Reconnaissance Optique des Caractères

Master 2 IA

LAURENT Thomas

Années: 2018 - 2019

Contents

Part I

Introduction

Chapter 1

l'écriture d'avant à maintenant

Depuis de nombreux siècles, l'Homme a su écrire des textes dans bien de nombreux langages.
Selon certain récits, l'écriture serait né en Mésopotamie durant l'an 3300 avant Jésus christ et étant composé de signes ou de pictogrammes représentant des mots ou des concepts. Ses textes était gravé dans la pierres, taillé dans le bois, écrit avec des roseaux taillées en pointes sur les murs des cavernes ou plus tard écrit sur du papier (ou des végétaux).
C'est ainsi que la connaissance et le patrimoine pu se conserver et être apprit par d'autres sociétés via dans un première instant la ré-écriture des textes. La ré-écriture n'étant jamais parfaite, la forme des caractères étaient différents d'un endroit à un autre, ce genre de transports ont notamment donné les chiffres tel qu'on l'ai connait aujourd'hui.
Depuis cette époque, de nombreux documents (de tout type) ont vue le jour, et même à ce jour nous continuons d'en créer.
Depuis la digitalisation et le format dématérialisé, l'informatique joue une part importante dans le stockage et le partagent de l'information, mais néanmoins la forme matérialisé des document ont une durée de vie limité, sur certaines surface l'ancre peut disparaitre ou la nature qui peut effacer certaines frises ou gravures, la restauration de ses documents est un processus utilisé mais qui a une limite.
L'invention du numériseur a permit de digitaliser certains document comme des anciens livres ou des documents officiel. Les documents officiel sont des pièces qui ont besoin d'être validé pour prouver leurs authenticité, mais nous ne pouvons pas simplement envoyer une image d'une pièce d'identité à une personne qui s'occupe de vérifier son authenticité. Pour vérifier qu'une série

de chiffres est valide, l'assistante devrait lire la série de chiffres recopier dans une base de donné, mais cette opération de peut être fait manuellement tout en conservant une vitesse de traitement rapide.
D'où l'informatique où l'intelligence artificiel donne une solution d'automatisation du traitement des images et de la récupération de données.

Chapter 2

Problématique

La reconnaissance des caractères est un problème de reconnaissance de patterns, dans une image l'OCR doit savoir distinguer un ensemble de formes formant des probables caractères. L'effort est de pouvoir analyser l'écriture humaine comportant beaucoup de différences d'un individu à un autre, la taille du caractère, sa longueur ou hauteur joue dans la reconnaissance de probable faux positif.
Prenons deux technique d'écriture, l'une étant l'écriture lié qu'on apprend à l'école et l'autre l'écriture que je vais appelé espacé qui reprend le format des caractères affiché sur l'écran d'un ordinateur. les deux textes ayant la même interprétation pour l'humain qui l'ai lit, l'OCR peut tout de même donner deux résultats différent lors du processus de transformation.
Comment l'intelligence artificiel a pu trouver une solution à ce problème via l'apprentissage et les réseaux de neurones.

Part II

État de l'art

On parle du premier OCR créé dans le monde en parlant des travaux de l'américain *Charles R Carey* en 1870 qui inventa un scanneur rétinien appliquant des patterns en mosaïque et étant appliqué sur une image envoyé en entré.

Dans certaines littératures on parle du premier OCR en 1900 par un russe du nom de *Tyurin* qui a tenté d'aider sur des travaux de reconnaissance de symbole pour les personnes handicapé.

On parle du premier OCR automatique en 1940 lors de l'introduction à l'air digital (4 ans après la machine de *Turing*), dans un premier temps les travaux d'automatisation OCR ont était réalisé directement des caractères issue d'une machine ou ont était réalisé via un petit ensemble de papier manuscrit où les caractères était finement bien représenté et distingué des autres. La conversion des papiers en binaire était faible, en effet l'extraction des caractères en format vectoriel était une procédure légère.