

IDENTIFICATION ET COMPTAGE DE PIÉTONS

Berta Bescós Torcal

Julien Guichon

Dimitri Gominski

ABSTRACT

The range of pedestrian detection techniques, in order to achieve better performance and efficiency, is getting more and more complex, and thus the implementation of these algorithms can become time-consuming in conception phase.

By using 2 different approaches, we demonstrate the impact of acquisition context on the performance of these methods, and study the viability of a "naive" algorithm versus a modern version of object identification applied to pedestrian detection.

1. INTRODUCTION

Le développement de techniques fiables pour la détection de piétons dans les systèmes numériques est une problématique récente, et porteuse d'enjeux pour de nombreux domaines tels que la sécurité, la robotique, la maîtrise des flux humains en urbanisme...

Le coeur de la chaîne de traitement visant à associer à une zone spécifique de l'image une classe (piéton/non-piéton) repose sur 2 opérations élémentaires : la description et la décision. La description extrait des caractéristiques de ladite zone en fournissant des valeurs numériques caractérisant la forme, l'intensité et la texture. De cette manière on associe un ensemble d'informations quantifiées à un objet pour l'instant non-identifié. La décision, en synthétisant toutes ces valeurs numériques, donne un résultat binaire sur l'appartenance de l'objet à une classe.

Il va de soi que l'exhaustivité des descripteurs et la précision des classifieurs (organes de décision) sont la clé d'algorithmes infaillibles, mais elles se traduisent invariablement par plus de complexité, et si des précautions ne sont pas prises, par un temps de calcul allongé. Le choix du descripteur et du classifieur est un sujet sensible à de nombreuses contraintes, et dépend fortement des données disponibles pour la conception, des conditions d'utilisation de l'algorithme, et des performances attendues. L'étude des différents descripteurs et classifieurs s'est faite de manière empirique dans les 20 dernières années, et une large bibliographie est disponible pour les caractériser et préciser les conditions dans lesquels ils fournissent les meilleurs résultats.

Ce rapport décrit les performances obtenues avec 2 algorithmes courants tirés de 2 approches différentes du problème de l'identification et du comptage d'un flux épars de piétons.

2. ÉTUDE BIBLIOGRAPHIQUE CHRONOLOGIQUE

Dès 1985, T. Tsukiyama et Y. Shirai [1] ont proposé une technique rudimentaire d'identification humaine en travaillant avec l'intensité pixel par pixel.

En 1997 est proposée au MIT [2] une technique innovante de description des formes basée sur les ondelettes de Haar, associée à un classifieur SVM. L'idée de séparer descripteur et classifieur est depuis devenue un standard dans ce domaine.

En 1998, cette technique est perfectionnée [3] pour la rendre universelle en incluant un apprentissage le plus complet possible avec un set de données dédiées, pour permettre l'exécution dans des conditions variées. L'article introduit également la notion d'échelle variable pour la fenêtre de détection.

Dalal & Triggs [4] publient en 2005 un article de référence dans le domaine (+6000 citations). Ils expliquent une méthode complète, d'implémentation relativement simple, pour mettre en place une chaîne de détection de performances correctes. Leur méthode repose sur l'histogramme des gradients orientés pour décrire la forme d'un piéton, et la décision se fait avec une SVM (Support Vector Machine) linéaire.

Depuis, de nombreuses nouvelles approches ont été proposées, citons entre autres l'utilisation des informations de couleur [5], du mouvement [6], du bootstrapping (réutilisation des résultats, re-training) [5], de très larges sets de données d'apprentissages acquises par data-mining [7] *etc.*, avec la problématique du temps de calcul toujours au coeur du problème [8].

3. CONTEXTE

Dans le cadre de l'étude de faisabilité d'un algorithme de détection en temps réel à 95% de précision, nous disposons d'un set de données d'entrées dans des conditions relativement favorables (voir Figure 1). La caméra fournissant ces images est fixe, les seuls événements d'occlusion concernent les croisements de piétons, et la zone concernée est une section de route fréquentée uniquement par des piétons, ce qui limite les risques de confusion avec du mobilier urbain ou des véhicules. Les seules contraintes sont l'orientation légèrement verticale de la caméra (ce qui limite la surface identifiable du corps humain) et la profondeur de champ qui implique de devoir gérer les changements d'échelle.

Nous réalisons notre étude avec Matlab, qui permet un



Fig. 1. Données d'entrée

prototypage rapide des algorithmes.

Prenons une approche naïve. Le fond de l'image est fixe, les piétons sont les seuls objets en mouvement (en éliminant les mouvements parasites des feuilles et le bruit d'acquisition) : peut-on faire l'association objet en mouvement \Leftrightarrow piéton ? La partie 4 traite de cette approche.

Mais les diverses études sur le sujet 2 montrent que recourir au couple descripteur/classifieur augmente fortement la précision de l'algorithme. Nous avons donc également implémenté (voir partie 5) cette approche pour évaluer sa pertinence dans notre cas.

4. APPROCHE MORPHOLOGIQUE

A travers différentes opérations de filtrages non-linéaires par des éléments structurants, l'approche morphologique permet d'effectuer des traitements rudimentaires sur l'image. L'objectif est d'extraire des régions dans l'image ("*blobs*" dans la documentation anglophone) en les séparant sur une image binaire.

4.1. Implémentation

Pour se détacher du fond il faut commencer par extraire le mouvement. Cela peut se faire par moyennage sur une plage d'images consécutives, par seuillage de la différence entre 2 images, par représentation probabiliste de chaque pixel, par création d'une image temporelle le long d'une ligne...

Nous avons choisi d'implémenter l'approche morphologique avec cette dernière méthode. On définit donc une ligne de détection sur l'image, sur laquelle on élimine les pixels correspondant au fond par moyennage glissant sur 100 images. On passe en image binaire par seuillage RGB réglable, afin de pouvoir appliquer les opérations morphologiques, qui se déroulent dans cet ordre :

- multiples itérations de l'opération $X_k = (X_{k-1} \oplus B) \cap A^c$ jusqu'à ce que $X_k = X_{k-1}$. Cette opération sert à combler les trous.
- ouverture $X = (X \ominus B) \oplus B$.

4.2. Identification

L'identification des régions se fait par analyse de connectivité sur l'image binaire obtenue : un filtre parcourt l'image pour associer un label à chaque groupe de pixels, puis un rapide traitement regroupe les labels coïncidents en blocs.

5. APPROCHE DESCRIPTIVE

Le descripteur qui semble s'être imposé dans le domaine de la reconnaissance humaine de par sa relative simplicité (sans être le plus performant) est l'histogramme des gradients orientés (*HOG*).

Introduit et plébiscité par Dalal & Triggs [4] en 2005, l'HOG parcourt des cellules de petite taille, calcule le gradient (orientation + intensité) en chaque pixel, et crée un histogramme des orientations des gradients sur la cellule, pondérées par les intensités.

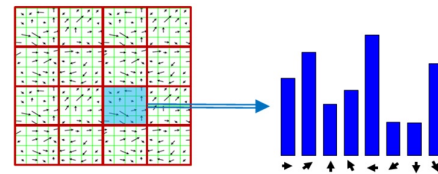


Fig. 2. Principe de l'HOG

L'HOG donne de cette manière une bonne idée de la forme générale sur la cellule, et surtout quantifie cette information sur un nombre de directions et une précision en intensité réglables.

5.1. Implémentation

Pour appliquer ce descripteur, on parcourt l'image avec une fenêtre de taille fixe, décalée à chaque itération d'un certain pas, qui a son importance pour éviter les détections redondantes tout en maintenant une bonne précision. Pour gagner en temps de calcul, et simplifier la gestion de l'échelle variable entre arrière-plan et avant-plan, nous faisons parcourir cette fenêtre sur une zone prédéfinie (milieu de l'image) où l'échelle peut être considérée comme constante (on peut facilement s'affranchir de cette hypothèse en utilisant une fenêtre à taille variable).

Le calcul de l'histogramme se fait sur des cellules, qui sont regroupées en blocs glissants sur lesquels on effectue une normalisation en norme 2 pour limiter les effets des changements de luminosité et de contraste sur l'image. L'histo-

gramme calculé sur N valeurs de directions, donne donc un vecteur de taille M , avec

$$M = NbBlocsParFenetre * TailleBloc * N$$

5.2. Classification

Pour associer à chaque vecteur de taille M un label binaire d'appartenance à la classe "piétons", nous avons choisi d'utiliser une classification via SVM. A partir d'un set de données d'apprentissage (vecteurs de taille M) associées à un vecteur de classe binaire, la SVM (linéaire dans notre cas) détermine le meilleur hyperplan pour séparer ces deux classes dans l'espace de dimension M correspondant.

L'apprentissage se fait avec des imagerie associées à des piétons et des imagerie associées au fond. Il est crucial d'être exhaustif dans cette phase d'apprentissage pour permettre à la SVM de gagner en sélectivité et donc en précision. Il a également été montré [4] que les imagerie fournies à la SVM dans le cas de la détection de piétons doivent fournir une information de contexte, il faut donc laisser de la marge autour des piétons.



Fig. 3. Imagerie d'apprentissage - piéton



Fig. 4. Imagerie d'apprentissage - fond

6. ÉVALUATION QUANTITATIVE

Pour évaluer et comparer les performances des algorithmes étudiés nous utiliserons l'outil classique de la matrice de confusion. Nous mesurons également le temps de calcul

pour estimer la compatibilité avec une exécution temps réel (cf partie 3). L'évaluation des performances se fait sur une séquence d'images des données d'entrée.

7. PAGE NUMBERING

Please do **not** paginate your paper. Page numbers, session numbers, and conference identification will be inserted when the paper is included in the proceedings.

8. ILLUSTRATIONS, GRAPHS, AND PHOTOGRAPHS

Illustrations must appear within the designated margins. They may span the two columns. If possible, position illustrations at the top of columns, rather than in the middle or at the bottom. Caption and number every illustration. All halftone illustrations must be clear black and white prints. Colors may be used, but they should be selected so as to be readable when printed on a black-only printer.

Since there are many ways, often incompatible, of including images (e.g., with experimental results) in a LaTeX document, below is an example of how to do this [?].

9. FOOTNOTES

Use footnotes sparingly (or not at all!) and place them at the bottom of the column on the page on which they are referenced. Use Times 9-point type, single-spaced. To help your readers, avoid using footnotes altogether and include necessary peripheral observations in the text (within parentheses, if you prefer, as in this sentence).

Fig. 5. Example of placing a figure with experimental results.

10. COPYRIGHT FORMS

You must include your fully completed, signed IEEE copyright release form when you submit your paper. We **must** have this form before your paper can be published in the proceedings.

11. REFERENCES

- [1] Toshifumi Tsukiyama and Yoshiaki Shirai, "Detection of the movements of persons from a sparse sequence of tv images.," *Pattern Recognition*, vol. 18, no. 3-4, pp. 207–213, 1985.
- [2] Michael Oren, Constantine Papageorgiou, Pawan Sinha, Edgar Osuna, and Tomaso Poggio, "Pedestrian detection using wavelet templates," in *Computer Vision and Pattern Recognition*, 1997, pp. 193–199.

- [3] Constantine Papageorgiou, Theodoros Evgeniou, and Tomaso Poggio, "A trainable pedestrian detection system," in *In Proceedings of Intelligent Vehicles*, 1998, pp. 241–246.
- [4] Navneet Dalal and Bill Triggs, "Histograms of oriented gradients for human detection," in *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Volume 1 - Volume 01*, Washington, DC, USA, 2005, CVPR '05, pp. 886–893, IEEE Computer Society.
- [5] Stefan Walk, Nikodem Majer, Konrad Schindler, and Bernt Schiele, "New features and insights for pedestrian detection.," in *CVPR*. 2010, pp. 1030–1037, IEEE Computer Society.
- [6] Paul Viola, Michael J. Jones, and Daniel Snow, "Detecting pedestrians using patterns of motion and appearance," *Int. J. Comput. Vision*, vol. 63, no. 2, pp. 153–161, July 2005.
- [7] Piotr Dollár, Zhuowen Tu, Hai Tao, and Serge Belongie, "Feature mining for image classification.," in *CVPR*. 2007, IEEE Computer Society.
- [8] Piotr Dollar, Serge Belongie, and Pietro Perona, "The fastest pedestrian detector in the west," in *Proceedings of the British Machine Vision Conference*. 2010, pp. 68.1–68.11, BMVA Press, doi :10.5244/C.24.68.