

Proposition de Stage

Campagne 2017



Sujet Stage :

Traitement de l'image pour la détection de faux documents

Contexte du travail proposé :

Les entreprises, les administrations et les particuliers échangent de très nombreux documents, qu'ils soient sous forme papier ou, de plus en plus, en format numérique. Ces documents (factures, certificats médicaux, bulletin de salaire, justificatifs...) peuvent être modifiés, ou falsifiés, très facilement à l'aide de logiciels de retouche d'image par exemple. Dans le cadre de la lutte contre la fraude, il est important de pouvoir certifier l'authenticité des documents échangés et de détecter les modifications apportées au document.

Pour cela, deux approches coexistent dans les recherches actuelles : la détection des modifications graphiques et la détection des fausses informations.

La première approche correspond à de l'analyse d'image et cherche à détecter des indices graphiques de fraude, comme la différence d'inclinaison, de taille, d'alignement ou de bruit d'un caractère par rapport aux autres, la différence de police ou d'espacements des caractères au sein d'un mot, le décalage d'une ligne par rapport aux marges du document ou encore l'inclinaison différente des lignes les unes par rapport aux autres. De nombreuses études sont également menées sur les encres utilisées pour l'impression des documents et la qualité des scanners, ce qui permet de détecter des insertions de fragments de textes ou d'images dans les documents analysés.

La deuxième approche consiste à analyser les informations contenues dans le document. Il s'agit pour cela de réaliser une reconnaissance de caractères sur les images de documents afin d'obtenir le texte, d'en extraire l'information, c'est-à-dire d'annoter le texte pour relever des éléments précis comme les entités nommées (noms d'organisations ou de personnes, lieux, dates...) ou les valeurs, et de modéliser les informations des documents afin de les comparer à des informations jugées vraies que l'on peut extraire du Web ou de bases de données.

Résumé du travail proposé

L'objectif du stage est de fournir des outils nécessaires à la réalisation d'une interface pour la détection des faux documents.

En collaboration avec le développeur de l'interface, le ou la stagiaire devra fournir :

- un module permettant de lier le texte extrait du document à son emplacement dans l'image. Pour cela, il/elle pourra s'appuyer sur les résultats d'OCR (reconnaissance de caractères) permettant de situer le caractère reconnu. Il s'agira par exemple de créer un

cadre en surcouche de l'image encadrant l'emplacement du texte sélectionné par l'utilisateur.

- un module permettant de localiser et de mettre en évidence sur l'image des informations selon un score de fiabilité pré-déterminé.

Le/la stagiaire sera amené(e) à adapter certains autres modules préexistants en collaboration avec le développeur de l'interface.

Déroulement du stage

Le stage débutera par une découverte de la problématique du traitement des documents et par la prise en main d'outils existants. L'apprentissage du langage Python sera probablement nécessaire. Le stagiaire devra donc être prêt à apprendre un nouveau langage de programmation, s'il n'a jamais fait de python.

Le développement des tâches décrites ci-dessus se fera selon une méthodologie Agile. Le stagiaire devra avoir une bonne capacité de compréhension des besoins utilisateurs, et devra travailler en équipe.

Le stage s'effectuant dans un laboratoire de recherche, le stagiaire aura la possibilité de participer à certaines activités du laboratoire (séminaires...)

Informations complémentaires :

Encadrante : Chloé Artaud

Date de début du stage : 17 avril 2017

Durée du contrat : 6 semaines

Contacts – liens :

e-mail : chloe.artaud@univ-lr.fr