Descripteurs pour la reconnaissance de piétons

Features for Pedestrian Recognition

L. Leyrit T. Chateau J.-T. Lapresté

LASMEA
UMR 6602, CNRS, Université Blaise Pascal 63000 Clermont-Ferrand
prénom.nom@lasmea.univ-bpclermont.fr

Résumé

La reconnaissance de piétons dans les images est une tâche à part entière qui requiert l'utilisation d'outils particuliers. Parmi les descripteurs récents utilisés pour la détection de piétons, on trouve les ondelettes de Haar, les histogrammes d'orientation de gradients et les descripteurs binaires. Ce papier présente une comparaison des performances de ces trois différents descripteurs à partir d'une base d'images commune et d'un même classifieur. Nous présenterons également comment associer ces descripteurs de façon simple pour améliorer les taux de reconnaissance de piétons.

Mots Clef

Reconnaissance des formes, classification, descripteurs, ondelettes de Haar, histogrammes de gradients, piétons

Abstract

Pedestrian recognition in images is a full task which requires specific and accurate tools. Among the features used for pedestrian detection, we find Haar wavelets, histograms of oriented gradients and binary descriptors. This paper presents a comparison of these three different features with the same images dataset and the same classifier. Moreover we present how to associate these features in a simple way to improve pedestrian recognition rate.

Keywords

Pattern recognition, classification, descriptors, Haar wavelets, histograms of oriented gradients, pedestrians

1 Introduction

La reconnaissance de piétons dans les images a pris un essor avec l'apparition des caméras dans la vie de tous les jours. On les utilise aujourd'hui pour faire de la vidéo surveillance dans les magasins ou dans les lieux sensibles (aéroport, banques,...), pour développer des aides à la conduite dans les véhicules (évitement de collision, activation d'airbags,...),... . Toutefois, un piéton reste un objet difficile à reconnaître pour tous les systèmes habituels de classification et de reconnaissance d'objets dans les images. En effet, c'est un objet ayant une grande variabilité (taille des personnes, habillement,...), sans oublier tous les problèmes classiques liés au changement de pose et d'illumination.
Des méthodes ont été développées afin de reconnaître en temps réel les piétons dans les images. Dans [10], les auteurs utilisent des ondelettes de Haar avec un classifieur de type SVM. Dans [13], c'est une cascade de classifieurs AdaBoost qui est mise en place. Le système présenté dans [2] emploie des histogrammes de gradients pour détecter des personnes. Des approches avec des descripteurs binaires ont aussi été développées soit pour des classifieurs de type AdaBoost [7], soit pour des machines à noyaux [5].
Dans tous ces systèmes, le challenge est de reconnaître les piétons en évitant des fausses détections, mais aussi de développer des méthodes qui soient facilement transposables pour des applications temps réel. Dans cette optique, tous les descripteurs utilisés sont relativement rapides d'exécution.
De nouvelles approches ont été développées afin de combiner les avantages de plusieurs descripteurs. Dans [3], deux types de descripteurs sont utilisés : des ondelettes de Haar et une variante des histogrammes de gradients. Ces deux descripteurs sont représentés par des valeurs réelles et des classifieurs faibles sont simplement créés en établissant une règle de décision à partir d'un seuil ; un algorithme AdaBoost est utilisé en cascade afin de créer le classifieur fort final. Dans [9], une association de deux descripteurs est présentée pour la reconnaissance de voitures. Il s'agit de concaténer un descripteur rectangulaire (de type onde-

lettes de Haar) à un descripteur d'histogrammes de gradients orientés dans le cadre d'une cascade de classifieurs AdaBoost. Dans les deux cas, la fusion de deux descripteurs permet d'augmenter les scores de reconnaissance.
Dans cet article, nous débuterons par une rapide présentation des descripteurs couramment utilisés pour la reconnaissance de piétons à savoir les ondelettes de Haar et les histogrammes de gradients, ainsi que d'un descripteur binaire. Nous introduirons ensuite le classifieur utilisé dans le cadre de cet article, qui se base sur les machines à noyaux, et les méthodes de combinaison des trois descripteurs. La section 5 présentera les expérimentations menées et les résultats obtenus. Enfin, une conclusion sera donnée dans la partie 6.

2 Les descripteurs

Dans cette partie, nous allons introduire les descripteurs qui seront utilisés dans la partie 5. Il s'agit tout d'abord des ondelettes de Haar, puis des histogrammes de gradients et enfin d'un descripteur binaire de comparaison de pixels.

2.1 Les ondelettes de Haar

Il s'agit du descripteur le plus utilisé pour la reconnaissance de piétons [10, 12]. Dans l'article [10], les auteurs définissent un dictionnaire complet d'ondelettes pour la classification. Il s'agit du calcul des composantes des ondelettes dans trois directions principales : horizontale, verticale et diagonale (voir figure 1). La taille de ces ondelettes doit ensuite être adaptée à la résolution de l'objet à décrire. En effet, une taille trop petite ne retient que du bruit, tandis qu'une taille trop grande englobe l'objet sans en définir ses caractéristiques.

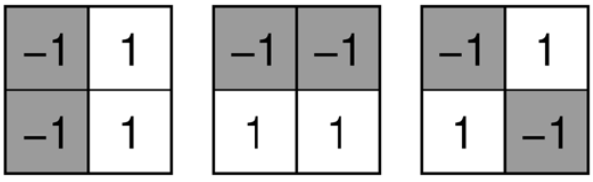

FIG. 1 – **Ondelettes de Haar utilisées pour la reconnaissance de piétons.**

Elles sont déclinées pour trois orientations (verticale, horizontale et diagonale).

2.2 Les histogrammes de gradients orientés

L'utilisation de ce type de descripteurs a été mise au goût du jour dans [2]. L'idée était d'utiliser une version simplifiée du descripteur SIFT [6] pour de la reconnaissance temps réel d'objet. Il s'agit d'utiliser un calcul de gradient de façon simple et efficace. L'image de l'objet à caractériser est découpée en plusieurs cellules pour lesquelles on comptabilise les occurrences de l'orientation du gradient dans un histogramme. Plusieurs versions existent : certaines agissent sur la normalisation des histogrammes [11], d'autres comptabilisent la magnitude du gradient au lieu des occurrences seules [1]. De même que pour les ondelettes de Haar, ce type de descripteur doit également être adapté à la taille et la résolution de l'objet à reconnaître, notamment pour la taille des cellules.

2.3 Le descripteur binaire de comparaison de pixels

Ces types de descripteurs simples se développent de plus en plus car ils ont l'avantage d'être rapides et d'assez bonne qualité pour des images de faibles résolution [7, 5]. Dans le cas de la reconnaissance de piétons, ils sont particulièrement bien adaptés au vu des applications : les piétons sont extraits d'images provenant de caméra avec des résolutions standard (640x480 pixels) et ne sont en général représentés que par quelques pixels. Nous utiliserons ici le descripteur présenté dans [5]. Il effectue une comparaison de l'intensité de certains pixels de l'image et renvoie une information binaire (voir figure 2). Les meilleurs couples de points sont choisis au préalable avec une sélection de variable par AdaBoost [4].

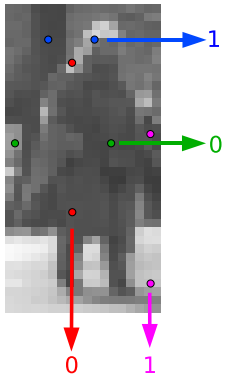

FIG. 2 – **Descripteurs par comparaison de niveaux de gris**

Pour un couple de points données, le descripteur retourne une valeur logique $d \in \{0,1\}$ *correspondant au résultat du test* $\{Intensité(point_1) \geq Intensité(point_2)\}$

3 Le classifieur

Généralement, les classifieurs utilisés en reconnaissance des formes sont soit des algorithmes de type Boosting, soit des machines à noyaux. Du fait de leur grande popularité

et des bons résultats obtenus en classification, nous utiliserons dans cet article une machine à noyau pour comparer les différents descripteurs présentés dans la section 2.
Dans cette partie, nous garderons des notations standard en représentant les étiquettes de sortie par un scalaire y qui pourra prendre deux valeurs discrètes possibles correspondant à la classe de l'objet : $y = -1$ pour les exemples négatifs (non-piétons) et $y = 1$ pour les exemples positifs (piétons). Le vecteur $\mathbf{x} \in \mathbb{R}^Q$ représente les caractéristiques d'entrée fournies par les descripteurs d'image. Notons $S \doteq \{(\mathbf{x}^i, y^i)\}_{i=1}^N$ l'ensemble d'apprentissage composé par N échantillons de vecteurs de primitives $\mathbf{x}^i$ associées à leur étiquette y^i correspondant à leur classe. La formulation générale des méthodes à noyaux est donnée par :

$$y = \text{sign}\left(\sum_{m=1}^{M} w_m \phi_m(\mathbf{x})\right) \tag{1}$$

$\{\phi_m(\mathbf{x})|m = 1...M\}$ sont des fonctions de base et $\{w_m|m = 1...M\}$ sont les poids associés. Nous proposons des fonctions de base non-linéaires telles que :

$$\phi_m(\mathbf{x}) = k(\mathbf{x}_m, \mathbf{x}) \tag{2}$$

où $k(\mathbf{x}_m, \mathbf{x})$ est une fonction noyau.
La règle de classification peut être écrite sous une forme plus compacte dans l'équation suivante :

$$y = \text{sign}(\mathbf{w}^T \phi(\mathbf{x})) \tag{3}$$

où $\mathbf{w}^T = (w_1, w_2, ..., w_M)$ est un vecteur poids et $\phi(\mathbf{x}) = (\phi(\mathbf{x}^1), \phi(\mathbf{x}^2), ..., \phi(\mathbf{x}^N))^T$. Pour entraîner le modèle (estimation de $\mathbf{w}$), nous avons l'ensemble d'apprentissage $S = \{(\mathbf{x}^i, y^i)\}_{i=1}^N$. Nous utilisons la norme Euclidienne pour mesurer l'erreur de prédiction dans l'espace des y, et donc le problème d'estimation revient à la formulation suivante :

$$\mathbf{w} := \arg\min_{\mathbf{w}} \{||\mathbf{w}^T \Phi - \mathbf{y}||^2\} \tag{4}$$

où $\Phi \doteq \left(\phi(\mathbf{x}^1), \phi(\mathbf{x}^2), ..., \phi(\mathbf{x}^N)\right)$ est la matrice de *design* et $\mathbf{y} \doteq (y^1, ...y^N)^T$ est le vecteur de classe de l'ensemble d'apprentissage.
L'estimation du vecteur de paramètres $\mathbf{w}$ défini dans l'équation (4) peut se faire en utilisant un critère au sens des moindres carrés :

$$\mathbf{w}_{ls} = \mathbf{y}\Phi^+ \tag{5}$$

où Φ^+ correspond à la pseudo-inverse de Φ.
Les vecteurs utilisés pour les fonctions de base sont généralement composés par un échantillon de l'ensemble d'apprentissage S^h. Il est également possible d'utiliser tout l'ensemble d'apprentissage et dans ce cas $M = N$. La matrice Φ est symétrique et le système peut être résolu plus efficacement par une décomposition de Cholesky. C'est ce classifieur qui sera utilisé dans la partie 5 ; nous avons fait le choix commun d'utiliser une Gaussienne comme fonction de base qui nous donne un modèle avec une fonction à base radiale (RBF) pour lequel le paramètre σ doit être ajusté :

$$\phi_m(\mathbf{x}) = \exp\left[-(\mathbf{x} - \mathbf{x}^m)^2 / \sigma^2\right] \tag{6}$$

4 Combinaison des descripteurs

Dans [3] et [9], deux types de descripteurs sont utilisés et concaténés afin d'améliorer les résultats de la classification. Au lieu de concaténer les différents descripteurs pour chaque image, ce qui donnerait un vecteur de caractéristiques de très grande dimension pour chaque objet, nous proposons dans un premier temps de fusionner les résultats des différents classifieurs (chacun utilisant un descripteur différent) par une simple moyenne algébrique. La règle de décision sera fournie par :

$$y_{combinaison1} = \frac{1}{C}\sum_{i=1}^{C} y_i \tag{7}$$

C représente le nombre de classifieurs fusionnés. Une deuxième méthode d'association consiste à choisir parmi les différentes réponses des classifieurs, celle dont la valeur absolue est maximale, soit :

$$y_{combinaison2} = y \text{ tel que } \|y\| = max\|y_i\|_{i=1...C} \tag{8}$$

5 Expérimentations et résultats

5.1 Méthodologie

Nous avons utilisé la base d'images présentée dans [8] dont quelques exemples sont montrés dans la figure 3. Cette base est subdivisée en cinq parties ; chacune contient 4500 images positives et 5000 négatives. Il s'agit d'images de luminance, de taille 36x18 pixels. Dans les images d'exemples positifs, les piétons se tiennent debout et sont entièrement visibles ; ils ont été pris dans différentes postures, et dans conditions d'illumination de fond variables. Chaque image de piéton a été aléatoirement décalée de quelques pixels dans les directions horizontale et verticale. Les images d'exemples négatifs représentent l'environnement urbain : bâtiments, arbres, voitures, panneaux de signalisations,... .
Pour les ondelettes de Haar, nous avons gardé les mêmes paramètres que dans [8]. Deux tailles d'ondelettes sont conservées (4x4 et 8x8) et calculées pour les trois orientations vues précédemment, et avec un recouvrement d'$\frac{1}{4}$ la taille de l'ondelette. Ce descripteur donne donc pour chaque imagette un vecteur de caractéristiques de taille 1x1755.
Pour les histogrammes de gradients, nous avons pris des

FIG. 3 – **Exemples de piétons (première ligne) et de non-piétons (deuxième ligne)** *(taille 36x18 pixels)* **.**

cellules de taille 3x3 sans recouvrement et des histogrammes à 8 barres. Nous obtenons donc un vecteur de caractéristiques de taille 1x576 pour une imagette donnée. Enfin, en ce qui concerne le descripteur binaire, la sélection de variables par AdaBoost a permis d'obtenir un vecteur de caractéristiques de taille 1x2468 pour chaque image.

5.2 Comparaison des trois descripteurs

Pour comparer les trois différents descripteurs, nous travaillons sur les trois bases d'apprentissage puis testons ces apprentissages sur les bases de test 4 et 5. Les images sont décrites par les trois descripteurs présentés dans la section 2. Des classifieurs sont entraînés à partir de chacune de ces descriptions sur les trois bases d'apprentissages, ce qui nous donne 9 apprentissages. Ces apprentissages sont ensuite testés en classification sur les bases de test 4 et 5. Les performances des apprentissages sont évaluées au moyen de courbes ROC (*Receiver Operating Curves*) qui représentent le taux de faux négatifs (des non-piétons pris pour des piétons) en fonction du taux de vrais positifs (des piétons bien reconnus en tant que tels) pour différents seuils de discrimination.
La figure 4 présente les courbes ROC pour ces apprentissages. Nous avons concaténé les résultats des trois classifieurs obtenus pour un type de descripteur donné. Nous obtenons ainsi trois courbes représentant les performances des apprentissages pour chacun des descripteurs testés (voir figure 4). Le premier constat est que ces trois apprentissages ont des performances proches. Toutefois, les histogrammes de gradients orientés semblent apporter de meilleurs résultats pour la classification de piétons. Le descripteur binaire, malgré son apparente simplicité, apporte des résultats quasi similaires. Et ce sont les ondelettes de Haar qui ont des performances moindres dans le cadre de cette application.

5.3 Combinaison de descripteurs

A partir des apprentissages réalisés, nous avons établis plusieurs combinaisons. La figure 5 présente l'association de trois classifieurs, chacun utilisant un descripteur différent. Les courbes de la figure 4 sont de nouveaux tracées afin d'établir une comparaison entre un classifieur seul et les deux types d'association des 3. Nous pouvons constater que les deux associations apportent un réel gain pour la reconnaissance de piétons. Mais il s'agit de la première méthode de combinaison par la moyenne qui atteint les meilleurs scores. Pour 10% de fausses alarmes, l'association moyennée des trois classifieurs (chacun basé sur un descripteur différent) atteint 90,20% de bonnes détections, le choix de la valeur maximale donne 84,04% de bonnes détections, tandis que l'apprentissage à partir des histogrammes de gradients n'obtient que 70,45% de bonnes détections.
On peut toutefois se demander si deux descripteurs ne suffiraient pas. La figure 6 montre l'association par la moyenne de deux descripteurs parmi les trois. Même si les résultats s'améliorent par rapport à l'utilisation d'un seul classifieur, c'est l'association des trois descripteurs qui est la plus performante.
Enfin, nous avons comparé cette méthode d'association par la moyenne algébrique aux résultats présentés dans [8]. Toutes les méthodes développées ont été apprises et testées sur la même base que nous avons utilisée. Nous avons retenu deux systèmes qui atteignent de bons taux de reconnaissance ; il s'agit d'une part d'un descripteur LRF (*Local Receptive Features*) avec un classifieur SVM (*Séparateur à Vaste Marge*) et d'autre part d'un descripteur avec ondelettes de Haar avec un classifieur SVM. Nous avons appris et testé la méthode de combinaison dans les mêmes conditions. Les résultats sont présentés en figure 7. Là encore, la méthode d'association par la moyenne obtient des scores de reconnaissance supérieurs à ceux présentés dans [8].

6 Conclusion et perspectives

Les résultats présentés dans ce papier ont permis d'établir une comparaison entre trois descripteurs pour une application particulière de reconnaissance de piétons dans des images de faible résolution. Les histogrammes de gradients orientés donnent de meilleurs résultats que les ondelettes

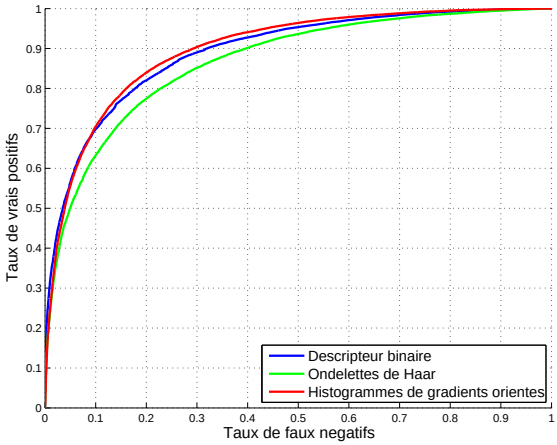

FIG. 4 – **Comparaison des trois descripteurs.**

Les histogrammes de gradients donnent de meilleurs résultats, même si le descripteur binaire s'approche des mêmes performances.

de Haar. On notera également que le descripteur binaire utilisé s'approche des bons résultats des histogrammes de gradients, et ceci, malgré leur apparente simplicité.
Nous avons également présenté deux méthodes d'association des différents descripteurs qui donnent des résultats très prometteurs. En effet, les taux de bonnes détections s'améliorent de façon significative. C'est la méthode de combinaison par moyenne algébrique qui donne de meilleurs résultats. Toutefois, cette méthode de combinaison est somme toute très simple et de futures travaux devraient permettre de confirmer ces résultats en améliorant la règle d'association.

Remerciements

Nous remercions D. Gavrila and S. Munder pour la mise à disposition de leur base de données.
Ce travail est présenté dans le cadre du projet LOVe (*Logiciels d'Observation des VulnérablEs*) qui propose de contribuer à la sécurité routière, et en particulier à celle des piétons. Le but est d'obtenir à terme des logiciels fiables d'observation des vulnérables.

Références

[1] Julien Bégard, Nicolas Allezard, and Patrick Sayd. Détection de piétons temps-réel en milieu urbain. In *Reconnaissance des Formes et Intelligence Artificielle (RFIA)*, 23 January 2008.

[2] Navneet Dalal and Bill Triggs. Histograms of Oriented Gradients for Human Detection. In *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 886–893, San Diego, California, USA, 2005.

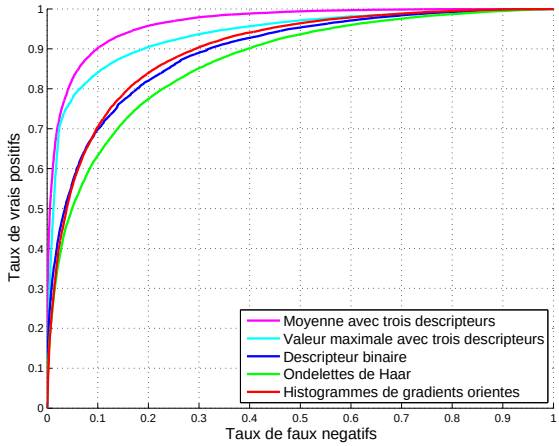

FIG. 5 – **Association de trois classifieurs par rapport à un apprentissage à partir d'un seul descripteur.**

L'association des trois descripteurs par la moyenne augmente très nettement les scores de reconnaissance de piétons.

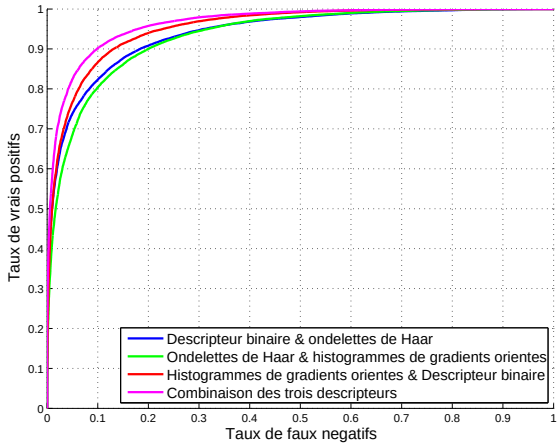

FIG. 6 – **Association par la moyenne de deux, puis trois classifieurs.**

C'est l'association des trois descripteurs qui permet d'obtenir le meilleur apprentissage.

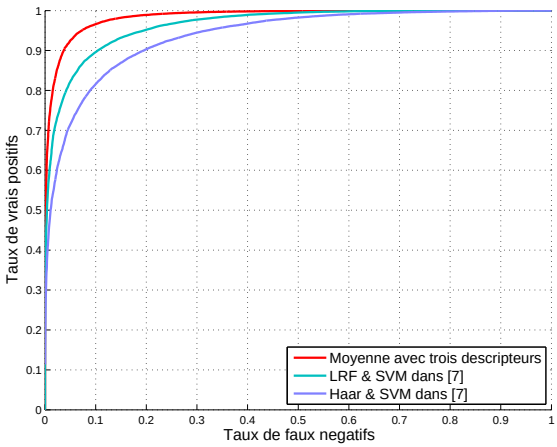

FIG. 7 – **Comparaison de la méthode d'association par la moyenne aux méthodes présentées dans [8].**

C'est la méthode de combinaison par la moyenne qui obtient les meilleurs taux de reconnaissance.

[3] David Geronimo, Antonio Lopez, Daniel Ponsa, and Angel D. Sappa. Haar Wavelets and Edge Orientation Histograms for On-Board Pedestrain Detection. In *Proceedings of the 3rd Iberian Conference on Computer Vison and Pattern Recognition*, New-York, USA, June 2006.

[4] Duy Dinh Le and Shin'ichi Satoh. Feature Selection by AdaBoost for SVM-based Face Detection. *Forum on Information Technology*, pages 183–186, 2004.

[5] Laetitia Leyrit, Thierry Chateau, Christophe Tournayre, and Jean-Thierry Lapresté. Association of AdaBoost and Kernel Based Machine Learning Methods for Visual Pedestrian Recognition. In *IEEE Intelligent Vehicles Symposium (IV 2008)*, Eindhoven, Netherlands, 4 June 2008.

[6] David G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(20) :91–110, November 2004.

[7] Fabien Moutarde, Bogdan Stanciulescu, and Amaury Breheret. Real-time Visual Detection of Vehicles and Pedestrians with New Efficient AdaBoost Features. In *International Conference on Intelligent RObots and Systems (IROS) workshop*, 22 September 2008.

[8] S. Munder and D.M. Gavrila. An Experimental Study on Pedestrian Classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 28(11), November 2006.

[9] Pablo Negri, Xavier Clady, Shehzad Muhammad Hanif, and Lionel Prevost. A Cascade of Boosted Generative and Discriminative Classifiers for Vehicle Detection. *Eurasip Journal on Advances in Signal Processing*, 2008, 2008.

[10] Constantine Papageorgiou and Tomaso Poggio. A trainable system for object detection. *International Journal of Computer Vision*, 38(1) :15–33, 2000.

[11] Frédéric Suard, Alain Rakotomamonjy, Abdelaziz Bensrhair, and Alberto Broggi. Pedestrian detection using infrared images and histograms of oriented gradients. In *Proceedings of the IEEE Conference of Intelligent Vehicles (IV)*, pages 206–212, Tokyo, Japan, June 2006.

[12] Paul Viola and Michael Jones. Rapid object detection using a boosted cascade of simple features. In *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 1, pages 511–518, 2001.

[13] Paul Viola, Michael Jones, and Daniel Snow. Detecting Pedestrians Using Patterns of Motion and Appearance. In *Proceedings of the Ninth IEEE International Conference on Computer Vision (ICCV)*, 13 October 2003.