Color Histogram	61.9	62.5	**62.7**	61.7
Color Texture	73.0	73.8	**81.3**	81.2
Color Moments	56.7	64.8	64.5	**64.9**
GLCM	64.1	65.7	66.5	**66.9**
Gabor	64.7	64.1	64.9	**71.1**
EHD	63.9	69.4	75.7	**77.2**
LBP	65.1	64.7	69.1	**71.0**
HOG	77.6	-	-	-

Zhang et al. presented in [17] an image retrieval approach based on Gabor filter. This technique is adopted for extracting textural features by applying a bank of filters where each of which captures the energy at specific frequency and direction. The extracted features are the mean and the variance of the gabor filtered image. To compare two images, the authors used the SSD distance between their corresponding feature vectors. The tests are conducted on textured and natural images.

In conclusion, several descriptors and similarity measurements have been proposed in the CBIR literature. These methods have been applied on different document datasets and have never been compared on the same corpus. Thus, to select the most suitable descriptors for identity document verification, an experimental study is performed. In addition, identity document verification incorporates some specifities that need specific adaptations or optimisations of the existing descriptors.

III. IDENTITY DOCUMENT VERIFICATION

An identity document has an invariable background in addition to textual zones that contain field labels (name, surname, address, etc.) and the ID's holder personal information and eventually his photo. The document background is often characterized by some visual patterns or texture.

There are two possible scenarios for the document verification. The first one takes into account the entire image content (including the variable zones). This has the advantage of not localizing automatically or manually specific zones to be compared. Nevertheless, the presence of variable text areas may lead to some false matches. The second scenario is to consider only fixed zones that contain discriminant patterns or texture. In this scenario, verification zones must be defined for each document class. Its advantage is that it offers a more accurate comparison. In our experiments, we tested these two scenarios.

A. *Experimental Study of existing features*

We carried out an experimental study of eight among the most used features in the literature (see Table I). A dataset of 471 identity documents (French identity cards) is used. It contains 361 authentic documents and 110 fraudulent ones (generated as in section IV).

Apart from HOG, all the tested features do not incorporate spatial information when extracted from the full image. In order to overcame this deficiency, a semi-local extraction of these features is explored. To do so, the input image is divided into $N \times M$ cells and the features are then extracted separately from each cell. The global feature vector is obtained by appending the local feature vectors. Table I shows the obtained accuracy for each feature using different grid sizes 3×3, 5×5 and 10×10. As shown in this table, all features perform better when extracted from individual cells rather than the full image.

It should be noted that the reported results are experimentally optimized based on the feature parameters and settings:

- For the three color features, RGB space is considered since only minor differences with other color spaces is observed, as has already been proved in [5].
- From the GLCM, the fourteen Haralick Features are computed [16]. The GLCM allows to emphasise relevant texture information in a given direction using a predefined displacement. Several combinations of orientation and displacement are tested and the obtained results show that the vertical GLCM with a displacement of 1 performs better than other combinations. This can be visually explained by the presence of regular changes of gray levels along the vertical direction of the studied document images.
- As Gabor features, we extract the local energy and the mean amplitude on the Gabor-filtered images. These images are calculated using a bank of filters tuned to various orientations and frequencies. We are based on [8] to design the filter bank: four values of orientations $(0°, 45°, 90°, 135°)$ combined with the following values of radial frequencies $1\sqrt{2}, 2\sqrt{2}, 4\sqrt{2}, ..., (W/4)\sqrt{2}$ are used.
- As for LBP-based features, the gray level histogram is computed on the result obtained by the LBP operator on the input image. The optimal histogram size is experimentally set up to 32.
- EHD is a block-based feature. It depends on the total number of image blocks and the threshold used to decide if there is an edge or not. In our experiments, the first parameter is arbitrary set to 1100 as in [12] and the second parameter is optimized (to a value of 15) using a grid search method.

Using the optimal configuration of each descriptor, we tested all the possible combinations of those features (255 combinations) using a supervised classification strategy (see section III-B2). One classifier is though trained for each combination where the features are aggregated in a unique features vector. The results of the 3 best combinations are reported in table II. It is worth noting that the obtained accuracy ($\approx 84\%$) still unsatisfactory in the context of a real world application where errors are very costly. Indeed, an error lead either to unjust sanctions or cheating operations.

TABLE II
THE THREE BEST COMBINATIONS OF THE STUDIED DESCRIPTORS

Combination	Accuracy
Color Moments, GLMCM, Gabor, LBP, HOG	83.85
Color Moments, Gabor, LBP, HOG	83.73
GLMCM, Gabor, LBP, HOG	83.54

B. Grid-3CD: a new descriptor for document verification

We think that the major weakness of the above descriptors is that they are not suitable to directly encode image's spatial information. Thus, a new descriptor: *Grid Color Connected Components Descriptor* (abbreviated as ***Grid-3CD***) is proposed.

1) Semantic of Grid-3CD: this descriptor is extracted from a color image (in the RGB color space) and based on the CC position, color, and shape (see Fig. 1). The image is firstly scaled into a predefined size $W \times H$ while preserving the original aspect ratio. Then, each color channel is quantified into 2^c colors where $c \in]0, 8[$. Thus, the total number of colors is reduced to 2^{3c}. The main advantage of this quantification is to obtain coherent adjacent pixels. A connected component labelling is then performed in order to identify color regions (adjacent pixels sharing the same color). Finally, a vector of 6-tuples $(c_{ij}, x_{ij}, y_{ij}, d_{ij}, h_{ij}, w_{ij})$ is calculated on a grid of pixels p_{ij} sampled using fixed horizontal and vertical strides (s_h and s_v), where:

- $i = \{n \times s_h\}_{n \in \mathbb{N}}$ and $j = \{m \times s_v\}_{m \in \mathbb{N}}$ where $i < W$ and $j < H$
- c_{ij} is the quantified color of the CC containing p_{ij},
- x_{ij} and y_{ij} are the position of the CC mass center,
- d_{ij} is the CC density, and
- h_{ij} and w_{ij} are the CC bounding box dimensions.

The proposed descriptor can be seen as a coarse version of the raw pixel values based descriptor used in [9]. Nevertheless it is much faster since it is calculated only on a grid of sampled pixels. Furthermore, it is less sensitive to small pixel variations, as similar pixels are merged in the same color CC thanks to the color quantification. In addition, the Grid-3CD is invariant to local pixel translation since it is extracted at the CC level. It has also an important advantage specific to the identity document verification problem. Indeed, it extenuates the inter-document variability which is due to the variable text zones. This text is composed of character CCs that are more or less similar according to Grid-3CD descriptor.

Grid-3CD has two parameters which are the quantification level c and the sampling strides. These parameters are application-dependant and can be experimentally set up. Table III shows the document verification accuracy on a validation set of documents, using different combinations of these two parameters. The best accuracy is obtained with a quantification level $c = 16$ and sampling strides $(s_h \times s_v) = (4 \times 4)$.

2) Similarity measurement: The document verification is performed based on Grid-3CD descriptor using two strategies:

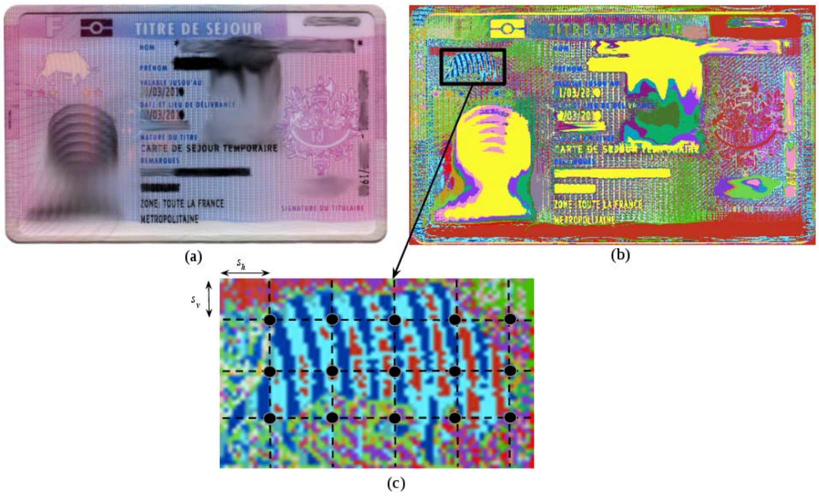

Fig. 1. (a) Identity document image scaled to a 512×320 (personal data are blurred for confidentialy reasons). (b) The extracted CCs on the corresponding quantified image. (c) A zoom on a snippet of the quantified image with the sampled pixels

TABLE III
DOCUMENT VERIFICATION ACCURACY USING GRID-3CD WITH DIFFERENT VALUES OF QUANTIFICATION LEVEL AND SAMPLING STRIDES

$s_h \times s_v$ c	3×3	4×4	5×5	7×7	10×10	12×12
2 (8 colors)	62.62	70.94	72.85	72.22	68.37	66.56
4 (64 colors)	77.85	78.80	78.16	77.79	77.75	75.58
8 (512 colors)	83.21	81.46	81.12	80.65	80.16	79.80
16 (4096 colors)	84.26	**85.09**	83.71	81.47	80.63	80.73

1) using feature vector distance: the verification is performed by computing the similarity between a given document image called query image (I_q) and a reference image (I_r) that represent an authentic document chosen by the user. Let $F_q = \{f_{q1}, ..., f_{qn}\}$ and $F_r = \{f_{r1}, ..., f_{rn}\}$ the feature vectors of I_q and I_r respectively. The similarity between I_q and I_r is obtained by computing the Canberra distance defined by the equation 1.

$$d(I_q, I_r) = \sum_{i=1}^{n} \frac{|f_{qi} - f_{ri}|}{|f_{qi} + f_{ri}|} \tag{1}$$

If the distance is below a given threshold, the document is considered as authentic, otherwise it is considered as fraudulent. The use of the Canberra distance is justified by the fact that it is normalized and thus provides a direct similarity measure even when using features taking their values into different intervals.

2) using a supervised method: the document verification is addressed as a classification task. A variant of SVM (SVDD [2]) allowing an efficient one-class classification is used. Thus, a given document is classified as belonging to the target class (authentic document) or as an outlier (fraudulent document). We opted for such classification and not for a two-class classification because the outliers (fraudulent documents) are rare and do not constitute a real class due to their large variability.

IV. EXPERIMENTATION

Datasets. Since there is no public dataset of identity documents, we used for the experimentation four private datasets (see Table IV). The fraudulent documents are artificially generated by adding clutters, scratches or by removing some parts of the document background. For each class of document, we define a set of fixed zones that contains fixed patterns and/or visual words.

Experimental protocol. As mentioned in section III, we evaluate the document verification considering firstly the full document image and secondly the fixed zones only. As previously said, the document verification is performed using:

- Canberra distance: in this case, each dataset is randomly divided into two subsets: validation set (composed of $1/4$ of the total documents) used to estimate the threshold decision and a test set (containing the $3/4$ remaining documents) used for the tests.
- one-class SVM classifier: in this case, we use a 4-fold cross validation evaluator.

As evaluation metric, the correct verification rate (R) defined by the equation 2 is used.

$$R = \frac{\#correctly\ verified}{\#total} \tag{2}$$

Results. The obtained results using Grid-3CD are summarized in Table V (in the grayed rows). It should be noted that in both comparison strategies, the results obtained using

TABLE IV
IDENTITY DOCUMENT IMAGE DATA SETS USED IN THE EXPERIMENTATION

Data set	Description	#images	
		authentic	fraudulent
IT_ID	Italian identity card	109	120
FR_ID	French identity card	1640	350
FR_RP	French residence permit	1385	300
SP_ID	Spanish identity card	122	121

TABLE V
DOCUMENT VERIFICATION ACCURACY ON FOUR DATA SETS: A COMPARISON OF GRID-3CD WITH BEST COMBINATION OF THE EXISTING DESCRIPTORS

Data set	Descriptor	Correct verification rate (%)			
		Full document		Fixed zones only	
		Canberra distance	SVM	Canberra distance	SVM
IT_ID	Grid-3CD	57.3	80.7	74.3	**86.8**
	Best Comb.	68.4	65.3	82.5	85.7
FR_ID	Grid-3CD	83.8	91.9	89.1	**92.2**
	Best Comb.	82.2	82.5	90.3	90.6
FR_RP	Grid-3CD	82.0	94.2	93.0	**94.5**
	Best Comb.	90.1	80.4	92.2	94.0
SP_ID	Grid-3CD	69.4	86.0	85.8	86.4
	Best Comb.	71.0	69.8	**90.2**	85.9

only the fixed zones are better than those obtained when the full image is used. This is due to the presence of variable zones containing text having different sizes. Indeed, as we already said, the Grid-3CD may absorb the variation between two different characters. Thus, two different words of the same size are equivalent. However, it is still difficult to cope with the problem of different text sizes without ignoring them. Whence comes the idea of considering only the fixed zones. In this latter case, the results are globally satisfactory. In particular, they are very satisfactory when enough documents are available for the SVM training or the threshold estimation; the case of the FR-ID and FR-RP data sets where the verification rate reaches 92.2% and 94.5% respectively.

It is also noted that, with the same features, the classification strategy performs better than the distance strategy. This is due to the large generalizability of the SVM. In addition, the distance-based strategy depend strongly on the reference image. This latter must be chosen so that it is similar to all authentic documents even those of bad quality and also different from all fraudulent documents even those having minor alteration.

Table V shows also the comparison of Grid-3CD with the best combination of the existing descriptors. It can be noted that Grid-3CD performs better on three datasets. The existing descriptors outperform Grid-3CD only on one dataset (SP_ID). However, this dataset do not contain enough documents and thus the obtained accuracy is less significant.

V. CONCLUSION

In this work, a detailed experimental study of the most common descriptors for document image verification is performed. The extraction level as well as the intrinsic parameters of each descriptor are discussed in order to find the best configuration of individual and combined descriptors. Once optimized, these features are used for identity document verification. However, the obtained results are unsatisfactory in the context of a real world application. For this reason, we proposed a new descriptor that incorporates simultaneously spatial, shape and color information. Globally, this descriptor performs better than the best combination of the existing features on four data sets of identity documents.

As future work, we would like to enlarge the datasets used for the experimentation and to study the combination of Grid-3CD with some of existing descriptors. Illumination invariance of Grid-3CD will also be investigated.

REFERENCES

[1] Y.-K. Chan and C.-C. Chang. “Image matching using run-length feature”. In: *Pattern Recognition Letters* 22.5 (2001), pp. 447–455.

[2] W.-C. Chang, C.-P. Lee, and C.-J. Lin. “A revisit to support vector data description”. In: *Dept. Comput. Sci., Nat. Taiwan Univ., Taipei, Taiwan, Tech. Rep* (2013).

[3] N. Dalal and B. Triggs. “Histograms of Oriented Gradients for Human Detection”. In: *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’05)*. CVPR ’05. 2005, pp. 886–893.

[4] L. David. “Object recognition from local scale-invariant features”. In: *ICCV* (1999), pp. 1150–1157.

[5] T. Deselaers, D. Keysers, and H. Ney. “Features for Image Retrieval: An Experimental Comparison”. In: *Information Retrieval* 11.2 (2008), pp. 77–107.

[6] M. Heikkilä, I. M. Pietikäinen, and C. Schmid. “Description of Interest Regions with Local Binary Patterns”. In: *Pattern Recognition* 42.3 (2009), pp. 425–436.

[7] D. Huang et al. “Local Binary Patterns and Its Application to Facial Image Analysis: A Survey”. In: *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews* 41 (2011), pp. 1–17.

[8] A.-K. Jain and F. Farrokhnia. “Unsupervised Texture Segmentation Using Gabor Filters”. In: *Pattern Recognition.* 24.12 (1991), pp. 1167–1186.

[9] D. Keysers et al. “Deformation models for image recognition”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 29.8 (2007).

[10] S.-R. Kodituwakku and S. Selvarajah. “Comparison of color features for image retrieval”. In: *Indian Journal of Computer Science and Engineering* 1.3 (2004), pp. 207–211.

[11] P. Mohanaiah, P. Sathyanarayana, and L. GuruKumar. “Image texture feature extraction using GLCM approach”. In: *International Journal of Scientific and Research Publications* 3.5 (2013), pp. 1–5.

[12] D.-K. Park, Y.-S. Jeon, and C.-S. Won. “Efficient use of local edge histogram descriptor.” In: *ACM Multimedia Workshops*. 2000, pp. 51–54.

[13] A. Shahbahrami, D. Borodin, and B. Juurlink. “Comparison between color and texture features for image retrieval”. In: *Proc.19th Annual Workshop on Circuits, Systems and Signal Processing*. 2008.

[14] A. Shrivastava et al. “Data-driven Visual Similarity for Cross-domain Image Matching”. In: *ACM Trans. Graph.* 30.6 (2011), 154:1–154:10.

[15] C. Tensmeyer and T. Martinez. “Analysis of Convolutional Neural Networks for Document Image Classification”. In: (2017), pp. 389–394.

[16] N. Zayed and H.-A. Elnemr. “Statistical analysis of Haralick texture features to discriminate lung abnormalities”. In: *Journal of Biomedical Imaging* (2015), pp. 268–279.

[17] D. Zhang et al. “Content-based image retrieval using Gabor texture features”. In: *Proceedings of the First IEEE Pacific Rim Conference on Multimedia*. 2000, pp. 392–395.