

A Filter Module Used in Pedestrian Detection System

Gengxin Miao, Yupin Luo, Qiming Tian, Jingxin Tang
Department of Automation, Tsinghua University, Beijing, 100084,
P.R.China

mgx03@mails.tsinghua.edu.cn, luo@tsinghua.edu.cn,
tianqm00@mails.tsinghua.edu.cn, tangjx@mail.tsinghua.edu.cn

Abstract. Most pedestrian detection systems are built based on computer vision technology and usually are composed of two basic modules: object detection module, and recognition module. This paper presents an efficient filtering module, which works between the two basic modules, based on extracting the 3-dimensional information from single frame images. The filter module removes the noisy objects extracted by object detection module and thus reduces the burden of the recognition module. 3-D information, such as height, width and distance are extracted from single frame images. Using this information, a Bayesian classifier is employed to implement the filter. The main contribution of this filter module is that it removed about 30% noisy objects detected by the object detection module. The total computing cost and error detection rate is reduced when this filter module is used in the pedestrian detection system.

1 Introduction

According to the report [9], there are, every year, lots of pedestrians killed in traffic crashes in the US. Also, the pedestrian security problem is very serious all over the world. Considering this, the demand for computer vision based pedestrian detection systems arises.

The hardware of the pedestrian detection system is composed of a camera mounted on the front part of vehicle and a computer processing the video signals captured by the camera. Most of the existing computer vision based systems used a two-stepped algorithm: detection module and recognition module. Objects in the video are detected in the detection phase, and then these objects are classified with

Please use the following format when citing this chapter:

Miao, Gengxin, Luo, Yupin, Tian, Qiming, Tang, Jingxin, 2006, in IFIP International Federation for Information Processing, Volume 204, Artificial Intelligence Applications and Innovations, eds. Maglogiannis, I., Karpouzis, K., Bramer, M., (Boston: Springer), pp. 212–220

various pattern recognition techniques in the recognition phase. Example systems can be found in other researchers' studies [4][5][7].

The simple two-phased systems may lead to heavy computing burden since the detection module marked too many image candidates for the classification module. Tian et al. [6] have made some rigid rules to filter the objects extracted by the detection module according to the position of the candidate images. Broggi et al. [1] concerned the projection process from a real world to a 2-dimensional image. Both of the methods recognized that the height of the candidates is crucial information for removing the noisy candidates. Neither of them, however, calculated the 3D information online. On one hand, the filtering rate is low; on the other hand, there are lots of candidates being filtered by mistake.

This paper focuses on a filter model which works between the object detection module and the recognition module used in a computer vision based pedestrian detection system. In figure 1, an example image from the video can illustrate the problem encountered in the existing system.



Fig 1. An example image from the video

In figure 1, the yellow rectangles show the results extracted by the detection module, and the red rectangles represent the output of classification module which are expected to be pedestrians. In fact, the smaller red rectangle is not an image of a pedestrian, and just because its scale is too small so that based on the image information the classifier cannot recognize it from the pedestrian objects. In this image, too many yellow rectangles influence both the efficiency and the effectiveness of the recognition module.

In this filter module, the camera is primarily calculated with mature techniques [8]; since the camera is fixed on the vehicle, its parameters can be regarded as constant when the vehicle is operating on the road after the coordinate systems are finely established. In the calculation, the image information is totally ignored since that is the task of recognition module; the 3-dimensional information is calculated just based on the position where the candidate objects exist. Similar approach can be

found in [2]. After that, a Bayesian classifier is employed to accomplish the filter task.

The rest of the paper is organized as follows. Section 2 presents the camera module used in this module; Section 3 describes the detail of this filter module; the experimental results are shown in Section 4; and, at last, Section 5 draws the conclusion and proposes some future work.

2 Camera Model and Definition of Coordinate System

A camera consists of optical elements and an image surface. By varying the optics of the camera, many different projection of a given scene can be obtained. Among this set of projection models, the pinhole model [3] is the simplest projection model, as shown in formula (1)

$$s \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = A \begin{bmatrix} r_1 & r_2 & r_3 & t \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad (1)$$

where s = coefficient related to the object distance
 $[u, v]'$ = image coordinates
 A = the camera's intrinsic parameter
 $[r_1, r_2, r_3, t]'$ = the camera's extrinsic parameter
 $[X, Y, Z]'$ = object coordinates in the world coordinate system

The definition of these symbols will not be described again for the space limitation.

For intrinsic parameter calibration, lots of researchers have developed a variety of methods to implement the calibration. Among them, Zhang, Z. [8] proposed a flexible technique to easily calibrate the intrinsic parameters. For its convenience and precision, we implemented Zhang's method in our experiment.

The camera is fixed on the vehicle, and, thus, the relative position of the two coordinate systems is invariable; in other words, the external parameter of the camera can be fixed via a convenient definition of the camera coordinate system and the real world coordinate system. To simplify the problem, the image coordinate system and the world coordinate system are established as shown in figure 2.

As shown in figure 2, both of the two coordinate systems have the constantly parallel direction. Therefore, the extrinsic parameter matrix should be like this:

$$B = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & -h_c \\ 0 & 0 & 1 & 0 \end{bmatrix} \quad h_c \text{ is the camera's height} \quad (2)$$

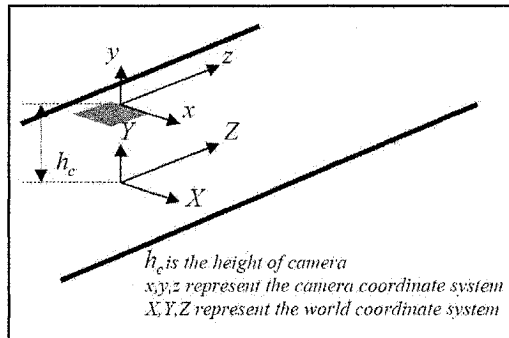


Fig 2. Camera and world coordinate systems

3 Our Approach

This section will mainly present two problems. 1) What are the data that the filter module will use to calculate the real size; and 2) What is the real size estimation process.

3.1 Online Size Estimation

The online size estimation algorithm is designed based on single frame images. To make this process feasible, some assumptions are proposed.

3.1.1 Assumptions: After estimating the camera parameters, the real size of the objects can be estimated based on some assumptions.

Assumption 1: Both the intrinsic and extrinsic parameters of the camera are invariable since the camera is fixed on the vehicle.

Assumption 2: The input objects are standing on the road plane. (Since the target of the module is pedestrian who are definitely standing on the ground)

Assumption 3: The road is planar.

3.1.2 Equations used for 3-dimensional information Estimation: Based on the camera module and the assumptions described above, the 3-dimensional information can be calculated from the single frame images.

Put the extrinsic parameter matrix (B described in Section 3.1) in formula 1. Thus:

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \frac{A}{s} \begin{bmatrix} X \\ Y - h_c \\ Z \end{bmatrix} \quad (3)$$

The formula can be further simplified as follows:

$$\begin{bmatrix} X \\ Y - h_c \\ Z \end{bmatrix} = s \begin{bmatrix} A^{-1}_{11} & A^{-1}_{12} & A^{-1}_{13} \\ 0 & A^{-1}_{22} & A^{-1}_{23} \\ 0 & 0 & A^{-1}_{33} \end{bmatrix} \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \quad (4)$$

Since A is an upper triangular matrix, its inverse matrix has the same character.

There are four points (named as bottom left, bottom right, top left, and top right) to present each object. Thus, according to assumption 2 and 3, the Y value of the bottom edge of the object should be zero. Thus:

$$\text{Equation 1: } \begin{bmatrix} X_l \\ -h_c \\ Z \end{bmatrix} = sA^{-1} \begin{bmatrix} u_l \\ v_b \\ 1 \end{bmatrix} \quad (5)$$

Furthermore, when examining the top and right edges of the object and the image, another equation can be achieved:

$$\text{Equation 2: } \begin{bmatrix} X_r \\ Y_t - h_c \\ Z \end{bmatrix} = sA^{-1} \begin{bmatrix} u_r \\ v_t \\ 1 \end{bmatrix} \quad (6)$$

Combining the two equations, the following real size related information can be calculated:

$$\begin{aligned} \text{Dis} = Z &= -h_c \frac{A^{-1}_{33}}{A^{-1}_{22}v_b + A^{-1}_{23}} \\ \text{Height} = Y_t &= h_c \frac{A^{-1}_{22}v_t - A^{-1}_{23}v_l}{A^{-1}_{22}v_b + A^{-1}_{23}} \\ \text{Width} = X_r - X_l &= -h \frac{A^{-1}_{11}u_r - A^{-1}_{11}u_l}{A^{-1}_{22}v_b + A^{-1}_{23}} \end{aligned} \quad (7)$$

3.3 Classifier

Three attributes (the distance, the height, and the width) can be computed to represent each object. Among of the three, the height is the most crucial cue to filter those non-pedestrian objects. Since the height to width ratio is constant (the objects are extracted based on template matching algorithm), the width information is omitted in the further calculation. The pedestrians can stand at any distance from the vehicle; in light of this, the distance information seems to be of less use either. However, when the distance increases, the estimation of height will be involved with

a larger error. Therefore, the two dimensional information of both height and distance are used as the input of the Bayesian classifier.

To assure the formula has an analytical solution, a Gaussian function is used to approximate the probability density function of the distribution of real height and distance of the candidate objects. All the samples should be firstly divided into two categories, the positive and the negative (which means the pedestrian objects and the non-pedestrian objects). Then the estimation of the mean value and the variance of the height and distance corresponding to each category can be achieved.

Thus, the discriminant functions and the decision surface can be expressed as follows.

$$g_p(x) = -\frac{1}{2} \begin{bmatrix} Dis - \mu_{dp} \\ Height - \mu_{hp} \end{bmatrix}^T \begin{bmatrix} \Sigma_{dp} & 0 \\ 0 & \Sigma_{hp} \end{bmatrix}^{-1} \begin{bmatrix} Dis - \mu_{dp} \\ Height - \mu_{hp} \end{bmatrix} - \frac{1}{2} \ln \begin{vmatrix} \Sigma_{dp} & 0 \\ 0 & \Sigma_{hp} \end{vmatrix} \quad (8)$$

$$g_{np}(x) = -\frac{1}{2} \begin{bmatrix} Dis - \mu_{dnp} \\ Height - \mu_{hnp} \end{bmatrix}^T \begin{bmatrix} \Sigma_{dnp} & 0 \\ 0 & \Sigma_{hnp} \end{bmatrix}^{-1} \begin{bmatrix} Dis - \mu_{dnp} \\ Height - \mu_{hnp} \end{bmatrix} - \frac{1}{2} \ln \begin{vmatrix} \Sigma_{dnp} & 0 \\ 0 & \Sigma_{hnp} \end{vmatrix}$$

According to formulae (7) and (8), the decision surface should be decided below:

$$f(x) = g_p(x) - g_{np}(x) = \frac{1}{2} \left((x - \mu_{np})^T \Sigma_{np}^{-1} (x - \mu_{np}) - (x - \mu_p)^T \Sigma_p^{-1} (x - \mu_p) \right) + C \quad (9)$$

4 Experimental Results

Our experiment is implemented based on a tape record derived from the real scenery in the streets during nighttimes. In the whole video file, there are 1,144,323 candidate objects extracted. Firstly, the video data are divided into two parts: some data for training and the other for test. There are 500,000 objects used for training and the rest for testing. In all the object samples, there are 9,333 positive objects and the others are negative ones.

4.4 Parameter Tuning

4.1.1 Probability Density Function Estimation: In order to design a classifier based on the height and distance information of the objects, the different distribution characters on these features of the two categories of objects should be examined first. In this section, the probability density function of the height and distance of both groups of objects is estimated.

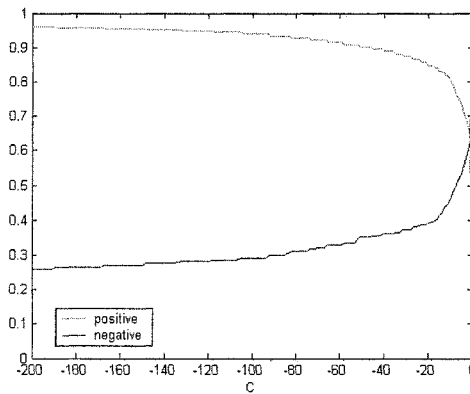
As mentioned above (see Section 3.3), this section uses Gaussian function to estimate the probability density function of the height and distance of both positive and negative objects. The results are shown in the following table.

Table1. Results of probability density function estimation

	Mean value μ	Variance Σ
Positive objects, distance	52.6m	4556.25
Positive objects, height	1.109m	1.9
Negative objects, distance	111m	77952.6
Negative objects, height	2.75m	51.7

4.1.2 Decision Surface Estimation: Theoretically, the height of object is independent to its distance from the vehicle. However, when the objects go farther from the camera, the image will be blurred and, consequently, the error in estimating its distance and height will be greater. Therefore, both the height and the distance information are useful to establish the Bayesian classifier.

Figure 3 shows the parameter tuning process in which the curve labeled as *positive* shows the retain rate of pedestrian objects when the parameter of the decision surface varies; and the *negative* curve shows the filter rate of non-pedestrian objects when the parameter of the decision surface varies.

**Fig 3. Parameter tuning graph**

The parameters should be chosen based on this principle: the most of the negative objects should be filtered and most of the positive ones should be retained. However, there is not a clearly optimal value shown in this figure, since when the negative objects' curve increases (which means the filter rate is growing), the positive objects' curve drops (which means the loss of pedestrian objects is also growing). As shown in figure 3, -100 is chosen as a decision surface.

5 Test Results

Once the decision surface is determined, it can be used to classify the test data collection. The result of the filter is shown below.

In figure 4, the vertical axis represents the number of candidate objects for both categories. The horizontal axis represents the parameter of the decision surface.

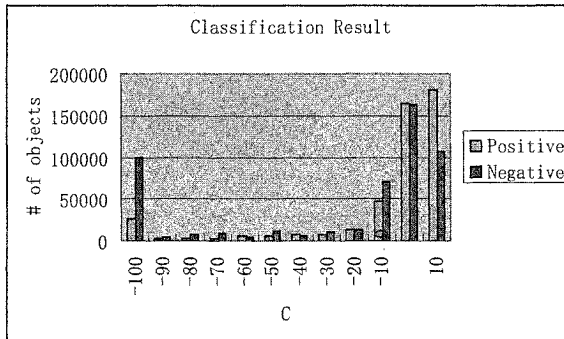


Fig 4. Test results

Because the pedestrian samples are too few to be comparable with the non-pedestrian samples, to amplify the result, this figure magnifies the number of pedestrian objects by 500 times. To make it clearer, the two ratio values (the two-ratio values, or the values of the two ratios) are calculated as follows.

$$\frac{\text{Left samples}}{\text{Total samples}} = 71.4\% \quad (10)$$

$$\frac{\text{Left positive samples}}{\text{Total positive samples}} = 94.1\%$$

In formula 10, the test results are shown clearly. The filter module finally removed 28.6 percents of the candidate objects, among which there are only 5.9 percents of the positive objects which are removed by mistake. This result is very satisfying for the further process. On one hand, by removing a lot of candidates, the following modules can save lots of time in processing them. On the other hand, most of the removed objects (326,716 among 327,276, almost 98.8%) are non-pedestrian objects. Thus, as a result, the wrong detection ratio of the following modules can be decreased.

5.1 Discussion

The wrong detection is mainly produced in two ways: 1) the approximated camera parameters import errors in calculating the real size information especially when the objects are in long distances. And 2) the mistaken segmentation of objects

offered by former detection module brings an inefaceable error in the following calculation. These two kinds of errors weakened the power of the filter module. To reduce the effect of the first type of error, the classification algorithm also had the distance information as an input; in this way, the results are improved remarkably. For the limitation of space, the former experimental results without considering the distance information are not included in this paper.

6 Conclusion and Future Work

To further improve the performance of this filter module, the two kinds of errors should be studied in depth. For the first type of error, the error of the camera parameters can be estimated, and consequently the error of the calculated real size information can be estimated as well as the height and distance information. Currently, the filter module avoided processing the raw image information, thus it can't get rid of the second type of error. Most of the time, the mistaken segmentation contains only a half part of a pedestrian dressing in different colors for the coat and the trousers. For future work, the filter module can detected a small range in the image and examine whether the pedestrian image is wrongly segmented. These will potentially improve the performance of this filter module.

References

1. Broggi, A. Fascioli, A. Carletti, M. Graf, T. Meinecke, M.: A multi-resolution approach for infrared vision-based pedestrian detection. 2004 IEEE intelligent vehicles symposium. Parma, Italy, (2004) Pages: 7-12
2. Criminisi, A.; Reid, I.; Zisserman, A.: Single view metrology. Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on Volume 1, 20-27, 1999.9 Pages:434-441 vol.1
3. David, A.: Computer Vision: a modern approach, ISBN:0130851981. Chapter 2
4. Gavrila, D.: Pedestrian detection from a moving vehicle. European Conference on Computer Vision. Trinity College Dublin, Ireland. (2000) Pages: 37-49
5. Nanda, H.: Probabilistic template based pedestrian detection in infrared videos. 2002 IEEE intelligent vehicles symposium. EPFL, Switzerland (2000) Pages: 15-20 vol.1
6. Qiming Tian, Yupin Luo, Dongcheng Hu: Pedestrian detection in nighttime driving. ICIG'04 Hongkong China (2004) Pages:116-119
7. Bin Zhang, Qiming Tian, Yupin Luo: A multi-layer classifiers framework for pedestrian detection in cluttered nighttime background Visual Communications and Image Processing Beijing, China (2005) Pages: 143-148
8. Zhengyou Zhang: Flexible Camera Calibration by Viewing a Plane from Unknown Orientations. ICCV 1999 Kerkyra, Greece, (1999) Pages: 666-673 vol.1
9. <http://www-nrd.nhtsa.dot.gov/departments/nrd-30/nrsa/> National center for statistics and analysis of the national highway traffic safety administration (accessed 12. May, 2005)