Meta-classifier for Pedestrian Attitude Recognition

Raluca Borca-Mureșan, Sergiu Nedevschi

Abstract

This paper presents an innovative pedestrian detection algorithm with applications to driving assistance systems. The novelty of the approach resides in the construction of an expert module for pedestrian recognition based on a meta-classification scheme applied on different pedestrian attitudes. The designed module is part of a real-time stereo based driving assistance system. The proposed meta-classification scheme learns the discriminant features of a partitioned pedestrian space. The complex pedestrian object is decomposed into different attitudes like pedestrian standing, running and for each attitude a classifier is trained. Our experiments show that the proposed meta-classification scheme outperforms a single classifier trained on the whole un-partitioned object space. For classification a probabilistic approach based on Bayesian Networks was used. Two types of features extracted from the image have been involved in the training process: magnitude of first order partial derivatives computed in four directions and histograms of gradient orientations (HOG).

1 Introduction

The detection of humans in still images and especially in traffic scenarios is an important problem for artificial vision and pattern recognition. A robust solution to this problem should have various applications to autonomous driving systems, video surveillance, image retrieval.

In general, the goal of pedestrian detection is to determine the presence of humans in images and videos and return information about their position. The problem of detecting pedestrians has a high degree of complexity because of the large intra-class variability, as pedestrians are highly deformable objects whose appearance depends on numerous factors like: pose, orientation, shape, attitude, occlusions, imaging conditions, background as shown in Figure 1.

Our aim is to build a reliable pedestrian detection system that works in highly cluttered urban environment. In our approach pedestrian hypotheses are generated by a stereo module [1] that uses 3D and 2D information for constructing the pedestrian object assumption. The existing stereo-based driving assistance system [1] performs several preprocessing operations, among which pedestrian hypotheses generation. The 3D pedestrian hypothesis is projected onto the image plane generating a 2D pedestrian assumption represented by a 2D image window. This 2D image window is scaled to a fixed size and sent as input to our meta-classification module whose purpose is to evaluate each pedestrian hypothesis and to provide a confidence measure.

A first step in the design of the meta-classifier is the choice of a robust feature set that allows the humanoid shape to be distinguished precisely in a cluttered background and under difficult illumination conditions. The proposed approach exploits descriptors based on histogram of gradient orientations and descriptors based on absolute value of first order partial derivatives computed in four different directions, as described in Figure 2.

As mentioned previously, pedestrians are complex objects with large intra-class variability. We have partitioned the pedestrian object space into categories based on different poses that people can have: running, walking, standing as shown in Figure 2. The division of attitudes reduced the complexity of the humanoid object set.

The contribution of this paper resides in the development of a pedestrian meta-classification scheme based on a partitioned object space. We have trained classifiers for different categories of pedestrian attitudes. We show that the obtained meta-classifier outperforms previous approaches that use the whole un-partitioned pedestrian space.

The structure of this paper is as follows: in section 2 we review the previous work on the pedestrian detection problem. In section 3 we give details on our meta-classification scheme. Section 4 presents the dataset and the methodol-
2 Related work

Over the last decade the field of pedestrian detection has been very active and had a real break-through. Most approaches use computer vision in the visible spectrum.

A hierarchical system is developed by [2], [3] that use Chamfer distance to compare edge maps of pedestrian templates to edges found in images. A variant of AdaBoost is used by [4] to build a pedestrian classifier that relies on spatial and temporal rectangle filters computed efficiently based on the integral image representation.

Methods based on local feature descriptors were also proved to give good results. A local feature detection scheme is used by [5] to generate a set of pedestrian hypotheses. From foreground pedestrian masks they generate for each hypotheses some segmentation masks that are further combined with Chamfer matching in a top-down verification step. [6] make use of Histogram of Oriented Gradient (HOG) descriptors and Support Vector Machines for building a pedestrian classifier. By tuning all the parameters of their HOG features, they compare the use of a variety of feature configurations to find the best configuration for pedestrian detection.

The problem of pedestrian classification with different features and classifiers has been studied by [7]. They find that local receptive fields can do a better job in representing pedestrians and also SVMs and AdaBoost classifiers outperform the other classifiers tested.

Other methods are based on stereo data and make use of depth information. Most systems rely on the disparity map and make some kind of segmentation of this map to detect objects [8], [9] or use a v-disparity approach [10], [11]. Infrared technology has also been explored in the field of pedestrian detection [12], [13]. A night vision system has been build by [13]. A stereo rig with infrared cameras is used to estimate the distance in the scene. The ego-motion is used to get the relative movement in the scene. Pedestrians are located in the image by simple mean intensity threshold, which finds the areas that are bright (hot) in the images.

3 Overview of the method

The aim of this research is to propose a 2D image based pedestrian recognition expert. This expert is a component of a probabilistic pedestrian detection framework exploiting 2D, 3D and motion information. The connection between the framework and the expert is bidirectional, the first providing pedestrian hypothesis with the associated 2D image window, and the second one returning the confidence score. The probabilistic pedestrian detection framework is described in [1] and [14].

The pedestrian recognition expert should provide accurate results for any pedestrian attitude. By consequence an accurate implementation based on a single classifier was not yet achieved. We propose a solution based on a meta-classifier that starting from the identification of main pedestrian attitudes generates a hierarchy of accurate classifiers.

The steps we performed for constructing the meta-classification scheme are: build a hierarchy of pedestrian attitudes; for each image from the hierarchy of attitudes, relevant features are extracted and selected; train a classifier for every defined pedestrian attitude. Perform classification and evaluate the accuracy of the detection. The flow of our pedestrian recognition algorithm is presented in Figure 3.

3.1 Feature description

In our work we use two sets of features: magnitude of first order partial derivatives computed in four directions and histogram of gradient orientations.

The magnitude of the first order partial derivatives is used because the sign of the magnitude of the first order partial derivative is uninformative due to varying clothing and background colors. In order to decrease the influence of
Gradient Orientation and Magnitude Computation

Weighted Histogram of Gradient Orientations

Normalization

Classifier Training

Method 1: use HOG

Method 2: use magnitude of first order partial derivatives

Figure 3. Flow of the pedestrian detection algorithm for two categories: pedestrians running and pedestrians standing.

small spatial shifts in the detection window, we locally average the first order partial derivatives in each direction by convolving their responses with a 2D averaging filter. For each image $I(x)$ in the training set we perform the following operation:

$$GL_d(x) = |I(x) * G_d| * B$$  \hspace{1cm} (1)

where $\ast$ denotes the convolution, $G_d$ is the derivative kernel ($[-1,0,1]$ or $[-1,0,1]^T$) used for obtaining the derivatives in direction $d \in D$. $B$ is a 2D averaging filter and $GL_d$ is the result image that captures the amount of first order partial derivative information at every pixel location in direction $d$. The set $D = \{0^\circ, 45^\circ, 90^\circ, 135^\circ\}$ represents the directions in which we have computed the partial derivatives. Figure 2 shows the resulting features for the four directions. Each feature is characterized by direction, absolute value of the first order partial difference and position in the image. We divide each image $GL_d$ into several blocks of different dimensions. The features in each block are normalized as described in section 3.2. For each block we perform a correlation feature selection (CFS) algorithm. The selected features from all the blocks form a vector of descriptors that represent the input to a classification algorithm.

Histogram of Oriented Gradients turned out to have good results for pedestrian classification [6], [15]. For each point of an image $I$ in the dataset we have computed the gradient magnitude, $M$ and orientation $\theta$ as follows:

$$GI_x = (I * B) * G_x$$  \hspace{1cm} (2)

$$M = \sqrt{(GI_x)^2 + (GI_y)^2}$$  \hspace{1cm} (3)

$$\theta = \arctan \frac{GI_y}{GI_x}$$  \hspace{1cm} (4)

where $B$ is a Gaussian smoothing kernel, $G_x = [-1,0,1]^T$, $G_y = [-1,0,1]$.

Next, we divide the image into non-overlapping rectangular cells of equal dimension as depicted in Figure 4. For each cell we compute a weighted histogram of gradient orientations. In each pixel location the orientation gives to the histogram a vote weighted by the gradient magnitude at the position of the respective pixel. The orientation bins are evenly spaced over $0^\circ$ - $360^\circ$.

The last step for HOG descriptors extraction is represented by normalization. The cells are grouped into overlapping blocks of various dimensions. Block overlapping ensures the fact that each scalar cell response contributes to several components of the final feature response. The normalization scheme we have used is described in section 3.2. The final vector of descriptors will contain all the components of the normalized cell responses from all of the blocks in the detection window.
3.2 Normalization scheme

As it turns out from the study of [6], effective local contrast normalization is essential for a good performance of pedestrian detectors based on gradient attributes. For both feature types we use L2-norm normalization algorithm applied on image blocks. Suppose that within a block we have a vector of k features denoted by \( f_d \). The value of a feature \( f_d(i) \) can either be the magnitude of the first order partial derivatives computed in four directions or a histogram value (for HOG attributes). The normalization equation is:

\[
 f_d(i) = \frac{f_d(i)}{\sqrt{\sum_{i=1}^{k} (f_d(i))^2 + \epsilon}}
\]

where \( \epsilon \) is a small constant.

3.3 Correlation-based feature selection for relevant magnitude of directional derivative attribute extraction

The purpose of correlation-based feature selection scheme is to eliminate redundant attributes as well as irrelevant ones. As presented by [16] and [17], CFS tries to select good feature subsets that contain attributes highly correlated with the class, yet uncorrelated with each other. As presented by [16] and [17], CFS tries to select good feature subsets that contain attributes highly correlated with the class, yet uncorrelated with each other. The correlation between two attributes \( A \) and \( B \) can be measured using the symmetric uncertainty [17]:

\[
 U(A, B) = 2 \times \frac{H(A) + H(B) - H(A, B)}{H(A) + H(B)}
\]

where \( H \) is the entropy function:

\[
 H(p_1, p_2, \ldots, p_n) = -p_1 \log p_1 - p_2 \log p_2 \ldots p_n \log p_n
\]

The symmetric uncertainty always lies between 0 and 1. Correlation-based feature selection determines the goodness of a set of attributes using:

\[
 \frac{\sum_j U(A_j, C)}{\sqrt{\sum_i \sum_j U(A_i, A_j)}}
\]

where \( C \) is the class attribute and the indices \( i \) and \( j \) range over all attributes in the set.

Attribute selection is normally done by searching the space of attribute subsets and evaluating each set. Search can be performed exhaustively, using a simple genetic algorithm, randomly, or by greedy hill-climbing with or without backtracking. In our experiments we have used the CFS implementation provided by [18].

3.4 Classification

Classification has a decisive role in any object detection system. We have implemented a Bayesian network learning scheme. Bayesian networks (or belief networks) have been used successfully for pedestrian classification in [19], [20].

Belief networks [21] are used to model the statistical dependencies among the component features. They take the topological form of a directed acyclic graph where each link is directional and there are no loops. They allow efficient and effective representation of the joint probability distribution over a set of random variables. In our algorithm each node (unit) represents one of the features described in section 3.1. For a given instance, the probability of each class value can be predicted using conditional probability tables that are given by the relative frequencies of the associated combinations of attribute values in the training data.

In order to build a learning algorithm for Bayesian networks two components must be defined: a function for evaluating a given network based on the data and a method for searching through the space of possible networks [17]. The quality of a given network is measured by the probability of the data given the network. The probability that the network accords to each instance is computed by adding the logarithms of the probabilities over all instances.

In our scheme, the nodes in the network are predetermined, one for each attribute (including the class). So, our Bayesian network \( U \) is a pair \( B = \langle G, \Theta \rangle \), where \( G \) is a directed acyclic graph whose vertices correspond to attributes in the training set, and whose edges represent direct dependencies between attributes. We model the set of attributes by the random variables \( A_1, \ldots, A_n \). The graph \( G \) encodes independence assumptions: each variable \( A_i \) is independent of its nondescendants given its parents in \( G \). \( \Theta \) represents the set of parameters that quantifies the network. It contains a parameter \( \theta_{A_i|\Pi_{A_i}} = P_B(a_i|\Pi_{A_i}) \) for each possible value \( a_i \) of \( A_i \), and \( \Pi_{A_i} \) of \( A_i \), where \( \Pi_{A_i} \) denotes the set of parents of \( A_i \) in \( G \). The Bayesian network \( B \) defines a unique joint probability distribution over \( U \) given by:

\[
 P_B(A_1, \ldots, A_n) = \prod_{i=1}^{n} P_B(A_i|\Pi_{A_i}) = \prod_{i=1}^{n} \theta_{A_i|\Pi_{A_i}}
\]

So, being given a set of attributes \( a_1, \ldots, a_n \) and an attribute describing the class, \( c \), the classifier based on \( B \) returns the label \( c \) that maximizes the posterior probability \( P_B(c|a_1, \ldots, a_n) \).

We have experimented several algorithms for learning Bayesian networks. The K2 [17] algorithm starts with a given ordering of the attributes then it process each node in turn and greedily considers adding edges from previously processed nodes to the current one. In each step it adds the edge that maximizes the posterior probability score. Another learning algorithm we have tested is the tree aug-
mented Naïve Bayes (TAN) [17]. As the name implies, it takes the Naïve Bayes classifier and adds edges to it. The class attribute is the single parent of each node of a Naïve Bayes network: TAN considers adding a second parent to each node. If the class node and all corresponding edges are excluded from consideration, and assuming that there is exactly one node to which a second parent is not added, the resulting classifier has a tree structure rooted at the parentless node. For this restricted type of network there is an efficient algorithm for finding the set of edges that maximizes the network’s likelihood based on computing the network’s maximum weighted spanning tree. This algorithm is linear in the number of instances and quadratic in the number of attributes.

4 Dataset and methodology

Dataset. We have considered two classes of attitudes for pedestrians: standing and running (Figure 1). For each class we have used a training set and a testing set. We have collected our samples in the category standing from the MIT pedestrian database [22] and from the INRIA pedestrian database [23] that contain images of pedestrians in city scenes. We have collected pictures for the category ‘running pedestrians’ from several images taken from the Internet or from our personal photos and some from the INRIA [23] database. We build a set of 300 images of different running pedestrians. In all the pictures the pedestrians occupy the central position. We have applied a 4 small in-plane rotations with 5°, 10°, -5°and -10°to each image from both classes of attitudes, hence we have obtained a dataset of 1500 pictures of running pedestrians and 2500 pictures of standing pedestrians.

Methodology. For each category we have considered images of dimension 18x36 pixels. We have divided the datasets for the standing pedestrian attitude into 2100 training samples and 400 testing samples and the dataset for running pedestrians was split into 1100 train images and the 400 test pictures. The initial negative training set had 12000 images (also of dimension 18x36 pixels) sampled randomly from person-free training photos. An initial training is made and the obtained detector is tested on a larger set of negative images that were not in the initial negative training set. All the false positives are added to the training set and the process is repeated until we reach a good accuracy of the detection. Hence, we obtain a classifier for each category of the considered pedestrian attitudes i.e. running and standing. For each category we have trained two separated classifiers, one that uses HOG features and another that exploits the absolute value of first order partial derivatives computed in four directions. We have compared the results obtained using these two feature sets, HOG providing the best results.

We have grouped the classifiers trained for each category in a meta-classifier or a hierarchy of classifiers. The next step was the comparison of our method based on the partitioning of the pedestrian space into several classes, with a method that trains a classifier using the whole pedestrian set. In order to perform this comparison we build a third classifier that used as positive training set 1100 images from the category running and the 2100 train images from the category standing and the same set of negative samples as the classifiers build separately for each category.

Considering the two test sets (for pedestrians running and for pedestrians standing) we evaluated the performance of the detector trained using the whole object space and the performance of the meta-classifier. As detailed in section 5 the best results of pedestrian recognition are obtained when applying the meta-classification scheme.

5 Experimental results

In this part we will present some results of our system. For both categories, pedestrians running and pedestrians standing, we have computed the two feature sets: histograms of gradient orientations and magnitude of first order partial derivatives computed in four directions. We experimented numerous combinations of parameters for each attribute set:

1. For magnitude of first order partial derivatives computed in four directions the possible parameters are:
   - block size: 3x6, 3x3, 6x6, 6x12 pixels
   - search strategy for correlation-based feature selection: best first, genetic search, random search

2. For HoG the set of parameters is given by:
   - cell size: 3x3, 3x6, 6x6, 6x12 pixels.
   - number of bins in the histogram: 4, 8, 16
   - block size in number of cells: 3x3, 2x2

We retained the parameters that provide optimal results for our detection window of 18x36 pixels:

1. For magnitude of first order partial derivatives computed in four directions: a block size of 6x12 pixels and best first search resulted in a feature set of 316 attributes.

2. For HOG: a cell size of 3x6 pixels with a histogram having 8 bins, and a block size of 3x3 cells resulted in a feature set of 288 attributes.

Concerning the parameters of the Bayesian Network we have used TAN as learning algorithm. It determines the maximum weight spanning tree and returns a Naïve Bayes network augmented with a tree (see section 3.4).
We have evaluated our method with 400 positive samples of running pedestrians, 400 positive images of standing pedestrians and 10000 non-pedestrians.

<table>
<thead>
<tr>
<th></th>
<th>TP</th>
<th>FP</th>
<th>Prec</th>
<th>Rec</th>
<th>ROC-A</th>
</tr>
</thead>
<tbody>
<tr>
<td>PedRun</td>
<td>0.901</td>
<td>0.05</td>
<td>0.92</td>
<td>0.901</td>
<td>0.964</td>
</tr>
<tr>
<td>NonPed</td>
<td>0.95</td>
<td>0.099</td>
<td>0.938</td>
<td>0.95</td>
<td>0.964</td>
</tr>
</tbody>
</table>

**Table 1. Recognition Rate for HOG of the classifier trained on the running set**

**Figure 5. ROC curve for pedestrians running classifier**

Classifiers trained on HoG features: Table 1 shows the detection rate of the classifier for the category pedestrians running(PedRun). The negative training set is shortly referenced with NonPed. We have depicted the true positive rate (TP), the false positive rate (FP), precision (Prec), recall (Rec) and area under ROC (ROC-A) for the classes running pedestrian and non-pedestrian.

Figure 5 presents the ROC for the classifier trained on pedestrian running set.

Table 2 shows the detection rate of the classifier trained for the category pedestrians standing (denoted in the table with PedStand).

<table>
<thead>
<tr>
<th></th>
<th>TP</th>
<th>FP</th>
<th>Prec</th>
<th>Rec</th>
<th>ROC-A</th>
</tr>
</thead>
<tbody>
<tr>
<td>PedStand</td>
<td>0.898</td>
<td>0.084</td>
<td>0.713</td>
<td>0.898</td>
<td>0.943</td>
</tr>
<tr>
<td>NonPed</td>
<td>0.916</td>
<td>0.102</td>
<td>0.975</td>
<td>0.916</td>
<td>0.943</td>
</tr>
</tbody>
</table>

**Table 2. Recognition Rate for HOG of the classifier trained on the set of standing pedestrians**

**Figure 6. ROC curve for pedestrians standing classifier**

**Table 3. Comparison of detection rates for the classifiers obtained using HOG features**

Classifiers trained on directional first order partial derivatives: Next we will present the results for the classifiers trained on feature vectors containing magnitude of first order partial derivatives computed in four directions. The results are not so promising as in the case of HOG, but these features can still be used in the future for other types of classifiers.

<table>
<thead>
<tr>
<th></th>
<th>TP</th>
<th>TN</th>
<th>TP</th>
<th>TN</th>
<th>TP</th>
<th>TN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Running test</td>
<td>0.901</td>
<td>0.95</td>
<td>0.55</td>
<td>0.91</td>
<td>0.76</td>
<td>0.91</td>
</tr>
<tr>
<td>Standing test</td>
<td>0.68</td>
<td>0.901</td>
<td>0.89</td>
<td>0.91</td>
<td>0.86</td>
<td>0.91</td>
</tr>
</tbody>
</table>

Table 4 makes a comparison of the recognition rates for the trained classifiers using HOG features. We depict the values of the true positive rate (TP) and of the true negative rate (TN). The results show that our approach which trained a classifier for each class of objects and formed a meta-classification scheme gives better results than the learner trained on the mixed set of pedestrians (running and standing).

**Table 4. Recognition Rate the classifier trained on the running set and on magnitude of first order partial derivatives**

Table 4 shows the recognition rate of the classifier for the category pedestrians running. We have depicted the
true positive rate (TP), the false positive rate (FP), precision (Prec.), recall (Rec) and area under ROC (ROC-A).

Figure 7 depicts the ROC of the classifier trained on using the direction derivative on pedestrians running set.

Table 5 shows the detection rate of the classifier for the category pedestrians standing trained using magnitude of first order partial derivatives.

<table>
<thead>
<tr>
<th>Classifier trained on</th>
<th>pedestrians running</th>
<th>pedestrians standing</th>
<th>running and standing</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>TP</td>
<td>TN</td>
<td>TP</td>
</tr>
<tr>
<td>PedStand</td>
<td>0.866</td>
<td>0.38</td>
<td>0.64</td>
</tr>
<tr>
<td>NonPed</td>
<td>0.947</td>
<td>0.134</td>
<td>0.995</td>
</tr>
</tbody>
</table>

Table 6 shows a comparison of the recognition rates for the trained classifiers using magnitude of first order partial derivatives as features.

We conclude that the recognition rates for the HOG are better than the ones for the magnitude of first order partial derivatives computed in different directions. Nevertheless, the performance of the detector trained on pedestrians standing using first order partial derivatives is quite accurate and its results can be improved. We have considered this set of features because our module is applied in the context of a real-time detection system in which fast computation of features is a must. Also, the CFS feature selection scheme can be replaced by more accurate feature selection methods.

For both sets of features, the resulting meta-classifier outperforms the detection rate of a classical learner trained on the whole pedestrian feature space.

Our detection window is of 18x36 pixels. We have not described the way in which the meta-classifier is applied for larger images because our recognition expert is designed to work within an existing probabilistic pedestrian detection framework[14], [1] which provides the pedestrian hypotheses.

6 Conclusion

This paper presented a method for pedestrian recognition based on the information provided by histograms of gradient orientations and magnitude of first order partial derivatives. The result of our research is a pedestrian recognition expert that is integrated in a stereo-based driving assistance system [14], [1].

Pedestrian attitudes have a high degree of complexity. We will study the space of pedestrian attitudes and investigate the parameters which provide a good partitioning scheme to offer more accurate recognition results.

The meta-classifier we have build runs its classifiers in parallel and the one that gives the best detection result labels an instance under test. Different ideas are currently under evaluation for improving the results. We want to build an automatic method for pedestrian attitude recognition, that will allow a single classifier from the meta-classification scheme to be active at one moment (ensuring by this a gain in speed, as not all the classifiers will run in parallel).
advanced classification algorithms can increase the accuracy of our results.

References


