HAL
open science

Lecture automatique d'un ticket de caisse par vision embarquée sur un téléphone mobile

Rizlène Raoui-Outach, Cécile Million-Rousseau, Alexandre Benoit, Patrick Lambert

To cite this version:

Rizlène Raoui-Outach, Cécile Million-Rousseau, Alexandre Benoit, Patrick Lambert. Lecture automatique d'un ticket de caisse par vision embarquée sur un téléphone mobile. RFIA 2016, Jun 2016, Clermont-Ferrand, France. hal-01319960

HAL Id: hal-01319960

https://hal.science/hal-01319960v1

Submitted on 23 May 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Lecture automatique d'un ticket de caisse par vision embarquée sur un téléphone mobile

Rizlène Raoui-Outach[1] Cécile Million-Rousseau[1] Alexandre Benoit[2] Patrick Lambert[2]

[1] AboutGoods Company 3 Espl. Augustin Aussedat 74960 Cran-Gevrier France
[2] LISTIC, Université Savoie Mont Blanc, B.P 80439 - 74944 Annecy-le-Vieux, France

rizlene.raoui@aboutgoods.net

Avril 2016

Résumé

L'objectif de ces travaux est de créer un système capable, à partir d'une simple vue prise par un smartphone, d'exploiter le contenu d'un ticket de caisse, et d'en extraire des informations telles que le point de vente, les produits achetés, leurs prix etc. On distingue deux étapes majeures : la reconnaissance optique de caractères sur des tickets qui peuvent être froissés, déchirés... et l'analyse sémantique afin d'identifier les produits achetés sans ambiguïté. Dans ce papier, nous présentons une solution qui effectue une acquisition "guidée" d'une image sur smartphone et une lecture du contenu sur un serveur distant.

Mots Clef

Segmentation de zone de textes, Lecture automatique de tickets de caisse

Abstract

This work addresses the automatic reading of sale receipts acquired with a smartphone's camera and the extraction of essential informations like the store brands, all the purchased products , their price etc. It is divided into two major tasks : the optical character recognition that is made complex due to the nature of the document often damaged, crumpled or torned and the semantic data analysis to identify purchased product without ambiguity. In this paper, we introduce a solution that enables a "guided" capture on smartphone and a receipt decoding on remote server.

Keywords

Text segmentation, automatic reading of sale receipts

1 Introduction

La start-up AboutGoods Company est spécialisée dans l'apport de solutions internet et mobiles pour les entreprises et le grand public afin d'améliorer la relation entre les consommateurs et les distributeurs. Pour y arriver, l'entreprise doit disposer de bases d'informations fiables et actualisées en permanence sur les produits de grande consommation. Pour cela, il est nécessaire de constituer une communauté d'utilisateurs déclarant leurs achats au quotidien, qu'ils soient consommateurs standards ou panélistes experts. Ces consommateurs produisent des données à forte valeur ajoutée. Actuellement, les instituts de sondage recrutent un nombre limité de consommateurs "panélistes" qui déclarent leurs achats au quotidien en les scannant un à un et en remplissant des formulaires. La réalisation d'un lecteur automatique de ticket de caisse permettrait d'élargir cette communauté tout en facilitant la tâche. Une application concurrente (skerou) proposait le même service, mais son temps de réponse était trop long pour l'utilisateur. Dans la suite, la section 2 décrit une méthode accompagnant l'utilisateur pendant l'acquisition de l'image du ticket de caisse en le localisant en temps réel, cette étape devant assurer une qualité d'image suffisante pour permettre l'analyse mais sans trop contraindre l'utilisateur. En section 3, nous décrivons une première méthode de localisation et de reconnaissance optique des caractères.

2 Localisation du ticket dans l'image

Dès que l'utilisateur positionne son smartphone au dessus du ticket, un algorithme cherche alors à détecter les bords et affiche par un cadre orange le résultat obtenu (Fig. 1).

FIGURE 1 – Détection en temps réel du ticket

Cette détection est effectuée sur une image recalée dynamiquement et sous-échantillonnée (en imposant une hauteur d'image de 256) obtenue par moyennage, ce qui permet non seulement d'alléger les calculs mais aussi d'atténuer le bruit d'image. L'objectif étant de détecter un objet blanc rectangulaire sur un fond quelconque, mais plus sombre que le ticket, et sachant que l'utilisateur sera forcé d'orienter le ticket dans le même sens que le capteur, le détecteur

a été élaboré en tirant parti de ces forts a priori.

Nous avons ainsi défini quatre masques (un pour chaque bord) (Fig. 2), sur lesquels sont définies 3 zones. Le fonctionnement du détecteur est expliqué pour le bord gauche, l'extension aux autres bords étant immédiate. Si l'on note μ_i et σ_i la moyenne et la variance de la zone i (notée z_i par la suite), un pixel appartiendra au bord gauche si :

— $\mu_1 < \mu_{1+3} + \Delta_1$ (z_1 bien plus sombre que z_{1+3})
— $\mu_3 > \mu_{1+3} + \Delta_2$ (z_3 bien plus claire que z_{1+3})
— $\sigma_3 < Th$ (z_3 homogène)

où Δ_1, Δ_2 et Th sont des constantes. La zone 2 est une zone neutre permettant une tolérance à la non-verticalité. Pour éviter les fausses et les doubles détections, nous ne gardons qu'un point par ligne, celui répondant le mieux aux critères choisis (maxi de $\mu_3 - \mu_1$).

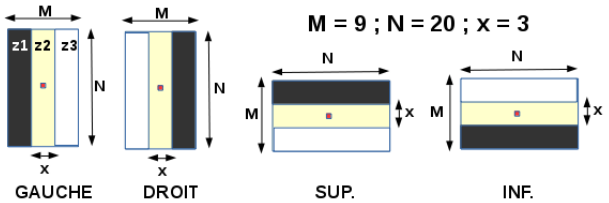

FIGURE 2 – Masques pour la détection des bords du ticket

Une fois les 4 bords obtenus, quelques considérations géométriques simples permettent d'obtenir le rectangle correspondant au ticket.

Cet algorithme fait intervenir un jeu de 7 paramètres : la composante couleur utilisée (R, V ou B), la hauteur M et la largeur N des fenêtres, la largeur x de la zone neutre, Δ_1, Δ_2 et le seuil Th sur l'écart-type. Pour valider un jeu de paramètres permettant d'obtenir un taux de détection maximal, nous avons créé une base d'une centaine d'images de tickets de caisse, prises dans des environnements différents et par des smartphones différents. Une large exploration des valeurs possibles nous a permis d'obtenir un taux de détection de 83% et de fausse alarme de 19% pour le jeu de paramètres suivant, avec la composante bleue :

$M = 9$, $N = 20$, $x = 3$, $\Delta_1 = 20$, $\Delta_2 = 40$, $Th = 35$

L'ajustement de l'image à une taille fixe et l'expansion de la dynamique assure la généricité de ces paramètres. L'algorithme fonctionne en temps réel, avec superposition du cadre détecté sur le flux image. L'utilisateur sait donc quand le ticket est correctement détecté et peut alors valider la saisie. Pour le développement de cette première étape, nous avons fait le choix d'utiliser la bibliothèque libre OpenCV de traitement d'images qui est facilement intégrable sur des terminaux mobiles.

3 Lecture du ticket

Une fois que l'acquisition de l'image du ticket à été faite sur le mobile, la localisation du texte et sa reconnaissance se font selon les étapes suivantes.

3.1 Localisation du texte

Dans notre situation particulière, un fond clair avec des caractères sombres, nous allons rechercher directement les boîtes englobantes contenant des lignes de texte. L'acquisition guidée de l'image assurant la quasi verticalité du ticket, nous avons utilisé des projections des intensités des pixels pour 1) localiser des blocs de texte sur toute la largeur du ticket (proj. horiz.), 2) découper ces blocs en sous-blocs (proj. vert.), 3) effectuer sur ces sous-blocs une segmentation en lignes (proj. horiz.).

3.2 Reconnaissance Optique de Caractère

Pour la reconnaissance du texte de chaque ligne, nous avons fait le choix d'utiliser un OCR libre : Tesseract [1]. Il nous permet d'obtenir des résultats relativement satisfaisants avec un taux de reconnaissance de 70%, dans des conditions favorables (mesures faites sur 15 tickets, distance de Levenshtein). Par la suite, une analyse sémantique permet de corriger un certain nombre d'erreurs (ex : "Tutal est facilement corrigé en Total"), en utilisant le modèle des N-grammes [3].

4 Conclusion et perspectives

Dans ce papier nous avons présenté une solution permettant d'optimiser l'acquisition d'une image d'un ticket de caisse en temps réel sur un terminal mobile, pour ensuite l'analyser et extraire son contenu sur un serveur distant. Ces premiers travaux ont permis de valider une démarche qui reste cependant sensible à la qualité du ticket. Pour gagner en robustesse, nous envisageons l'utilisation d'une approche basée sur le Deep Learning (réseau de neurones à grand nombre de couches). Des premiers tests avec un réseau convolutionnel profond [2], entraîné pour la détection et la reconnaissance de caractère dans des images de scènes naturelles (22000 caractères), sans réentraînement, nous ont permis d'obtenir, dans des conditions favorables, des résultats satisfaisants.

Une fois la lecture effectuée, il restera à mettre en place la dernière étape : l'analyse sémantique. Celle-ci a pour objectif d'améliorer la performance de lecture, en corrigeant les erreurs de l'OCR, mais elle permettra surtout une interprétation des libellés courts, étape rendue difficile par l'absence de normalisation dans l'utilisation de ces libellés courts (ex : SDW CLUB JBN = Sandwich club au jambon).

Références

[1] R. Smith, D. Antonova, and D. Lee, Adapting the tesseract open source ocr engine for multilingual ocr, *ACM International Workshop on Multilingual OCR*, July 2009.

[2] Jaderberg M. and Vedaldi A. and Zisserman A., Deep Features for Text Spotting, *ECCV*, 2014.

[3] I. Biskri, S. Delisle, Les n-grams de caractères pour l'aide à l'extraction de connaissances dans des BDD textuelles multilingues,*TALN* 2001.