Model Features and Horizon Line Estimation for Pedestrian Detection in Advanced Driver Assistance Systems

David Gerónimo, Antonio López, Angel D. Sappa and Daniel Ponsa

Computer Vision Center
Edifici O Campus UAB
08193 Bellaterra, Barcelona, Spain
E-mail: david.geronimo@cvc.uab.es

Abstract In this paper we propose a novel stereo–based technique to estimate the horizon line, which corresponds to a pitch angle, to determine a set of candidate windows to be classified. Then, we study the best feature sets (Haar wavelets, Differential Invariants and Local Jet, Edge Orientation Histograms and Structure Tensor Orientation) that will define a pedestrian model to be used by an AdaBoost classifier. The experiments, which have used our own database, show that combining Simple Haar sets together with Edge Orientation Histograms provide the best detector performance. Finally, we show some snapshots of the system in real urban scenarios.

Keywords: Pedestrian Detection, Advanced Driver Assistance Systems, learning, AdaBoost, horizon line, pitch estimation.

1 Introduction

In the last decade, research to improve traffic safety has moved towards intelligent onboard systems able to anticipate and prevent accidents, or at least, minimize their effects when unavoidable. Some examples of these systems, referred as Advanced Driver Assitance Systems (ADAS), are Adaptive Cruise Control, Lane Departure Warning, Headlights Automatic Control, etc. In this context we focus our work on Pedestrian Protection Systems (PPS).

Common difficulties for ADAS applications arise from dealing with a camera mounted on a mobile platform, so everything from backgrounds to obstacles are in movement in the image. In addition, objects intra–class variability is high as a result of this camera movement (i.e., distance, size and view angle variations) and working in outdoor scenarios (i.e., illumination, temperature conditions, etc.). Moreover, real–time requirements usually go from 5Hz to 25Hz. In this context, PPS must deal with non–rigid aspect–changing objects, and since pedestrian detection makes sense mainly in urban areas, the degree of complexity is much more high than other scenarios like highways.

The high relevance of pedestrian detection has attracted the attention of many researchers. For instance, *Gavrila et al.* [2] present the PROTECTOR System. First, a depth map coming from a stereo pair is multiplexed in 2D pedestrian–sized regions to be classified by a hierarchical template matching (Chamfer System), followed by a texture classification Neural Network technique. Finally, tracking is also incorporated to filter out spurious detections. In [8], *Shashua et al.* first use a recognition–by–components based on SIFT descriptor, and then apply a multiframe approval process that takes advantage of temporal coherence of the windows.

The work presented in this paper addresses the pedestrian detection by first using a horizon line estimation technique to determine a set of windows in the ground likely to contain a pedestrian, and then using AdaBoost to classify these windows as pedestrians or non–pedestrians. The outline of this paper is as follows. Section 2 introduces the acquisition system that is used to construct our pedestrian database. In section 3 we explain our stereo–based pitch estimation technique, which helps us to determine the candidate windows to be classified. Section 4 presents the features used to construct a pedestrian model, used to classify the mentioned windows. Section 5 shows some experimental results of the complete system, which prove that the technique works well in complex urban scenarios. Finally, Section 6 summarizes the conclusions and proposed future work.

2 Acquisition system

In order to make the experiments we have mounted a PtGrey Bumblebee stereo pair on two different sedans (Fig. 1) (thus, camera position can vary in different acquisition sessions). For the stereo process involved in the horizon line technique, RGB images from the two cameras are used; for the classification purposes, we just use the right image using grayscale. Some of the relevant parameters of our system are: effective field of view of one camera is 43.07° (horizontal) and 32.97° (vertical), the stereo baseline is $12cm$, and the resolution is 640×480 (no downsampling applied).

Our database consists in $1,000$ positives, i.e., pedestrians, and $5,000$ negatives, which correspond to windows in the ground area and containing non–pedestrians, i.e., vehicles, trees, benchs, etc..

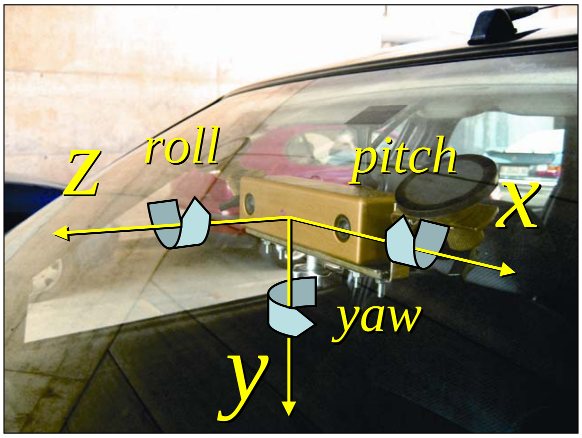


Figure 1: Acquisition system with the coordinates. Roll and yaw can be assumed to be constant, but pitch orientation varies when the vehicle brakes or accelarates, when road slope changes and even as a result of asphalt rugosity.

3 Horizon Line Estimation

The first processing in our system consists in determining the windows to be classified. The most intuitive approach is to make an exhaustive scan of the input image with windows of all the possible positions and sizes, as used in [6]. Obviously this method is too time–consuming for our purposes. A more sofisticated approach, often used in ADAS, consists in fixing the pitch angle and then adjusting the window sizes according to the real size that pedestrians would have at certain distances [5]. Although it works quite will in highways, in urban scenarios we will have problems because the road slope is not constant, thus the pitch angle and height of the camera varies. Our proposal is to dynamically estimate these parameters [7].

3D data acquired from the stereo pair (Fig. 2a) is projected and downsampled to cells in the $Y - Z$ plane (Fig. 2b). Then, noise is filtered out by discarding cells without a minimum number of points. The first 3 cells in each column are then selected. The resulting set of points is noted as ζ (Fig. 2c). Next, a plane solution (a, b, c), consisting in 3 random points, is calculated by Least Squares Fitting in an iterative manner (using RANSAC). In this way, the solution with highest number of inliers in a $\pm 5cm$ band is chosen as the correct one, and the road orientation (Fig. 2d), thus the horizon line in the image, is computed.

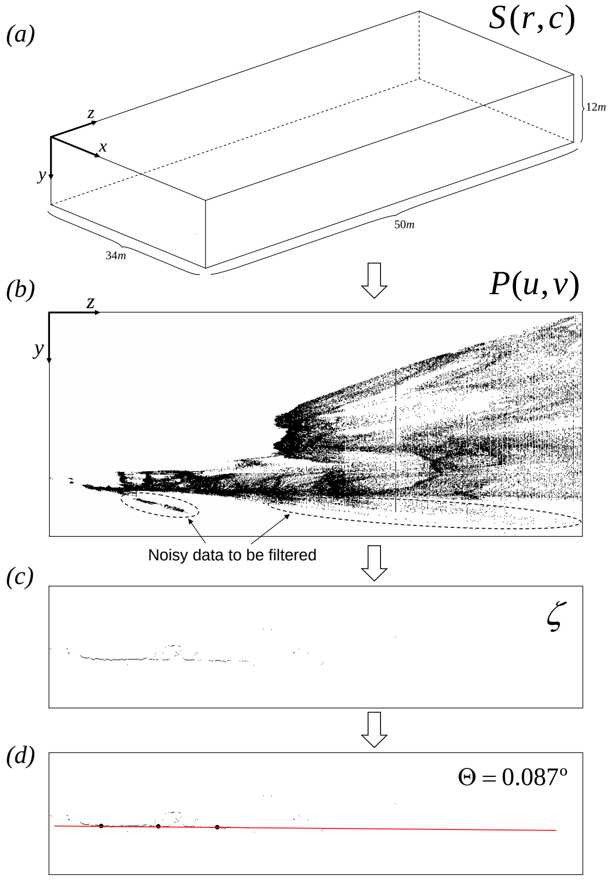


Figure 2: Pitch estimation technique [7].

4 Model features

Once the set of candidate windows is determined, they can be classified as pedestrians or non–pedestrians. In our case, we propose to use AdaBoost learning algorithm to select relevant features to construct a pedestrian model. Then, we also use AdaBoost to classify each candidate window by using these selected features.

In our work, we have focused on the analysis of the features that provide the best performance rates at pedestrian classification. In order to make the experiments, 700 positive and $4,000$ negative samples are selected as the training set, and passed to AdaBoost in order to use the 100 most discriminant features as the weak rules of the classifier. Then, the remaining 300 positives and $1,000$ negatives are used as the testing set.

4.1 Haar wavelets

A Haar wavelet feature is defined as a region of a given size and position where a filter is applied. A filter is computed as the difference of illumination between two areas, one represented as black and other as white. The first set of Haar filters (*(a–c)* in Fig. 3) are proposed by Oren et al. and used to perform pedestrian detection [6] (Basic Haar set). Later, Viola and Jones [10] added two more filters (*(d) and (e)*)to this set in order to detect faces (Simple Haar set). Finally, Lienhart and Maydt [4] extended the set to also perform face detection (Extended Haar set) by adding filters *(f),(g),(h)*, and the $45°$ rotation of the whole set.

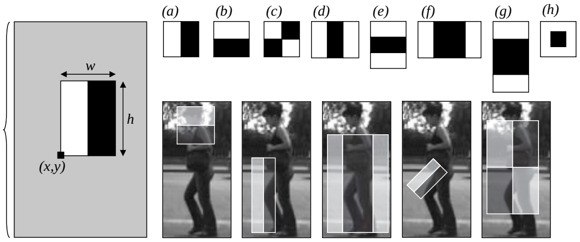


Figure 3: *Haar filters*. *Left*: xample of filter in a candidate window. *Top–right*: Basic forms of the Extended Haar set. *Bottom–right*: Examples of filters that give high response in regions containing pedestrians.

4.2 Differential Invariants and Local Jet (DI&LJ)

Next, we take advantage of the Simple Haar set features both to use them as the local jet until order 2 and to calculate several differential invariants:

$$\text{GradientMagnitude} = \sqrt{I_x^2 + I_y^2}$$

$$\text{Laplacian} = I_{xx} + I_{yy}$$

$$\text{Isophote Curvature} = \frac{2I_xI_yI_{xy}-I_x^2Iyy-I_y^2Ixx}{(I_x^2+I_y^2)^{3/2}}$$

$$\text{Flowline Curvature} = \frac{I_xI_y(I_{yy}-I_{xx})+I_{xy}(I_x^2-I_y^2)}{(I_x^2+I_y^2)^{3/2}} \quad . \tag{1}$$

The features needed to compute the local jet and the invariants are already calculated. For instance, I_x and I_y are the first order partial derivative in x and y axes, respectively, so they are equivalent to the two first filters of the Simple Haar set. The same happens with the second derivative I_{xx}, I_{yy} and the cross I_{xy}. In their original context [9], the features would be computed in the gaussian space–scale, whilst in our approach the scale of the feature is given by the size of the filters.

4.3 Edge Orientation Histograms

In 2004, Levi and Weiss present a face detector based on orientation histograms [3]. These features seem also to be interesting for our work, since pedestrians usually present strong edges in zones like the legs or the trunk. In addition, these features not only maintain invariance to global illumination changes but also are able to extract information usually difficult to capture by Haar filters.

The feature is defined in a given region with a specific size and position. First, a Sobel mask extracts the edges in x and y to compute the edge orientation of each pixel. Then, image pixels in a region are stored in bins according to their orientation. Therefore, a pixel in bin $k_n{\in}K$, where $\#K = 4$ in our case, contains its gradient magnitude if its orientation is inside k_n's range, otherwise is null. In our case, instead of giving all the gradient magnitude value to a concrete bin, the value is bilinearly interpolated between the neighbouring bin centers as [1] proposed with other similar features (HOGs) since the results are certainly improved. Finally, the feature value is defined as the relation between two different orientations for the given region (Fig. 4).

In the results plotted in Section 5, the number of bins, K, has been fixed to 4, since it gives better results than using $K = 9$ (as proposed for HOGs in [1]) for our database.

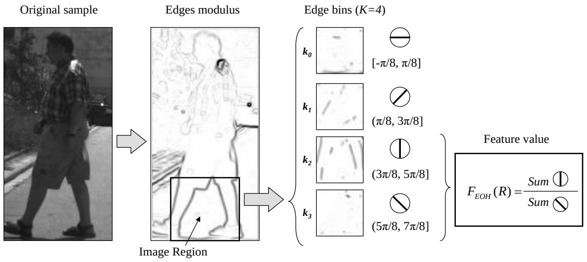


Figure 4: *EOH features*. The feature is defined as the relation between two orientations of a region. In this case, vertical orientations are dominant with respect to the diagonal orientations (k3), so the feature will have a high value.

4.4 Structure Tensor Orientation

In the search for a feature that filters out spureous pixel orientations and maintains the dominal orientations of neighborhoods, the structure tensor has also been tested.

The feature is computed as follows. Let us note the image as I, which is convoluted with a gaussian G_{σ_D} (where σ_D is the *differentiation scale*) to form I_{σ_D}. Then, the first order derivative in x and y axes result in $I_{\sigma_{D,x}}$ and $I_{\sigma_{D,y}}$. Finally, the structure tensor S is constructed with the elements $s_{11} = (I_{\sigma_D,x})^2_{\sigma_I}$, $s_{22} = (I_{\sigma_D,y})^2_{\sigma_I}$ and $s_{12} = (I_{\sigma_D,x})_{\sigma_I}(I_{\sigma_D,y})_{\sigma_I}$, where σ_I corresponds to the *integration scale* of the gaussian that is applied to $(I_{\sigma_D,x})^2$, $(I_{\sigma_D,y})^2$ and $I_{\sigma_D x} I_{\sigma_D y}$:

$$\mathbf{S} = \begin{pmatrix} s_{11} & s_{12} \\ s_{12} & s_{22} \end{pmatrix} \quad . \tag{2}$$

Supposing that S is not a diagonal matrix, the steps are the following:

$$\begin{aligned}
&\Delta_1 = s_{11} - s_{22} \\
&\Delta_2 = s_{12} \\
&\lambda_\Delta = (\Delta_1)^2 + (\Delta_2)^2 \\
&\Delta_3 = (\Delta_1 + \sqrt{\lambda_\Delta})/\Delta_2 \\
&\Delta_4 = \sqrt{1 + (\Delta_3^2)} \\
&u_1^x = \Delta_3/\Delta_4 \\
&u_1^y = 1/\Delta_4 \quad .
\end{aligned} \tag{3}$$

Finally, the resulting orientation is noted as $\theta =$arctan$\left(\frac{u_1^y}{u_1^x}\right)$. The feature is defined as the relation between two orientations for a given area, the same approach as Edge Orientation Histograms. In the results plotted in Section 5, we have used $\sigma_D = 1$ and $\sigma_I = 10$, since they provide the best results for this feature.

4.5 Combination

Finally, we have tested the combination of several sets: Simple Haar together with EOH, Extended Haar with EOH and DI&LJ with EOH. In this case, the 100 features of the classifier are a combination of the best features from the two sets.

5 Experimental Results

First, we are going to analyze the performance of the classification itself, to then show some snapshots of the whole detection system (consisting of both the pitch estimation technique and the classifier).

Fig. 5 illustrates ROC curves for classifiers based in 100 AdaBoost weak hypothesis; i.e., 100 discriminatory features, learnt for each feature set. The curves correspond to the average of four experiments, each one with a random training and testing set[1].

By taking a $FPR = 0.01$, which represents one false positive for one hundred tested samples, in Fig. 5a it can be seen that the Extended Haar

[1] Note that Area Under the Curve (AUC) is computed by using the $[0, 0.1]$ interval of False Positive Rate (FPR) and not the whole ROC space for two reasons: first, we think the important part of the curve is at low FPRs, the rest is useless for an ADAS application; and second, the curves tend to stabilize passed the 0.1, so the resulting area differences beyond it would just hide the results in the intersting range.

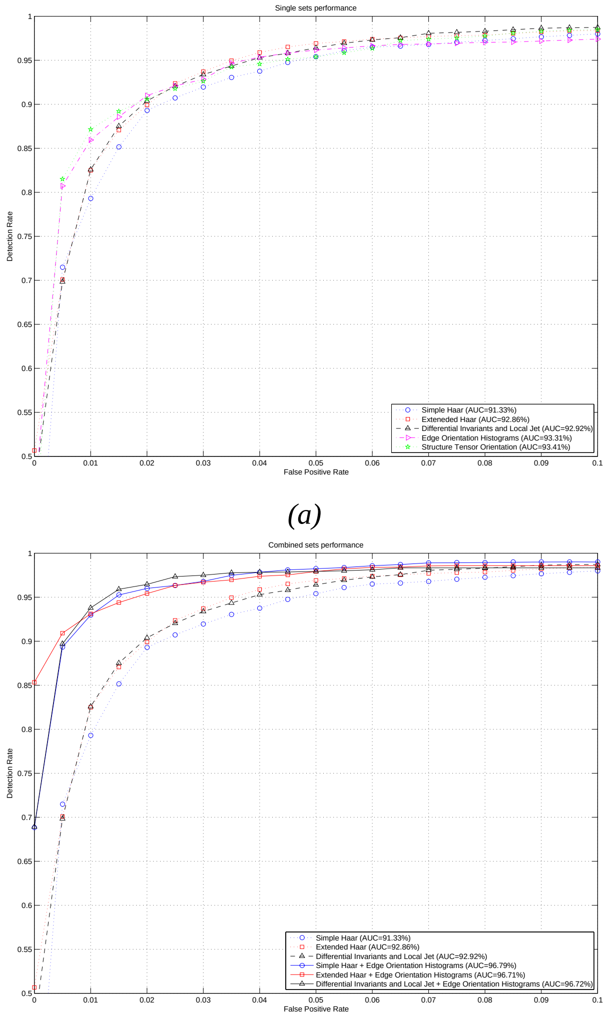
set improves the performance of the Simple set in 4%. Differential Invariants and Local Jet provide the same improvement. EOH also improve the rates in 7% and 4% compared to Simple and Extended Haar sets respectively. The most intuitive idea here is that orientation–based features generalize better, since they use more concrete information than just graylevel differences. Finally, although providing an improvement of about 1.5% over EOH, there is not a significant improvement in all the plot when using STO. Then, as can be seen in Fig. 5b, the combination of any Haar set and EOH provides a great performance improvement. For instance, the combination of Simple Haar and EOH represents an improvement in correct detection of 15% over the Simple Haar alone, and about 11% for Extended Haar and EOH compared to the Extended Haar alone. The same happens with Differential Invariants and Local Jet. However, not all the combinations provide the same improvement in terms of detection rates. It can be stated the combined sets' cuve is not the sum of the single sets' curves alone. Contrary, the resulting curve seems to reflexct the degree of complementarity between the two combined sets, i.e., if two given sets extract different information from the image, their combination would suppose a higher improvement than two sets that extract similar information when used alone.

Finally, in Fig. 6 we illustrate the results of the detection system, tested on different urban environments, providing good performance in different illumination conditions and ground profiles. In this case, we have used the combined Simple Haar and EOH feature sets in the classification stage, and of course, the horizon has been dynamically estimated by the exposed technique. As can be seen, the overall performance of the detector is satisfactory: pedestrians are correctly detected even without any tracking process involved. Notice that no postprocessing was applied to the detections, so redundant detection windows are present for single pedestrians. Spurious false positive detections will be removed in future works by adding cascades to the classifier and attaching a tracking module.

6 Conclusions and Future Work

We have presented a system for pedestrian detection in urban environments. The study of the feature sets performance made relevant that combining sets with complementary information improve the performance of the single sets alone when facing this problem. In our case, the Simple Haar set together with EOH were used in the final detection system. In addition, a stereo-based pitch-estimation technique allows us to determine the searching area as well as a sampling grid from where meaningful windows for applying the pedestrian classifier are obtained, thus, avoiding an exhaustive search on the whole image.

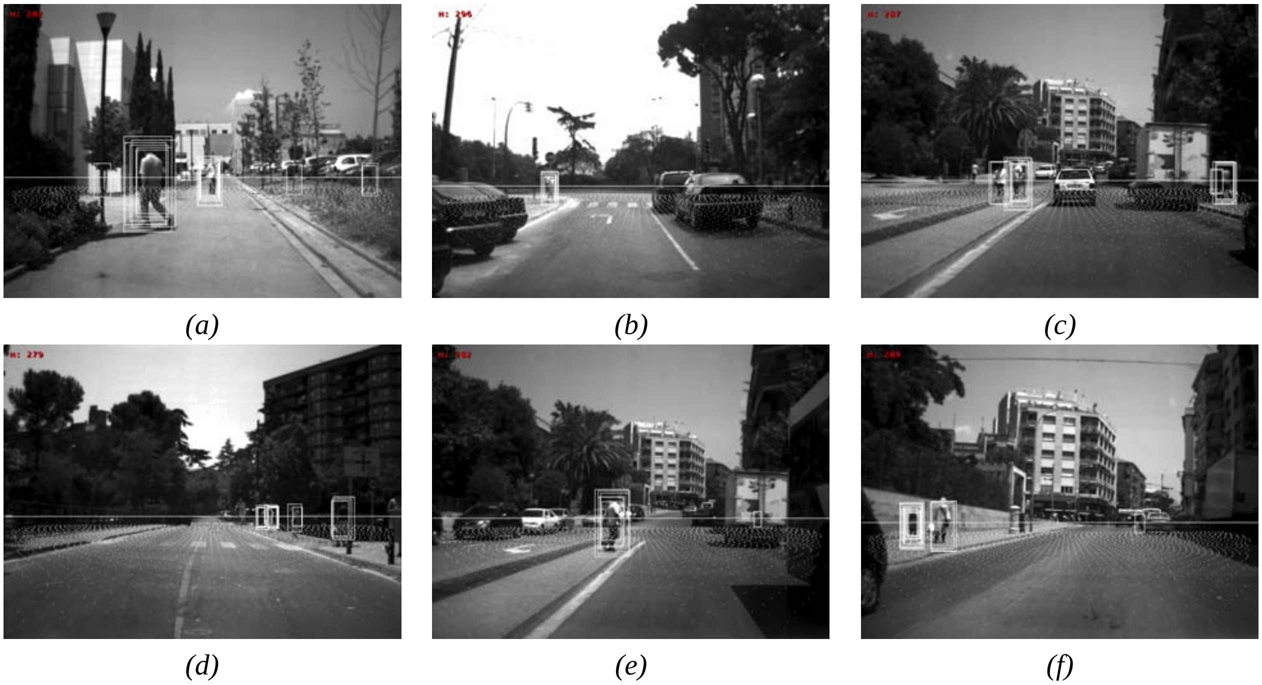
As future work, we plan to improve the system by using free–space analysis, implementing methods to remove redundant windows and attach a tracking module to filter our spurious false positive detections.



(a)

(b)

Figure 5: Clasifier performance. ROC curves for *(a)* single and *(b)* combined feature sets.



(a) (b) (c)

(d) (e) (f)

Figure 6: Pedestrian detection system results. The estimated horizon is used to adjust the candidate windows botton–left–corners (white dots). Positive detections are marked as white boxes.

References

[1] N. Dalal and B. Triggs, ”Histograms of Oriented Gradients for Human Detection”, *IEEE Proc. Conf. Computer Vision and Pattern Recognition*, 2:886–893, 2005.

[2] D.M. Gavrila, J. Giebel and S. Munder, ”Pedestrian Detection: The PROTECTOR+ System”, *IEEE Proc. Intelligent Vehicle Symposium*, 2004.

[3] K. Levi and Y. Weiss, ”Learning object detection from a small number of examples: the importance of good features”, *IEEE Proc. Intl. Conf. on Computer Vision and Pattern Recognition*, 53–60, 2004.

[4] R. Lienhart and J. Maydt, ”An extended set of Haar–like features for rapid object detection”, *IEEE Proc Intl. Conf. on Image Processing*, 2002.

[5] D. Ponsa, A. López, F. Lumbreras, J. Serrat and T. Graf, ”3D vehicle sensor based on monocular vision”, *Proc. Intelligent Transportation Systems*, 2005.

[6] M. Oren, C. Papageorgious, P. Sinha, E. Osuna and T.Poggio, ”Pedetrian detection using wavelet templates”’, *IEEE Proc. Conf. Computer Vision and Pattern Recognition* 193–199, 1997.

[7] A.D. Sappa, D. Gerónimo, F. Dornaika and A. López. ”On–board camera extrinsic parameter estimation”, IEE Electronic Letters, 42(13), pp. 645747. 2006.

[8] A. Shashua, Y. Gdalyahu and G. Hayun, ”Pedestrian Detection for Driving Assistance Systems: Singleframe classification and System Level Performance”’, *IEEE Proc. Intelligent Vehicle Symposium*, 2004.

[9] J. Sporring, M. Nielsen, L. Florack and P. Johansen (Eds), *Gaussian Scale–Space Theory*, Kluwer Academic Publishers, 1997.

[10] P. Viola and M. Jones, ”Rapid Object Detection using a Boosted Cascade of Simple Features”, *IEEE Proc. Intl. Conf. on Computer Vision and Pattern Recognition*, 2001.