Searching Space Reduction for On-Board Pedestrian Detection

David Gerónimo*, Angel D. Sappa*, José M. Alvarez* and Antonio López**

*Computer Vision Center and  **Computer Science Department
Universitat Autònoma de Barcelona (UAB), 08193, Bellaterra, Barcelona, Spain
E-mail: {dgeronimo,asappa,jalvarez,antonio}@cvc.uab.es

Abstract This paper studies two different searching space reduction techniques for a vision-based pedestrian protection system. Both of them start from a successful candidates generation algorithm based on road plane fitting, and go one step further by trying to reduce the resulting set of candidates to be later classified. The first approach proposed exploits information from a color based road segmentation algorithm to filter out the candidates laying on the road. The second one uses 3D points to generate a probability map of vertical objects on the road, so candidates not laying on highlighted areas are discarded. Experimental results show that the stereo based proposal provides a great reduction (around 95%) in the number of candidates, whilst in the color based one the reduction is lower (15%).

Keywords: Pedestrian Detection, Stereo, Foreground Segmentation, Color.

1 Introduction

During the last decades, traffic safety has progressively become a major concern for motor companies, governments and the research community. Road security has increased along with the appearance of seat-belts, airbags, and other driver-centered mechanisms. However, during the last years a more sophisticated line of research pursues safety both inside and outside of the vehicle, the so-called Advanced Driver Assistance Systems (ADAS). These systems not only increase safety in the own vehicle, like adaptive cruise control (ACC) or lane departure warning (LDW), but also take care of other road users like people or vehicles, as pursued by pedestrian protection systems (PPSs).

PPSs aim at detecting the presence of both static and moving people in a specific area of interest around the host vehicle in order to warn the driver, perform braking actions or even deploy external airbags if a collision is unavoidable. Attending to the number of people involved in vehicle-to-pedestrian accidents, i.e., 150,000 injured and 7,000 killed each year in the European Union [1], it is clear that any improvement in this systems can save lots of lifes. However, the central task of a PPS (i.e., to detect the pedestrians) is far from being trivial. First, their variability in pose, clothes and sizes is considerable. Second, they must be identified in outdoor dynamic urban scenarios with cluttered background and illumination changes, in a wide range of distances and viewpoints. And finally, the requirements in terms of misdetections and computational cost are hard.

A common approach in PPSs consists in generating a set of regions of interest (ROIs), namely candidates, likely to contain a pedestrian and then label them using a classifier. These two steps are mapped to two modules of the standard PPS architecture proposed in [2]. In fact, they are the only two mandatory modules: foreground segmentation and object classification. This paper studies two searching space reduction approaches (filtering of candidates) that reduces the number of ROIs generated by the successful foreground segmentation module proposed in [3, 4]. The first algorithm relies on a color based road segmentation technique previously applied to non-urban scenes [5]. The second one exploits 3D data to localize vertical objects in the scene. In order to find examples of the object classification module the reader may refer to [6, 7, 4].

The remainder of the paper is as follows. Sect. 2 describes the approaches followed to perform candidates generation. Sect. 3 and Sect. 4 present the two methods proposed. Experimental results are shown in Sect. 5. Sect. 6 contains the final conclusions.
2 Related Research

The simplest candidates generation scheme is an exhaustive scan [8, 6]. The idea is to generate ROIs at all possible scales and in all possible locations of the image. For a typical $640 \times 480$ image, a dense scan consists of hundreds of thousands of ROIs, depending on the ROIs stride and their minimum and maximum size (Fig. 1(a)) Although this approach can be useful (and even necessary) in some applications like surveillance or database retrieval, the real-time requirements of ADAS make it inapplicable for PPS.

Recently, Sappa et al. [3] have proposed a method to restrict the scan just to the road plane, thus not only avoiding parts of the image where it is not necessary to search (e.g., the sky) but also discarding candidates according to their position and size in the image (e.g., small ROIs must be in the center of the image, i.e., where the furthest pedestrians are). The idea is to estimate the road plane position by making use of 3D points computed from a stereo pair. Once the plane is computed, 3D ROIs can be placed using a uniform sampling of the road and then projected to the image plane (Fig. 1(b)).

In this paper we propose to go one step further and present two techniques to discard ROIs from the list provided by the previous road fitting technique. The first one relies on a color-based road segmentation algorithm to discard candidates. The second is based on vertical projection of 3D points. They are explained in the following sections. Reducing the number of candidates to be classified is twofold advantageous. First, given the increasing complexity of classifiers, the less ROIs to be classified represents a lower computational cost required. Second, by discarding regions where pedestrians do not appear, the system avoids a potential false positive detection by the classifier.

3 Color Based

The first approach assumes that ROIs lying on the road do not contain pedestrians, whereas ROIs lying on areas with a small percentage of asphalt are likely to contain an object.

The road detection process is performed under the assumption that the bottom region of the image is road. In fact, the lowest row of the image corresponds to a distance of about four meters away from the camera placement. The algorithm combines the lighting–invariant properties of the color space introduced by Finlayson et al. [10] with a model–based classifier [9] (Fig. 2). Each RGB image is converted into an intensity $I(q)$ by projecting the log–chromaticity $\{\log(R/B), \log(B/G)\}$ values onto an invariant direction $\theta$. This direction is considered an intrinsic parameter of the camera and can be computed off-line [10]. Following, each pixel $q$ in $I$ is classified as road or background accordingly a road model $P(I(q)|Road)$, and a fixed threshold $\lambda$,

$$\begin{cases} q \text{ is Road,} & \text{if } P(I(q)|Road) \geq \lambda, \\ q \text{ is Background,} & \text{otherwise}. \end{cases}$$ (1)

Finally, only those road regions spatially connected to several seeds placed at the bottom part of the image are considered. The results presents some small holes that are filled by standard mathematical morphology.

The remaining is estimating $P(I(q)|Road)$ and adjusting $\lambda$. The road model is built using the nor-
Figure 2: Road detection algorithm.

ormalized histogram formed with the surrounding region of those seeds placed at the bottom part of $J$. In order to adjust $\lambda$ a set of images has been processed using different values of $\lambda \in [0...1]$ and their results have been compared against their ground truth using the measure proposed in [5]. The optimal value of $\lambda$ is the one which produces the maximum averaged accuracy.

Segmented road pixels are mapped to the estimated road plane [3] through the following equations system:

$$
\begin{align*}
  u &= u_0 + f \frac{Y_C}{Z_C} \\
  v &= v_0 + f \frac{X_C}{Z_C} \\
  n_x X_C + n_y Y_C + n_z Z_C &= 1,
\end{align*}
$$

where the known parameters are the estimated plane normals $n_x, n_y, n_z$, the image coordinates $u, v$ of the mask and the focal length of the camera in pixels $f$, whereas $X_C, Y_C, Z_C$ are the unknown mapped coordinates in the world. Continuity obtained in the projected region is achieved by linear interpolation.

After this, a given ROI is discarded if it lays on top of the projected mask, i.e., if the ratio between road and foreground exceeds a given threshold $\tau$. Experimentally, $\tau$ is set to 0.75. Fig. 3(a) shows a sketch of the mapping and filtering process.

4 Stereo Based

The second technique exploits stereo information. The idea is to project computed 3D points of the scene into the estimated road plane so to create a probability map of the regions with vertical objects in the scene, and then select the ROIs according to such probability values.

The method starts by rotating the camera coordinate system $(X_C, Y_C, Z_C)$ using the pitch angle $\theta$ computed according to [3], in order to make the estimated road plane parallel to the new XZ-axis. Roll angle is set to zero by an offline calibration process while yaw angle is zero given that $z$ axes of camera and world are assumed to be contained in the same plane. Hence, the coordinates of a given point $p(x, y, z)$, referred to the new coordinate system are computed as follows:

$$
\begin{align*}
  p_{xR} &= p_x \\
  p_{yR} &= \cos(\Theta)p_y - \sin(\Theta)p_z \\
  p_{zR} &= \sin(\Theta)p_y + \cos(\Theta)p_z.
\end{align*}
$$

Then, rotated points located over the road\footnote{Set of points placed in a band from 0 to $2m$ over the road plane, assuming that this is the maximum height of a pedestrian} are projected on a uniform grid $G_P$ in the plane, where each cell has a size of $\sigma \times \sigma$. A given point $p(x_R, y_R, z_R)$ votes into the cell $(i, j)$, where $i = \lfloor x_R/\sigma \rfloor$ and $j = \lfloor z_R/\sigma \rfloor$. Experimentally, $\sigma$ has been set to 0.2 meters. The resulting map $G_P$ is shown in Fig. 4(a). As can be seen, cells far away from the sensor tend to have few projected points. This is caused by two factors. First, the number of projected points decreases directly with the distance, as a result of perspective projection (Fig. 5). Second, the uncertainty of stereo reconstruction also increases with distance. Thus the points of an ideal vertical and planar object would spread wider into $G_P$ as the distance of these points increases. In order to amend this problem, the number of points projected into each cell in $G_P$ are reweighted and redistributed. The reweighting function is

$$
G_{RW}(i, j) = j\sigma G_P(i, j),
$$

where $j\sigma$ corresponds to the real depth of the cell. The redistribution function consists in propagating the value of $G_{RW}$ to its neighbours as follows:

$$
G(i, j) = \sum_{s=-\eta/2}^{i+\eta/2} \sum_{t=-\eta/2}^{j+\eta/2} G_{RW}(s, t),
$$

where $\eta$ is the stereo uncertainty at a given depth (in cells): $\eta = uncertainty/\sigma$. Uncertainty is computed as a function of disparity values:

$$
uncertainty = f \cdot baseline \cdot \frac{\mu}{\text{disparity}^2},
$$

where $baseline$ is the baseline of the stereo pair in meters, $f$ is the focal length in pixels and $\mu$ is the correlation accuracy of the stereo. The resulting (i.e.,
Figure 3: The two algorithms proposed. (a) Color based and (b) stereo based candidate generation proposals.

after reweighting and redistribution processes) map $G$ is illustrated in Fig. 4(b). Finally, a given ROI is discarded if the corresponding cell contains less than $\zeta$ points; an illustration is presented in Fig. 3(b).

Figure 4: Probability map of vertical objects on the road plane. (a) Raw projection $G_P$. (b) Reweighted and redistributed vertical projection map of the frame 3D points. (c) Original frame.

Figure 5: The number of points in a constant range $h$ projected on the road plane decreases according to the depth $z$ (darker cells correspond to fewer accumulated points.

5 Experimental Results

In order to evaluate the performance of both proposals 30 representative (i.e., typical complex urban scenes containing pedestrians and other vertical objects) frames have been tested.

The road scan is made by placing a ROI every $10\,cm$ in a range from $-10\,m$ to $10\,m$ on the sides of the camera ($X$–axis) and every $50\,cm$ from 0 to $50\,m$ in front of the camera ($Z$–axis). For a given scan position in the plane two ROIs are tested in order to be robust to slight misalignments of the estimated road plane: one ROI over the road and another one moved up $0.2\,m$. In a real detector this sampling should also include additional ROIs to cover multiple pedestrian sizes for the same position, e.g., from $1.5$ to $2.2\,m$ at intervals of $10\,cm$, but we have omitted this to make the figures more clear.

Fig. 6 illustrates the performance of the two algorithms. Starting from the list of ROIs provided by the road fitting, which provides around 30,000, the color based road segmentation discards around $15\% \ (\sigma = 14.3\%)$ of them. The algorithm filters out quite well the ROIs that just contain asphalt, as can be clearly appreciated in the last two rows. However, two disadvantages can be pointed out. First, the road segmentation results are not perfect as a result of the big complexity of urban roads compared to highways: zebra crossings and other painted signals are special cases in the road color, there can exist not uniform and eroded paintings and different asphalt patches along the road. This can make the ROI filtering process lose a potential pedestrian, as occurs in the fourth example of the figure. On the contrary,
the stereo algorithm seems more robust as it does not lose any pedestrian in any of the test images. In addition, the filtering discards around 95% ($\sigma = 3.7\%$) of the ROIs, which translates in a very significant speed gain when the ROIs are processed by the classifier. Regarding the computational cost, it can be said that both algorithms are of similar complexity $O(n)$.

6 Conclusions

We have presented two proposals to reduce the number of candidates to be classified in a PPS: a color based road segmentation and a stereo based vertical map of vertical objects. The algorithms start from a road fitting technique [3] that places 30,000 candidate ROIs, and achieve a mean reduction of 15% and 95%, respectively, which lets us think that the proposed algorithms, specially the stereo based one, can be essential to speed up the whole system.

However, since experience in ADAS makes us be conservative in every step of the system, i.e., the required robustness of ADAS must be so high that a false negative (wrongly discarding a pedestrian ROI) cannot be tolerated. Hence, regardless of the impressive results of the stereo based algorithm, we believe that in the future the two approaches will need to be combined in order to achieve robustness and significant searching space reduction.

Acknowledgements

This work was supported by Spanish Ministry of Education and Science (MEC) under Project TRA2007-62526/AUT, Research Program Consolider Ingenio 2010 Project CSD2007-00018 and MEC grant BES-2005-8864 (first author).

References


Figure 6: Some example results of the two proposed algorithms. Number of ROIs is noted in the bottom-left corner of the image.