

Proposition d'IMPACT, Hologram. Industries

Entreprise : Hologram. Industries, <http://www.hologram-industries.com/>,
<https://keesingreferencesystems.com/>

Encadrants : Amine OUDDAN, amine.ouddan@keesing-technologies.com, 0164763100

Mathieu HOARAU, mathieu.hoarau@keesing-technologies.com, 0643004962

Clarisse MANDRIDAKE, clarisse.mandridake@keesing-technologies.com

Condition de réalisation : Convention Impact/Prise en charge des 3 déplacements prévisionnels : lancement, point intermédiaire et présentation des résultats/livraison des codes.

Projet : Classification automatique de type de pièce d'identité

Le but du projet est d'explorer les méthodes de classification du type de documents d'identité à partir de scans multi-canaux (UV/IR/visible) parmi une base de plusieurs centaines de documents (voire milliers). Le projet constitue la phase exploratoire de la première brique d'une solution d'authentification de documents d'identité et de son porteur.

a) Redressement et rognage (crop) du document dans le scan :

Il faudra avant tout extraire le document d'un scan élargi afin d'en sélectionner uniquement la zone comportant le document et le redresser (deskewing).

b) Première classification parmi les 3 premiers types principaux

Les documents d'identité sont catégorisés en 3 formats primaires principaux (les plus communs)

- Type passeport
- Type carte d'identité nationale française
- Type nouveau permis de conduire (ou format carte vitale)

De simples critères sur les dimensions du document devraient suffire pour cette étape qui permet de faire un premier tri.

c) Calcul des descripteurs du document

Etant donné la diversité des types de documents, il faudra prévoir plusieurs types de descripteurs qui s'avéreront plus pertinents sur certains types de documents que sur d'autres et vice-et-versa.

d) Décisionnel

Nous pouvons imaginer une recherche séquentielle dans la base de documents de référence et des comparaisons des descripteurs qui aboutira à une décision basée sur le meilleur score mais le temps de calcul pourra s'avérer rédhibitoire alors dans ce cas une stratégie plus intelligente pourra être envisagée.

Note : dans le cadre de ce projet, les parties les plus intéressantes pour l'étudiant sont les c et d. Nous pouvons donc imaginer nous concentrer séparément sur les 3 types pour lesquels candidats et suspects sont correctement normalisés et entamer le travail sur c et d, sur la fin du projet l'étudiant pourra compléter par les étapes a et b pour achever la chaîne.

Livrables :

- To + 40h : Plan de travail et choix des approches à mettre en œuvre
- To + 80 h : Avancement, résultat obtenu sur un sous- ensemble de documents.
- To + 120 : Documentation complète décrivant l'approche et les résultats obtenus sur un Corpus de 50 documents.
 - Ce document doit décrire les résultats expérimentaux (paramétrage, taux de réussite (Rappel et Précision))
 - Eventuellement les contraintes matériels (3 types de scanner : PRMC, ComboSmart, 3M)
 - Codes sources en C++ fonctionnel, sous opencv 2.4, Windows 64 bit, Visual Studio 2012)

Compétences désirées : Traitement d'images, Classification d'images, méthodes probabilistes d'aide à la décision multi-critères, librairie OpenCV, développement C++.