Classification et extraction des documents complexes à partir des images issues d'un périphérique mobile
Application aux documents d'identité

Ahmad Montaser Awal — Abdullah Almaksour

AriadNEXT - 80 Avenue des Buttes de Coësmes 35700 Rennes
{montaser.awal, abdullah.almaksour}@ariadnext.com

RÉSUMÉ. *Nous proposons dans cet article une méthode de classification des images de documents d'identité. Contrairement à plupart des systèmes existants, l'approche proposée permet de localiser le document et de reconnaître son type simultanément. Premièrement, une base de modèles est construite à partir d'une seule image de référence et éventuellement des exemples d'apprentissage. Ensuite, des points caractéristiques sont extraits de l'image requête et comparés aux modèles. Une fois le modèle de document identifié, une comparaison plus fine est effectuée afin de localiser précisément la zone de document dans l'image requête. Nous proposons plusieurs optimisations afin de garantir un temps de traitement indépendant de nombre des modèles. De plus, la qualité de document extrait est estimée ce qui permet de rejeter les documents inconnus. Notre système a été évalué sur une base de 2494 documents réels.*

ABSTRACT. *We propose in this paper a document image classification method. In contrary to most of existing systems, the proposed approach allow locating the document and recognizing its type simultaneously. First, a knowledge base of document models is created from reference images. Training images are not indispensable and though only one reference image is enough to create a document model. Then, key-points are extracted from the request image and matched against the models in the knowledge base. Once the document model is found, a more accurate matching is applied in order to extract the document from the image. The matching process is optimized in order to guarantee an execution time independent from the number of document models. Finally, the document quality is estimated in order to reject unkown documetns. Our system has been evaluated on a set of 2494 real documents*

MOTS-CLÉS : *Classification, Document d'identité, Descripteurs locaux, SURF, Flann, KDTree.*

KEYWORDS: *Classification, Identity documents, Local descriptors, SURF, Flann, KDTree.*

1. Introduction

Le travail présenté dans cet article a été réalisé au sein du projet ANR ***IDFRAud***. La fraude d'identité présente des risques très importants aux niveaux sociaux et sécuritaires. Selon l'Institut National des Hautes Études de la Sécurité et de la Justice (INHESJ), plus de 12 000 cas de fraude d'identité ont été détectés par la police nationale et près de 9200 par la Police Aux Frontières (PAF) en 2012. Ce type de fraude a coûté plus de 2 milliards d'euros à la sécurité sociale en France.

Le but de ce projet est de proposer un plate-forme permettant un déploiement des solutions mobiles de vérification et d'analyse des documents d'identités. Dans ce type de solutions, une première étape consiste à trouver le modèle du document présenté (pays, type, série, ...) pour ensuite effectuer les contrôles et les analyses adaptés. Cette étape nécessite une technique de classification automatique à partir de l'image de document. Cette technique doit être robuste à la qualité de la prise de vue du document notamment pour les solutions destinées à un usage sur un périphérique mobile (ex. smartphone). De plus, aucune information à priori n'est disponible en entrée. Néanmoins, les méthodes classiques de classification ne sont pas directement adaptées à ce problème.

En effet, un système de classification des documents d'identité doit adresser les challenges suivants :

– La grande variété des qualités d'images.

– Les images sont prises sans contraintes, et les documents doivent être extraits des fonds qui peuvent être très complexes.

– Le document doit être extrait avec une très grande précision afin de pouvoir lire des informations utiles du document.

– Passage à l'échelle : 9 classes sont nécessaires pour une couverture minimale des documents d'identité français (2 passeports, 1 carte nationale d'identité, 2 titres de séjour, 1 permis de conduire). Pour une couverture minimale au niveau européen le nombre des classes à reconnaître s'élève à une centaine (et à un millier pour une couverture mondiale).

– Rareté des images d'apprentissage : les documents d'identité ne sont pas disponibles en masse pour faire des apprentissages classiques. Très souvent, un seul exemplaire de chaque classe est disponible.

Nous allons dans la section suivante présenter l'état de l'art de classification des documents. Ensuite, nous adressons ces challenges en détaillant notre approche de classification et extraction des documents d'identité. L'évaluation de notre système sera effectuée sur un ensemble des documents réels dans la section 4. Nous conclurons ce papier avec une discussion des limitations et nos perspectives.

2. État de l'art

On peut distinguer trois grandes familles de méthodes de classification des documents : analyse de la disposition structurelle, analyse textuelle et analyse visuelle.

La classification par disposition structurelle de contenu d'un document est adaptée aux documents bien structurés très riches en textes tel que les articles de journaux ou les factures. D'une manière générale, un document est décrit par des blocs textuels, figures, tables, etc. (Hu *et al.*, 1999). L'image de document est donc caractérisée par la disposition spatiale des différents composants du document. Ainsi, une distance de type Manhattan (ou une autre méthode de classification classique) peut être utilisée pour regrouper les documents similaires dans la même catégorie. Une similarité visuelle entre deux documents peut être calculée à partir de ces blocs (Eglin et Bres, 2003). Les auteurs dans (Bagdanov et Worring, 2001) proposent de représenter la structure du document par un graphe relationnel d'attribut (ARGs). Cette représentation permet de regrouper les documents de la même classe par des graphes aléatoires de premier ordre (FORGs). Les informations spatiales relatives (position, taille, ratio) sont utilisées dans (Shin et Doermann, 2006) pour trouver les documents similaires avec l'image de document à analyser. Cette disposition spatiale peut être estimée par des arbres X-Y qui permet la construction des modèles probabilistes de la distribution des classes de documents (Cesarini *et al.*, 2001) (Diligenti *et al.*, 2003).

La disposition spatiale n'est plus discriminante quand les documents à classer ont des structures similaires ou si les documents de la même classe ne partagent pas la même structure spatiale. Dans ce cas, la transcription du texte présent dans le document est utilisée pour construire des BoW (Bag of Words). La plupart des méthodes combinent des classifieurs classiques avec les BoWs (Sebastiani, 2002) (Yang et Liu, 1999).

Les méthodes ci-dessus ne sont pas applicables aux documents complexes (comme par exemple les documents d'identité). La classification de ce type de document apporte des difficultés supplémentaires car le document n'est pas localisé à priori et les informations du fond peuvent perturber la classification. De plus, eles informations textuelles sont difficiles à extraire de avant connaître le modèle et le type de document. Dans ce contexte, on se base souvent sur des descripteurs visuels indépendants du contenu de document. L'extraction de ce type de descripteurs à partir des points d'intérêt a été prouvée très efficace pour la recherche et mise en correspondance des images de scène naturelle (Yang *et al.*, 2007). La méthode de BoW a été étendue pour représenter des descripteurs visuels au lieu de ceux textuels (Kumar et Doermann, 2013). Des descripteurs locaux de type SURF sont combinés à des histogrammes hiérarchiques et ensuite regroupés par un K-mean algorithme en un vocabulaire des mots visuels (de las Heras *et al.*, 2015). Ensuite, une image est représentée par un histogramme de ces mots ce qui permet d'entraîner un classifieur classique (SVM, MLP, ...) sur les différentes classes de documents (Chen *et al.*, 2012).

Certains travaux supposent que le document est localisé à priori ce qui réduit la difficulté des fonds complexes. Dans ce cas, des descripteurs globaux peuvent

être utilisés pour caractériser l'image de document (Simon *et al.*, 2015). (Augereau *et al.*, 2012) ont proposé un système de reconnaissance et de localisation des pièces d'identité dans des documents numérisés (scannés). Un modèle référence est construit pour chacune des deux classes (ID et passeport français) à partir d'une image de très bonne qualité. Ensuite, les points d'intérêt de l'image requête sont mis en correspondance avec les modèles pour détecter la classe la plus similaire.

Dans le cadre de ce projet, on n'impose aucune contrainte sur les images des documents à reconnaître. Elles peuvent varier des scans en haute qualité à des photos de mauvaises qualités prises par un téléphone portable. Nous avons dans un premier temps implémenté une méthode basée sur les BoW extraits à partir des descripteurs SURF (Bay *et al.*, 2006). Néanmoins, l'inconvénient principale de cette méthode est le besoin d'une base d'apprentissage assez grande pour entraîner le classifieur. Cette hypothèse n'est pas toujours valable et nous avons souvent très peu (voir un seul) d'exemple(s) pour la création des modèles.
Nous allons dans la section suivante détailler notre approche de classification de document inspirée de (Augereau *et al.*, 2012). Cependant, nous proposons plusieurs améliorations :

– La création du modèle de document à partir d'une seule image de référence.

– Un schéma de mise en correspondance indépendant du nombre de classes de document ce qui permet le passage à l'échelle. Nous évaluons notre approche en considérant 9 classes de documents.

– Une estimation de la qualité du quadrilatère de document localisé ce qui permet de rejeter les documents inconnus. La capacité de rejet est très importante dans un système complètement automatisé.

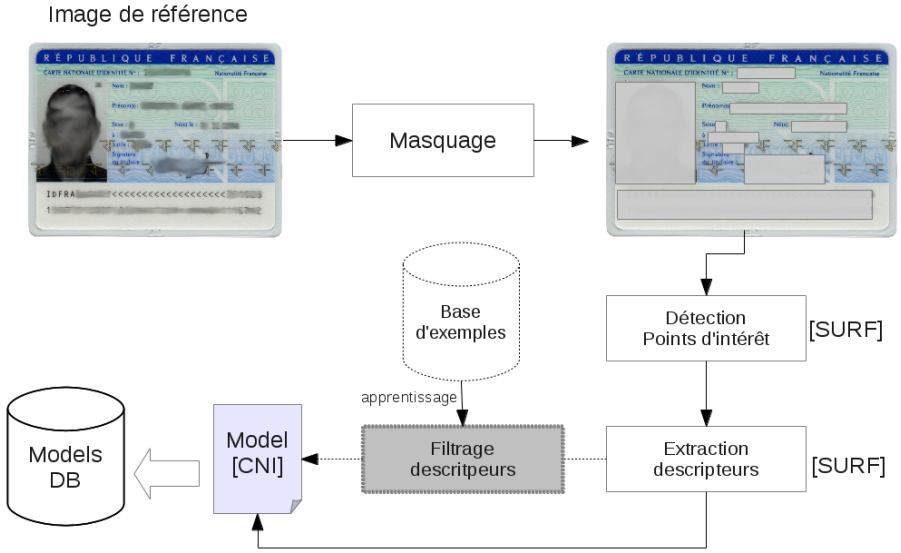
3. Approche proposée : Classification des documents complexes

3.1. *Création des modèles de référence*

Tout d'abord, chaque classe de document est représentée par un modèle pour en construire une base des modèles. La problématique principale est la non-disponibilité des documents de référence pour construire les modèles des classes. Nous proposons une méthode flexible qui permet la création du modèle à partir d'une seule image de référence ou à partir de plusieurs images de référence (si disponible).

La création d'un modèle se déroule comme suivant (voir la figure 1). Les points d'intérêt de l'image de référence sont extraits et caractérisés par SURF (Bay *et al.*, 2006) (ou d'autres descripteurs locaux). L'avantage de ces descripteurs est qu'ils sont invariants à la translation, au changement d'échelle et, plus généralement, aux transformations affines. La particularité des documents d'identité est la présence des informations fixes (les mentions : nom, prénom, etc.) et qu'ils partagent le même fond documentaire. En conséquence, il sera favorable de masquer les mentions variables ainsi que la photo pour éviter les faux mises en correpondance des points d'in-

Figure 1. *Le schéma de la création du modèle CNI*



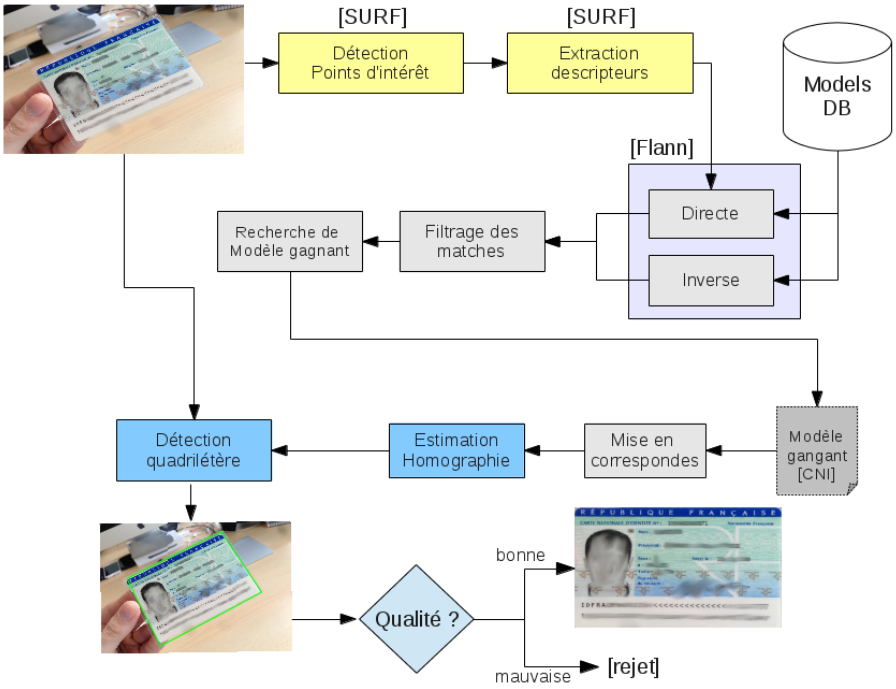
térêt. Le même processus est répété sur toutes les images d'apprentissage si elles sont disponibles pour obtenir un ensemble des points d'intérêt. Ensuite, cet ensemble est utilisé pour filtrer les points inutiles dans le modèle en gardant que les points de référence ayant des correspondances dans toutes les images d'apprentissage. En effet, avoir un nombre assez grand des images d'apprentissage pourra être équivalant au masquage des zones variables du document.
On dénote un modèle i par $M_i = \{D_{i1}, D_{i2}, ..., D_{in}\}$ où $|M_i| = n$ est le nombre des points d'intérêt retenus et D_{ij} sont les descripteurs extraits du point j.

3.2. *Chaîne de traitement*

3.2.1. *Classification*

L'idée globale du fonctionnement de notre classifieur de documents est illustrée dans la figure 2. Nous cherchons dans cette étape à trouver le modèle du document qui correspond au mieux à l'image requête (la classe gagnante). La classe gagnante est celle du modèle qui a le maximum des points d'intérêt mis en correspondance avec ceux de la requête. Les points d'intérêt de l'image requête de document sont caractérisés par la même méthode utilisée pour construire les modèles. Les points d'intérêt sont mis en correspondance contre tous les modèles dans la base. Nous utilisons les arbres KD aléatoires (Silpa-Anan et Hartley, 2008) dans l'algorithme Flann (Muja et Lowe, 2009) pour mettre en correspondance l'image requête avec un modèle. Il est

Figure 2. *La structure du classifieur de documents*



évident que la mise en correspondance modèle par modèle va être très longue et ne permettra pas le passage à l'échelle. Pour contourner ce problème, tous les modèles sont mis en concurrence dans la phase de mise en correspondance. De plus, nous appliquons une mise en correspondance bidirectionnelle :

– Les points de l'image requête contre l'ensemble des points des modèles. Cette mise en correspondance produit des *correspondances directs* ($m = \{m_i\}$)

– L'ensemble des points des modèles contre les points de l'image requête produisant des *correspondances inverses* ($m' = \{m'_i\}$).

Pour un modèle i, nous ne gardons que des correspondances valides V_i ; où :

• S_i est l'ensemble des correspondances symétriques où pour chaque correspondance directe dans m entre le point p_q dans l'image requête et le point p_m dans le modèle il existe la correspondance inverse dans m'.

• On calcule l'histogramme de la différence d'orientation de chaque couple des points des correspondances dans S_i. O_i est l'ensemble des correspondances ayant une différence d'orientation correspondant à la différence la plus fréquente dans l'histogramme.

• Finalement, l'algorithme RANSAC est appliqué (Chum *et al.*, 2003) pour trouver une transformation géométrique parmi les différents correspondances O_i en filtrant celles erronées, on obtient en sortie l'ensemble des correspondances finales V_i.

Les correspondances valides ne sont pas calculées préalablement pour toutes les classes. Premièrement, les modèles sont ordonnés par le ratio des points du modèles mis en correspondance avec les points requêtes $r_i = \frac{|m_i|}{|M_i|}$. Le modèle avec le ratio le plus élevé est le premier candidat à être la classe gagnante : $g = \underset{i}{argmax}(r_i)$. Ainsi, on examine les autres modèles dans un ordre décroissant des ratios :

1) si $|m_i| < |V_g|$ alors on examine le modèle suivant ; sinon
2) on calcule S_i, si $|S_i| < |V_g|$ alors $i = i + 1$ et on revient à l'étape 1 ; sinon
3) on calcule O_i, si $|O_i| < |V_g|$ alors $i = i + 1$ et on revient à l'étape 1 ; sinon
4) on calcule V_i, si $|V_i| < |V_g|$ alors $i = i + 1$ et on revient à l'étape 1 ; sinon
5) $V_g = V_i$, $i = i + 1$;
6) répéter pour tous les modèles

L'avantage de cet algorithme est d'éviter l'application systématique de tous les filtres, et surtout de l'algorithme RANSAC qui peut être assez long. Bien que cette première étape permet de trouver la classe de document, une deuxième mise en correspondance entre les points requêtes et les points du modèles (seuls) est nécessaire afin de retrouver les appariements avec plus de précision. Ensuite, le document sera localisé et extrait par une transformation géométrique obtenue de l'étape RANSAC.

3.2.2. *Extraction et estimation de la qualité (rejet)*

La méthode RANSAC permet d'estimer l'homographie H de la transformation géométrique des points d'intérêt de l'image requête à ceux du modèle gagnant :

$$H = \begin{pmatrix} a_{11} & a_{12} & t_1 \\ a_{21} & a_{22} & t_2 \\ v_1 & v_2 & 1 \end{pmatrix} \qquad [1]$$

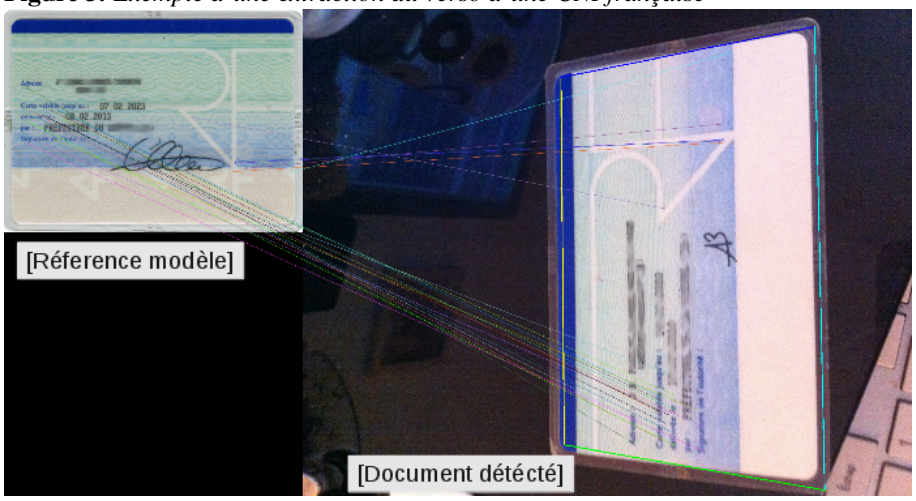
La figure 3 illustre un exemple de cette détection (les correspondances + localisation du document). L'objective n'est pas seulement la localisation du document mais aussi :

a) extraction des informations utiles (nom, prénom, date de naissance, ..),
b) vérification de l'authenticité de ce document.

De plus, notre système de lecture des documents d'identité n'impose aucune contrainte au type de document en entrée. En conséquence, le classifieur peut faire face à des classes inconnues qui n'apparaissent pas dans la base des modèles. Pour ces deux raisons, il est très important d'avoir une capacité de rejet. On se base sur la homographie H et le quadrilatère de document pour estimer si on doit rejeter le document. Le quadrilatère du document à extraire est calculé par :

$$Q_r = H \times Q_m \qquad [2]$$

Figure 3. *Exemple d'une extraction du verso d'une CNI française*



où Q_r est le quadilatère détécté et Q_m est celui du modèle. On dénote les quatre sommets du quadilatère par $A, B, C\ et\ D$ et les quatre côtés $[AB], [BC], [CD]\ et\ [DA]$. L'homographie calculée doit préserver l'orientation des sommets du modèle dans le quadrilatère détecté. Ainsi, l'image requête est *rejetée* (considérée comme classe inconnue) si le déterminant d de l'homographie est négatif (voir figure 4.b). Le déterminant est calculé par :

$$d = a_{11} \times a_{22} - a_{12} \times a_{21} \qquad [3]$$

En outre, le quadrilatère idéal doit s'approcher d'une forme rectangulaire. Toutefois, la perspective associée à la prise de photo déforme souvent ce rectangle en trapèze isocèle (une seule perspective) ou en parallélogramme (deux perspectives). Une image requête sera *rejetée* si le quadrilatère détecté ne répond pas aux conditions suivantes (comme l'exemple a) de la figure 4) :

1) Au moins un couple des côtés opposés est parallèle (avec une tolérance de cinq dégrées)

$$(\widehat{\overrightarrow{[AB]}} - \widehat{\overrightarrow{[CD]}}) < 5^\circ \Rightarrow [AB] \parallel [CD]\ \ ou \qquad [4]$$

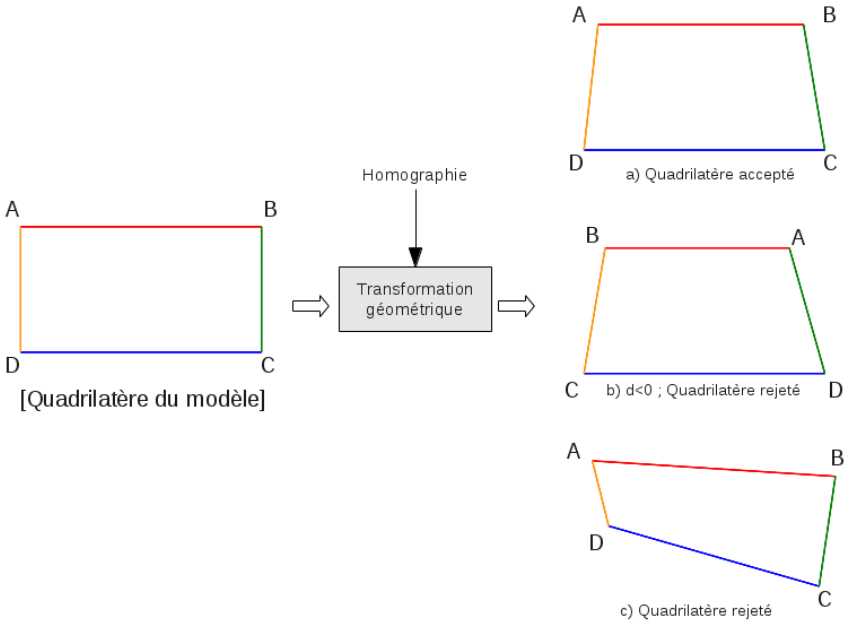
$$(\widehat{\overrightarrow{[AD]}} - \widehat{\overrightarrow{[BC]}}) < 5^\circ \Rightarrow [AD] \parallel [BC] \qquad [5]$$

2) La différence moyenne entre les angles opposés est inférieure à dix dégrées

$$\frac{(\hat{A} - \hat{B}) + (\hat{C} - \hat{D})}{2} < 10^\circ;\ \ si\ \ [AB] \parallel [CD]\ \ et \qquad [6]$$

$$\frac{(\hat{A} - \hat{D}) + (\hat{B} - \hat{C})}{2} < 10^\circ;\ \ si\ \ [AD] \parallel [BC] \qquad [7]$$

Figure 4. *Exemples des quadrilatères calculés à partir du quadrilatère modèle*



3) La Perpendicularité moyenne des quatre sommets est inférieure à vingt cinq dégrées

$$\frac{(\hat{A}-90)+(\hat{B}-90)+(\hat{C}-90)+(\hat{D}-90)}{4} < 25^{\circ} \qquad [8]$$

4. Expérimentations

4.1. *Bases de tests*

Nous considérons neuf classes dans nos expérimentations :

1,2 : les recto/verso de la carte nationale d'identité française (CNI)
3,4 : la première page du passeport français ancienne/nouvelle génération (PAS1/2)
5,6 : les recto/verso du titre de séjour français ancienne génération (TS1)
7,8 : les recto/verso du titre de séjour français nouvelle génération (TS2)
9 : le recto du permit de conduire français nouvelle génération (DL)

Tableau 1. *Évaluation de l'extraction des documents par classe*

Classe	#Exemples	Chevauchement %	Erreur %
CNI_recto	35	96.2	0
CNI_verso	12	84.6	25
PAS1	9	95.9	0
PAS2	13	96.7	0
TS1-recto	18	94.6	5.6
TS1-verso	9	93.5	11.1
TS2-recto	12	93.2	0
TS2-verso	15	94.4	8.3
DL	10	93.5	0
total	*133*	*94.1*	*4.5*

Un modèle est créé par classe à partir d'une image de référence scannée à 300 dpi. Nous disposons des images réelles des documents venants principalement des photos d'un appareil mobile. Ces images constituent deux bases de tests :

– Base1 : un total de 133 images des neuf classes.

– Base2 : un total de 2004 images des neuf classes en plus de 490 exemplaires des classes inconnues (rejet).

4.2. *Évaluation de l'extraction de documents*

Nous avons manuellement étiqueté tous les documents de la Base1 afin de définir le quadrilatère de vérité de chaque document dans son image. La vraie classe est associée à chaque image afin d'évaluer l'étape de l'extraction indépendamment du comportement du classifieur. On calcul le pourcentage de chevauchement entre le vrai quadrilatère et celui détecté en forçant la bonne classe. Un document est considéré comme bien extrait si ce pourcentage dépasse 90%. On peut constater du tableau 1 que pour les 133 documents on obtient 94.1% de chevauchement en moyenne pour seulement 4.5% d'erreurs. Nous remarquons que l'extraction est moins performante quand il s'agit des versos des pièces d'identité (CNI_verso, TS1_verso, TS2_verso). Cela est dû au fait que les versos contiennent souvent très peu d'informations visuelles comparés aux rectos.

4.3. *Évaluation de la classification de documents*

Le classifieur attribue une classe à chacune des images dans la Base2. L'ensemble de 490 documents inconnus au classifieur contient une grande variété des documents d'identité européens et internationaux. La figure 5 montre quatre exemples des documents. Les deux premiers appartiennent aux classes connues (Titre de séjour et Carte

Figure 5. *Exemple des résultats du classification et de la localisation des documents connus et inconnus*

TS2-recto

DL

[Documents correctement classés et localisés]

[Documents correctement rejetés]

Nationale d'identité correctement classés), pendant que les deux autres sont des pièces inconnues au classifieur (correctement rejetés). En effet, seulement une seule pièce inconnue n'a pas été rejetée par le classifieur. En plus, le classifieur a une précision global de 99.95% pour un rappel de 94.76%. L'évaluation du classifieur est présentée dans le tableau 2, où on mesure l'accuracy pour chacune des classes. Nous remarquons de ces résultats que le classifieur est moins performants pour les versos où il y a moins d'informations visuelles ce qui conduit souvent à des mauvaises mise en correspondance des points d'intérêt.

5. Conclusion et travaux futurs

Nous avons proposé dans ce papier un système de classification d'images de documents sur un fond complexe. Le document est associé à une classe donnée en maximisant le nombre d'appariements entre l'image requête et l'image référence (ou le modèle) de chaque classe. Cet appariement est effectué entre les points d'intérêt détectes sur les deux images avec une métrique de distance basée sur les descripteurs locaux calculés autour des points. Cette étape d'appariement global et grossier qui vise à associer le document à une classe gagnante est suive par une étape d'appariement

Tableau 2. *Évaluation de la performance du classifieur*

Classe	#Exemples	Accuracy%
CNI_recto	324	99.4
CNI_verso	325	87.1
PAS1	86	98.8
PAS2	318	95.3
TS1-recto	136	99.3
TS1-verso	143	93.7
TS2-recto	304	93.4
TS2-verso	360	94.4
DL	8	100
total	*2494*	*95.75*

fin qui, quant à elle, permet de localiser de façon précise le document dans l'image requête. Le système a été évalué sur un problématique de classification des document d'identité français. Les résultats montrent des très bon taux de rappel et de précision.

Nos travaux futurs serons consacrés à l'amélioration de l'approche proposé sur plusieurs axes :

– trouver un mécanisme de détection de points d'intérêt permettant d'augmenter le nombre de points détectés sur des documents avec peu d'éléments visuels marquants (comme pour le verso de la carte nationale d'identité française). Nous allons évaluer, entre autre, la pertinence des méthodes d'échantillonnage basé sur une grille sur ce type de document.

– améliorer la qualité des correspondances valides des points d'intérêts en appliquant des filtres supplémentaires (comme par exemple en utilisant l'histogramme de l'échelle).

– limiter dans certains cas les transformations géométriques prises en compte dans la mise en correspondance des points d'intérêt par RANSAC aux transformations rigides (rotation, translation, échelle). Ceci aide à éviter d'appliquer, à tort, des corrections de perspectives dans les cas où le périphérique d'acquisition ne produit pas des transformations projectives sur l'image de document (ex. scanner).

Remerciements

Ce travail est subventionné par le projet ANR IDFRAud [1] (ANR-14-CE28-0012) et cofinancé par la direction générale de l'armement (DGA). Les auteurs tiennent à remercier les membres du pôle Document de la société AriadNEXT pour leurs participations.

1. http ://www.idfraud.fr

6. Bibliographie

Augereau O., Journet N., Domenger J.-P., « Reconnaissance et Extraction de Pièces d'identité », *Colloque International Francophone sur l'Écrit et le Document*, Bordeaux, France, p. 179-194, March, 2012.

Bagdanov A., Worring M., « Fine-grained document genre classification using first order random graphs », *Proceedings. Sixth International Conference on Document Analysis and Recognition*, p. 79-83, 2001.

Bay H., Tuytelaars T., Van Gool L., « SURF : Speeded Up Robust Features », *in* A. Leonardis, H. Bischof, A. Pinz (eds), *Computer Vision – ECCV 2006*, vol. 3951 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, p. 404-417, 2006.

Cesarini F., Lastri M., Marinai S., Soda G., « Encoding of modified X-Y trees for document classification », *Proceedings. Sixth International Conference on Document Analysis and Recognition*, p. 1131-1136, 2001.

Chen S., He Y., Sun J., Naoi S., « Structured document classification by matching local salient features », *2012 21st International Conference on Pattern Recognition*, p. 653-656, Nov, 2012.

Chum O., Matas J., Kittler J., « Locally Optimized RANSAC. », *in* B. Michaelis, G. Krell (eds), *DAGM-Symposium*, vol. 2781 of *Lecture Notes in Computer Science*, Springer, p. 236-243, 2003.

de las Heras L.-P., Terrades O. R., Llados J., Fernandez-Mota D., Canero C., « Use case visual Bag-of-Words techniques for camera based identity document classification », *2015 13th International Conference on Document Analysis and Recognition*, IEEE, Tunis, Tunisia, p. 721-725, 2015.

Diligenti M., Frasconi P., Gori M., « Hidden tree Markov models for document image classification », *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 25, n° 4, p. 519-523, April, 2003.

Eglin V., Bres S., « Document page similarity based on layout visual saliency : application to query by example and document classification », *Proceedings. Seventh International Conference on Document Analysis and Recognition*, p. 1208-1212, Aug, 2003.

Hu J., Kashi R., Wilfong G., « Document image layout comparison and classification », *Proceedings of the Fifth International Conference on Document Analysis and Recognition*, p. 285-288, Sep, 1999.

Kumar J., Doermann D., « Unsupervised Classification of Structurally Similar Document Images », *2013 12th International Conference on Document Analysis and Recognition*, p. 1225-1229, Aug, 2013.

Muja M., Lowe D. G., « Fast Approximate Nearest Neighbors with Automatic Algorithm Configuration », *International Conference on Computer Vision Theory and Application VISSAPP'09)*, INSTICC Press, p. 331-340, 2009.

Sebastiani F., « Machine Learning in Automated Text Categorization », *ACM Comput. Surv.*, vol. 34, n° 1, p. 1-47, March, 2002.

Shin C., Doermann D., « Document Image Retrieval Based on Layout Structural Similarity », *International Conference on Image Processing, Computer Vision, and Pattern Recognition*, p. 606-612, 2006.

Silpa-Anan C., Hartley R., « Optimised KD-trees for fast image descriptor matching », *CVPR 2008. IEEE Conference on Computer Vision and Pattern Recognition*, p. 1-8, June, 2008.

Simon M., Rodner E., Denzler J., « Fine-grained Classification of Identity Document Types with Only One Example », *14th IAPR International Conference on Machine Vision Applications (MVA)*, IEEE, Tokyo, Japon, p. 126-129, 2015.

Yang G., Stewart C. V., Sofka M., ling Tsai C., « Alignment of challenging image pairs : Refinement and region growing Starting From Single keypoint correspondence », *IEEE TRANS. PATTERN ANAL. MACHINE INTELL*, 2007.

Yang Y., Liu X., « A Re-examination of Text Categorization Methods », *Proceedings of the 22Nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '99, ACM, New York, NY, USA, p. 42-49, 1999.