FIR pd etc etc etc

A. Broggi, M. Bertozzi, M. Felisa, M. Del Rose and Frederic Suard

***Abstract*— This paper presents... udda udda udda**

I. Bounding boxes analysis

In this part, we will describe in details the method to classify the content of bounding boxes. As we said in previous section, the result of stereovision process is a list of bounding boxes. Each bounding box produces one image with an object.

First we will describe the image with a descriptor which extract the discriminant information contained in this image.

Then we analyse the descriptor with a classifier, in our case the Support Vector Machines.

A. HoG descriptor

In 2005, Shashua et al. [4] had already introduced a descriptor using the information of gradient orientation. But he proposed to computed local histograms within small regions corresponding to the human morphology. Dalal et al. [3] has extended the use of histograms but with a dense approach.

The computation of a descriptor is done according the following steps :

1) compute horizontal G_H and vertical G_V gradient of image by filtering image with $[-1\ 0\ 1]$
2) compute both norm and orientation of the gradient :
   - $N_G(x,y) = \sqrt{G_H(x,y)^2 + G_V(x,y)^2}$ (figure 1-b)
   - $O_G(x,y) = atan\left(\frac{G_H(x,y)}{G_V(x,y)}\right)$ (figure 1-c)
3) split image into cells (figure 1-d),
4) compute one histogram for each cell (figure 1-e),
5) normalize all histograms within a block of cell.

The last step is a specificity of this descriptor, since the normalization can reduce the illumination variability. The final descriptor is obtained by adding all normalized histograms into a single vector.

B. SVM classifier

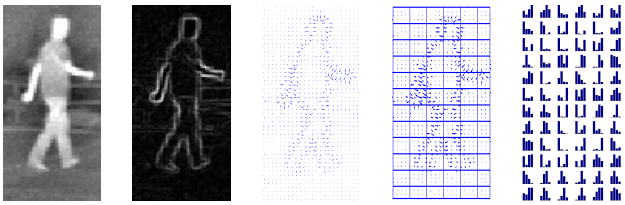
The recognition system is based on a supervised learning technique. Hence, we have used a set of training image examples with and without pedestrians, and described by their HoG, to learn a decision function. In our case, we have used a Support Vector Machines classifier.

This work has been supported by the European Research Office of the U. S. Army under contract number N62558-05-P-0380.

A. Broggi, M. Bertozzi, M. Felisa are with the Dip. Ing. Informazione, Università di Parma, ITALY. http://www.vislab.it

Frederic Suard is with Laboratoire LITIS (PSI), INSA de Rouen, France.

M. Del Rose is with XXXXXXXXXXXXXXX



a b c d e

Fig. 1. Image characterization using HoG : original image (a), gradient norm (b), gradient orientation (c), cell splitting (d) and histogramm computation (e).

The Support Vector Machines classifier is a binary classifier algorithm that looks for an optimal hyperplane as a decision function in a high-dimensional space [1], [5], [2]. Thus, consider one has a training data set $\{\mathbf{x}_k, y_k\} \in \mathcal{X} \times \{-1, 1\}$ where $\mathbf{x}_k$ are the training examples HOG feature vector and y_k the class label. At first, the method consists in mapping $\mathbf{x}_k$ in a high dimensional space owing to a function Φ. Then, it looks for a decision function of the form : $f(\mathbf{x}) = \mathbf{w} \cdot \Phi(\mathbf{x}) + b$ and $f(\mathbf{x})$ is optimal in the sense that it maximizes the distance between the nearest point $\Phi(\mathbf{x}_i)$ and the hyperplane. The class label of $\mathbf{x}$ is then obtained by considering the sign of $f(\mathbf{x})$. This optimization problem can be turned, in the case of L_1 soft-margin SVM classifier (misclassified examples are linearly penalized), in this following way :

$$\min_{\mathbf{w},\xi} \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{k=1}^{m} \xi_k \tag{1}$$

under the constraint $\forall k, \quad y_k f(\mathbf{x}_k) \geq 1 - \xi_k$. The solution of this problem is obtained using the Lagrangian theory and it is possible to show that the vector $\mathbf{w}$ is of the form :

$$\mathbf{w} = \sum_{k=1}^{m} \alpha_k^* y_k \Phi(\mathbf{x}_k) \tag{2}$$

where α_i^* is the solution of the following quadratic optimization problem :

$$\max_{\alpha} W(\alpha) = \sum_{k=1}^{m} \alpha_k - \frac{1}{2}\sum_{k,\ell}^{m} \alpha_k \alpha_\ell y_k y_\ell K(\mathbf{x}_k, \mathbf{x}_\ell) \tag{3}$$

subject to $\sum_{k=1}^{m} y_k \alpha_k = 0$ and $\forall k, 0 \leq \alpha_k \leq C$, where $K(\mathbf{x}_k, \mathbf{x}_\ell) = \langle \Phi(\mathbf{x}_k), \Phi(\mathbf{x}_\ell) \rangle$. According to equation (2) and (3), the solution of the SVM problem depends only on the Gram matrix K.

AUC		10	50	100	500
Tetranight	FIR	0.9554	0.9602	0.9662	0.9704
	VIS	0.9364	0.9447	0.9523	0.9550
Tetravision	FIR	0.7304	0.8374	0.8622	0.8935
	VIS	0.7416	0.8460	0.8618	0.8977

TABLE I

VALUE OF AREA UNDER CURVE FOR EACH SEQUENCE, WHEN THE SIZE OF THE LEARNING SET VARIES.

Recognition rate		10	50	100	500
Tetranight	FIR	0.8727	0.8993	0.9060	0.9144
	VIS	0.8546	0.8832	0.8893	0.8999
Tetravision	FIR	0.6583	0.7555	0.7784	0.8121
	VIS	0.6702	0.7603	0.7822	0.8165

TABLE II

RECOGNITION RATE FOR EACH SEQUENCE.

II. RESULTS

In this last part, we will presents some results of our system.

A. Stereovision

B. HoG

We evaluated the HoG method with 2 video sequences. The first : Tetravision05, was taken during day with good luminisity conditions. The second : Tetranight01, was taken during night. For each sequence, we used both a visible and infrared stereovision system.

Thanks to the tetravision system which was described previously, a list of bounding boxes can be extracted from all sequences. For each bounding box, we extracted the corresponding image and labeled it manually as pedestrian or non-pedestrian. An image is labeled as a pedestrian if it contains only one person, which is centered. The size of the pedestrian should also be the size of the bounding box. Figure 2 shows some examples of pedestrians and non-pedestrians. The table below shows the number of pedestrians and non-pedestrians images which were labeled :

	Tetravision05		Tetranight01	
	FIR	VIS	FIR	VIS
pedestrian	2255	1860	1678	1359
non-pedestrian	20246	20520	2933	3262

All images are then resized to the same size : 128×64 pixels. This operation is due to the fact that the descriptor works on images with the same size.

We evaluated independantly each category : tetravision visible, tetravision infrared, tetranight visible and tetranight infrared. We extracted randomly a subset of images to build a learning set. We learn the linear Support Vector Machine classifier on this subset and test on a random subset of 500 images of pedestrian and 500 non-pedestrian. It should be notice that images of the test dataset are not images used during the learning step.

To improve the reliability of these results, each test was iterated 10 times, and we renewed randomly the learning set and test set for each iteration. The given results are the average of all tests.

To evaluate the performance of our system, we compute the rate of true positives against the rate of false positives, and we compute the area under the curve (AUC) which is obtained. The table I shows our results. We also determined the good recognition rate on the table II. We also evaluate the performance when the size of the learning set varies.

We can note that we obtained very good results, particularly during the night (Tetranight) with up to 91% of good recognition rate for the infrared night sequence. We can explain this by the fact that with the infrared images the pedestrian is warmer than the background. Concerning the visible images, all pedestrian detected are pedestrian located in front of the car, so images are well defined. Moreover, due to the headlight, the pedestrian could be easy detached from the background.

During the day, shapes are less contrasted, so that the characterization is not facilitated. This fact could explain the lower performance achieved during day.

If we consider the generalization capacity of this method, the performance is quite optimal for a learning set containing only 50 pedestrian and 50 non-pedestrian. When the the learning set becomes larger, the system performs better.

REFERENCES

[1] B. E. Boser, I. Guyon, and V. Vapnik. A training algorithm for optimal margin classifiers. In *Computational Learing Theory*, pages 144–152, 1992.

[2] N. Cristianini and J. Shawe-Taylor. *Introduction to Support Vector Machines*. Cambridge Univeristy Press, 2000.

[3] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In C. Schmid, S. Soatto, and C. Tomasi, editors, *International Conference on Computer Vision and Pattern Recognition*, volume 2, pages 886–893, June 2005.

[4] A. Shashua, Y. Gdalyahu, and G. Hayon. Pedestrian detection for driving assistance systems: Single-frame classification and system level performance. In *Proceedings of IEEE Intelligent Vehicles Symposium*, 2004.

[5] V. Vapnik. *The Nature of Statistical Learning Theory*. Springer, N.Y, 1995.

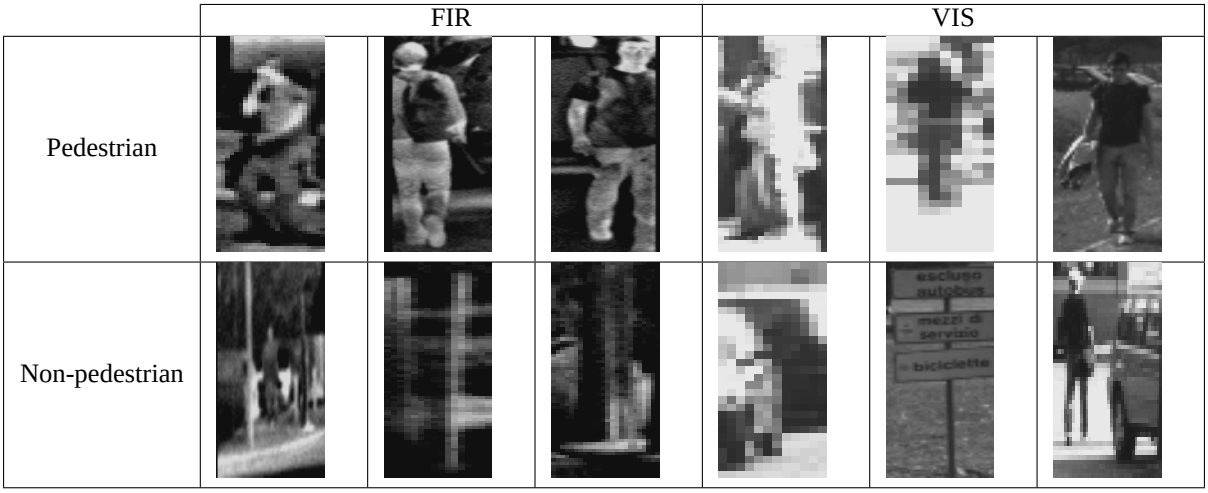


Fig. 2. Examples of pedestrian and non pedestrian images extracted for the tetravision sequence.