2004 IEEE Intelligent Vehicles Symposium
University of Parma
Parma, Italy • June 14-17, 2004

Pedestrian Localization and Tracking System with Kalman Filtering

M. Bertozzi, A. Broggi, A. Fascioli, A. Tibaldi
Dipartimento di Ingegneria dell'Informazione
Università di Parma
Parma, I-43100, ITALY
{bertozzi,broggi,fascal,tibaldi}@ce.unipr.it

R. Chapuis, F. Chausse
LASMEA UMR 6602 UBP/CNRS
Université de Clermont-Ferrand
Clermont-Ferrand, FRANCE
{chapuis,chausse}@lasmea.univ-bpclermont.fr

Abstract— **This work presents an implementation of a vision-based system for recognizing pedestrians in different environments and precisely localizing them with the use of a Kalman filter estimator configured as a tracker. Pedestrians, in various poses and with different kinds of clothing, are first recognized by the vision subsystem through the use of algorithms based on edge density and symmetry maps. The information produced in this way is then passed on to the tracker module which reconstructs an interpretation of the pedestrians positions in the scene. An appropriately configured indoor system setup with an accurate measurement of the imposed human trajectory has been realized. This setup has permitted an accurate evaluation of the accuracy of the results, when the new auxiliary tracker is activated.**

I. INTRODUCTION

Vehicles that automatically perform safety tasks like detection of pedestrians will have an important role in the future of an intelligent transportation system. The possibility to appropriately equip an high number of vehicles will allow to reduce the casualties derived from accidents, specially in the urban environment. Pedestrian localization in outdoor scenes is a challenging task because of the variety of the environments and of the clothes. A moving vehicle has to deal with a lot of problems: noise produced by the presence of buildings and human artifacts, different illumination conditions, obstacles and so on.

Widely used approaches for addressing vision-based pedestrian detection are: the search of specific patterns or textures [1], shape detection [2], [3], [4], [5], [6], [7] and neural nets-based methods [8].

This work presents the system introduced in [9] that is aimed at the localization of pedestrians by means of vision. This system has been designed to be installed on board of moving vehicles in order to provide the driver with warning signals. In particular, the implementation of a new tracking layer based on Kalman filtering [10], [11], [12] for this system is examined and the article mainly deals with performance measurements of the system activity.

This paper is composed of the following sections:

- section II presents the system scheme,
- section III introduces the new tracking functionality of the system,
- section IV summarizes the most valuable numerical results obtained,
- section V discusses the results and outlines the possibilities for future improvements of the system.

II. SYSTEM STRUCTURE

In this section the components of the pedestrian localization system are briefly explained. Fig. 1 depicts the relationships between the system components that perform the following tasks:

- "Preattentive Phase" - low level vision elaboration,
- "Symmetry Detection" - symmetry maps evaluation,
- "Bounding Boxes Generation" - pedestrians outlining,
- "Bounding Boxes Filtering" - pedestrian boxes selection,
- "Pedestrian Localization" - spatial position estimation for pedestrian boxes,
- "Bounding Boxes Tracking" - state variables and associated accuracy evaluation.

A. Preattentive phase

The knowledge of the vision system's extrinsic parameters and the flat scene assumption allows to reduce the search for candidates to one limited part of the image, and reasonable ranges and steps are considered for dealing with different pedestrian dimensions.

Fig. 2 (*a*) presents the image clustering and edge extraction performed at this stage of the processing. Besides the obvious advantage of avoiding false detections in wrong areas, this technique, combined with an undersampling procedure, strongly reduces the computational time needed for frame elaboration and shows excellent temporal results.

B. Symmetry detection

After the low level preprocessing and the analysis of vertical symmetry maps derived from gray-level and horizontal gradient image values, the identification of regions that can be characterized as human shapes takes place. Since pedestrians evidence an high symmetry, especially vertical, image columns can be considered as possible symmetry axes and edges can be used as a discriminant in a preattentive *filtering* stage (see fig. 2 (*b*)).

Approaches based onto this kind of maps have already been illustrated with the Generalized Symmetry Transform (GST) [13].

0-7803-8310-9/04/$20.00 © 2004 IEEE

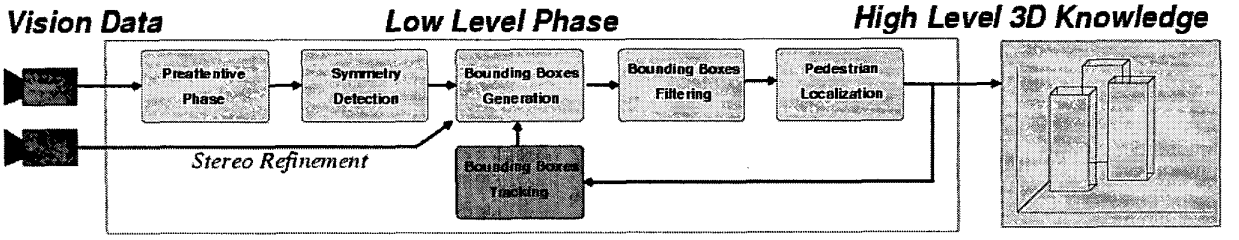


Fig. 1. The system architecture.

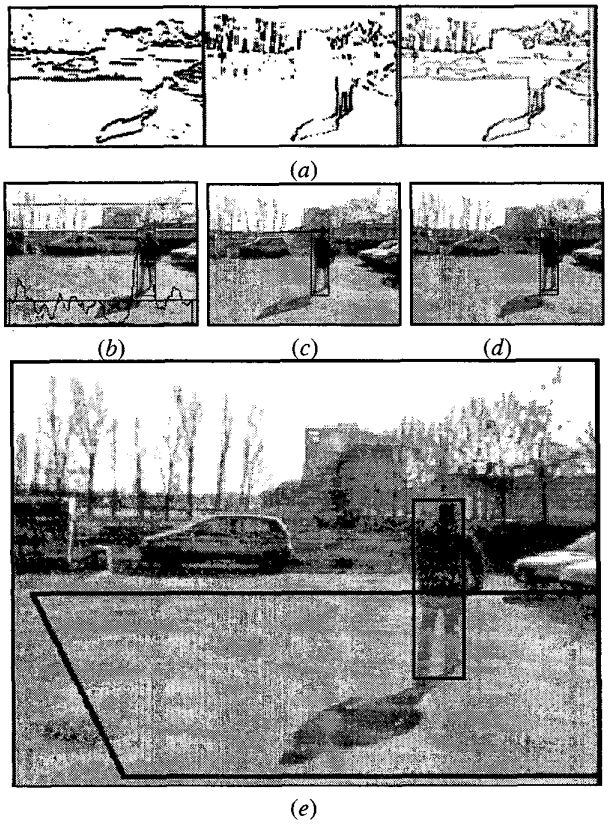


Fig. 2. The vision algorithm processing stages for an example outdoor image acquired from a moving vehicle: (*a*) low level horizontal, vertical and combined edges; (*b*) preattentive filtering; (*c*) search range for the homologous box; (*d*) stereo refinement for the base of the box; (*e*) result (the stereo search area is surrounded with a border).

C. Bounding boxes generation

The axis-based approach is followed by maps analysis for the extraction of the boxes and after this a particular *stereo refinement* technique is used to improve the accuracy of the identification of the boxes' bases (see fig. 2 (*c,d*)). Fig. 2 (*e*) presents the box generation result for an image acquired from a moving vehicle.

This processing level produces boxes with an high probability to fit one pedestrian, and candidates are characterized by problem specific dimensions in pixels and symmetry axes placed nearby the peaks of relative maximums in the axial weighted symmetry sum.

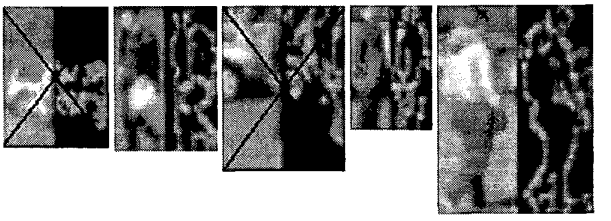


Fig. 3. Bounding boxes filtering: the discarded pedestrian candidates are marked with a black "x", each example shows the original and the edges inside the candidate bounding box.

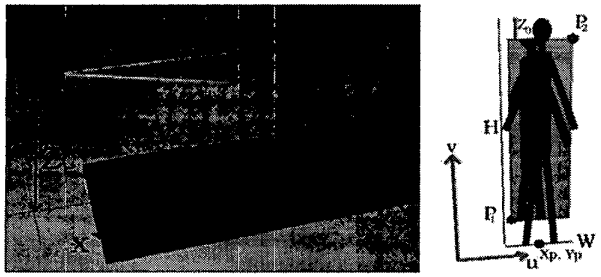
D. Bounding box filtering

Unfortunately symmetrical objects other than pedestrians may happen to be detected as well. In order to get rid of such false positives a number of filters based on regionalization have been devised and are still under development. Fig. 3 illustrates some examples of how the filters check the eligible candidates and eliminate some of them that do not actually represent a human shape. These filters evidence promising results with artifacts such poles, road infrastructures, traffic signs and buildings that cause the box generation to fail.

E. Pedestrian localization

This module estimates the position of the pedestrians in the scene in the chosen coordinate reference system. The contact point (X_p, Y_p) of each pedestrian vertical axis with the ground assumed flat is associated with opportune state variables for this purpose. The height from the ground Z_c, the tilt angle α of the camera observing the scene and a set of intrinsic calibration parameters represented by e_u and e_v must be known. Fig. 4 (*a*) shows the coordinate system in which the contact point is defined according to the road plane and also the position of the camera.

The estimation uses an original modeling that takes explicitely into account the unavoidable difference of the vision-detected bounding box of a pedestrian with the real ideal one, defined by a height H and by a width W with fixed average and standard deviation values, realistic enough to represent a human shape ($H = 1.65\ m$, $\sigma_H = 0.1\ m$, $W = kH$ with $k = 0.3$ a realistic width/height



(a) (b)

Fig. 4. Setup scheme and problem variables: (*a*) world coordinate reference system; (*b*) image coordinates of the scene bounding box.

ratio, $\sigma_W = 0.1\ m$). Z_0 represents half of the difference between the scene bounding box height and the real pedestrian height, with average value zero and standard deviation $\sigma_{Z_0} = 0.1\ m$.

Considering a perspective projection of the scene onto the image, the relationship between the coordinates of the corners $P_1 = (X_1, Y_1, Z_1)^t$ and $P_2 = (X_2, Y_2, Z_2)^t$ of a pedestrian bounding box in the camera coordinate system can be linked in a linear way, thanks to small angle approximation for α, to the planar position coordinates X_p and Y_p (1). The correspondent image coordinates $p_1 = (u_1, v_1)$, $p_2 = (u_2, v_2)$ and the observation system $\mathbf{Y} = H \cdot \mathbf{X} + \mathbf{v}$ (2) are then easily deduced.

$$\begin{cases} X_1 = X_P - \frac{W}{2} \\ Y_1 = Y_P - \alpha(Z_0 - Z_c) \\ Z_1 = \alpha Y_P + (Z_0 - Z_c) \end{cases} \text{and} \begin{cases} X_2 = X_P + \frac{W}{2} \\ Y_2 = Y_P - \alpha(H - Z_0 - Z_c) \\ Z_2 = \alpha Y_P + (H - Z_0 - Z_c) \end{cases} \tag{1}$$

$$\begin{pmatrix} -e_v(Z_0 - Z_c) \\ e_u \frac{W}{2} \\ -e_v(H - Z_0 - Z_c) \\ -e_u \frac{W}{2} \end{pmatrix} = \begin{pmatrix} 0 & -v_1 \\ e_u & -u_1 \\ 0 & -v_2 \\ e_u & -u_2 \end{pmatrix} \begin{pmatrix} X_P \\ Y_P \end{pmatrix} + \underline{v} \tag{2}$$

The contribution of all the parameters subject to error is also taken into account with the use of the covariance matrix of the noise vector $\underline{v}$, in order to improve the estimation of the positions of the pedestrians, concretely realized with a Kalman filter.

More details on how this modeling deals with the pedestrian spatial positioning are available in [9]. A new bounding box tracking stage now completes the approach.

III. Bounding Boxes Tracking

In this section the implementation of the tracker is explained in terms of design choices, box management queues, tracking politics and Kalman filtering integration.

Each new pedestrian identified by the localization is provided with an unique `id`. This is used to drop the box if the timeout for joining with an appropriately new detected pedestrian expires, to log the history of the pedestrian path, to differentiate it from the others and also to render clearly all the graphical information (see fig. 6).

The tracker presents a flexible politic for data logging, box processing, matrixes allocation and an efficient method encapsulation for complex procedural sections; the subsystem presents two possible working modalities: *single tracking* and *multi tracking* mode; these discriminates the way in which the rejoining of lost traces is managed.

The main tasks of the tracker are: the *merge* of visual localizations with the internal state representation, the calculations relating to the evolution of the state of each pedestrian (through Kalman filtering), and the prediction projection for the triggering of the elaboration. Input and output buffering queues are used for filtering purposes too, in order to implement insertion and removal politics that enhance the reliability of noisy sensor data.

Graphics are used to illustrate the state variables history of the pedestrian boxes. This is done for sake of an efficient and constant system check by the human supervisor, both in the perspective image (fig. 6 (*a*)) and in the road *top view* plane (fig. 6 (*b*)). The current box position and the position prediction, in the form of probability-blended image projection areas, are drawn in the perspective representation. Moreover the error ellipsis for each box is represented on the experimental road plane image.

The *merge function* performs one feedback task related to the association of newly detected pedestrians with the set of spatially localized and tracked ones. This approach solves problems related to wrong estimations and temporal mismatches. It is based on box areal overlapping and Mahalanobis distance estimation, and is responsible for updating the tracked set of pedestrians. The overlapping criterion is based on probability image areas after Kalman prediction and the metric criterion exploits the state of each tracked pedestrian. Instead of a Mahalanobis distance classification *tout-court*, the product $H \cdot \mathbf{X}$ is used as observations for the evaluation of the metric r.

The most effective formula for the extraction of r at the iteration κ for the vision observation n and the consequent classification has been found to be the match criterion in (3),

$$\begin{aligned} & r_n^{(k,)}(i) = \Delta^\top \cdot C^{(k-1,i)^{-1}} \cdot \Delta \\ & \Delta \triangleq [H_n^{(k,)} \cdot \mathbf{X}_\mathbf{n}^{(\mathbf{k},)} - H^{(k-1,i)} \cdot \mathbf{X}^{(\mathbf{k-1,i})}] \\ & \texttt{match} \rightleftharpoons i \mid \min\{r_n^{(k,)}(i)\} \leq t^* \end{aligned} \tag{3}$$

where a generic matrix denoted as $A^{(h,p)}$ refers to the tracked pedestrian p at the h-th iteration of the tracker and t^* is an opportunely chosen threshold.

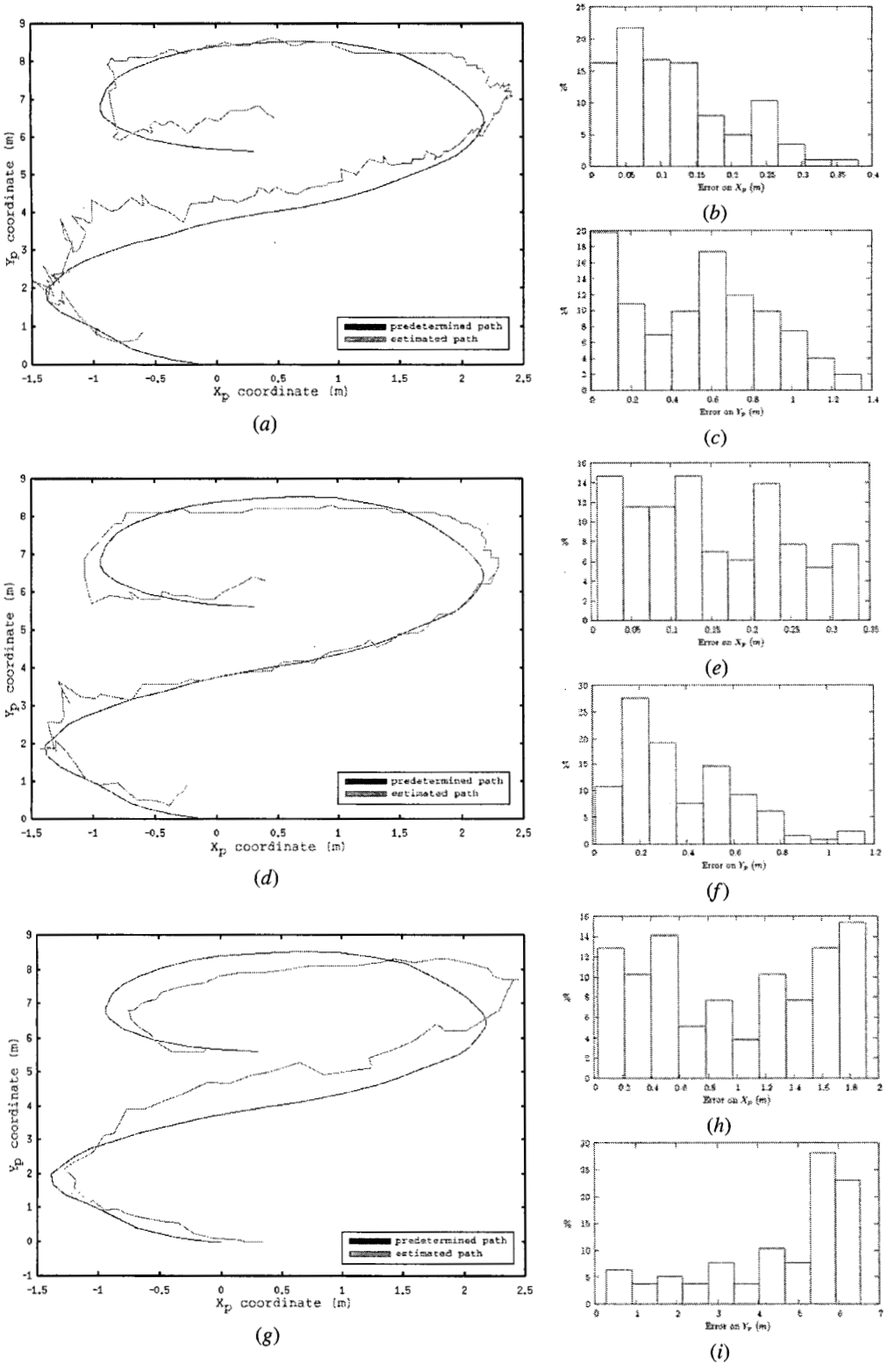
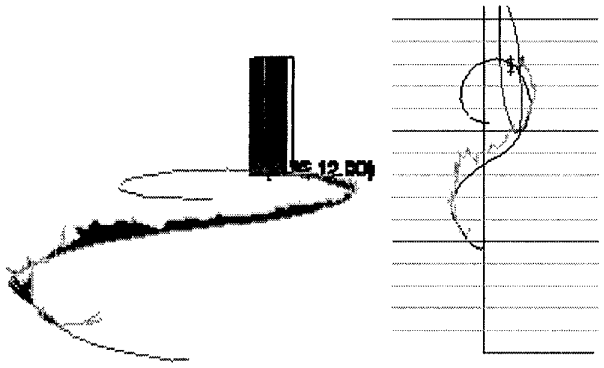


Fig. 5. Comparison between the estimated paths and the trajectory: (*a*) Planar estimation for the slowly forward walking experiment; (*b*) normalized histogram of the error in the X_p coordinate for the slowly forward walking experiment; (*c*) histogram of the error in the Y_p coordinate; (*d*,*e*,*f*) analog data representation for the regular speed forward walking experiment; (*g*,*h*,*i*) data for the backwards running experiment.



(a) (b)

Fig. 6. Pedestrian path and reference trajectory for the indoor acquisition as reported by the tracker when the *single-tracking* mode is selected: (*a*) perspective projection; (*b*) top view of the ground plane, the predefined trajectory is shown in black and the measured trajectory is shown in green.

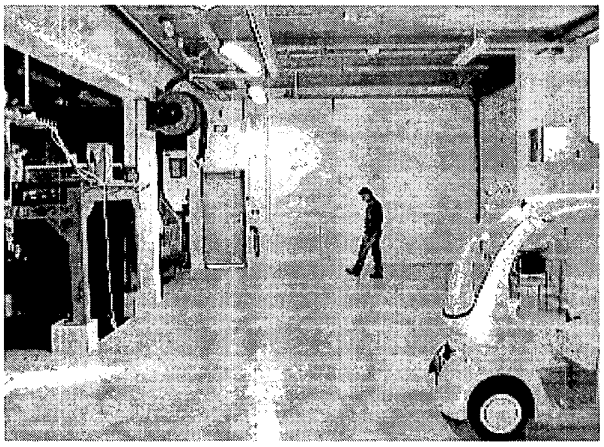


Fig. 7. Indoor test setup.

IV. PRECISION EVALUATION

Indoor experiments have been realized in order to verify the correctness of the estimated coordinates of the pedestrian path with the use of the tracker.

A reference trajectory has been set up simply by defining a set of way-points on the ground sufficiently close one to the next. The X and Y coordinates of all these points have been measured with classical measurement instruments in the chosen reference system. A calibrated camera has been positioned to look at this trajectory, with height and orientation as if it was installed inside a car. The position of the camera observing the reference points in the coordinate reference axes has been determined. Fig. 6 (*a*) shows a camera view of the pedestrian trajectory including perspective and fig. 6 (*b*) shows the associated bird-eye view.

The test includes the movement of a pedestrian along the predefined trajectory and the acquisition of the corresponding images in order to post-process them with the vision algorithm. An example of image acquired in this way is provided in fig. 7. For convenience the experiments have been realized indoor; due to this the images presented many vertical edges that lead to the generation of additional noise caused by the indoor structure. To solve this problem, a simple background subtraction has been applied in the middle of the processing; of course this is not needed in outdoor scenes: it is only used to make the localization verification possible.

TABLE I

MEAN AND MAXIMUM COORDINATE ERRORS (*m*)

Sequence	$\bar{e}_{X_p}$	Max e_{X_p}	$\bar{e}_{Y_p}$	Max e_{Y_p}
forward slowly	0.12	0.38	0.53	1.35
forward regular	0.15	0.34	0.37	1.16
forward running	0.47	1.48	0.47	2.04
forward natural	0.17	0.44	0.39	1.98
backwards slowly	0.66	1.54	4.07	6.89
backwards regular	0.65	1.38	4.19	6.98
backwards running	0.98	1.87	4.47	6.46
backwards natural	0.76	1.67	2.66	5.38

Fig. 6 (*b*) shows a superposition example of the reference trajectory and of the trajectory estimated by the system. The current pedestrian position and its covariance ellipsis are also drawn together with the estimated trajectory.

Thanks to the use of a digital camera, the experiments have been characterized by a known intraframe temporal gap, so that the temporal synchronization of the estimated and of the reference trajectories has been made possible through a parameterization. Since the estimator provides the values of X_p and Y_p separately, it has been possible to compare the X and the Y coordinates independently. The maximal and average errors measured for the X_p and Y_p planar pedestrian coordinates of the experiments are reported in table I; one time plot example of the euclidean error is in fig. 8. Fig. 5 shows plots and error composition histograms regarding the resulting estimated paths for various ways of covering the trajectory. Considering that the Y coordinate is the one related to the depth of the scene relatively to the camera, the fact that the error on Y_p is greater than the one on X_p is not surprising and is a rather obvious conclusion in the field of computer vision; however, the overall precision is remarkable.

Another significant result obtained with these experiments is that the average errors on X_p and Y_p obtained by measurements coincide with the *a priori* error estimated at the output of the Kalman filter. This has the important meaning that the errors provided by the estimator can be considered reliable.

V. CONCLUSIONS

The high-level tracking module for *environmental understanding* has evidenced with its filtering capabilities a good accuracy in the spatial localization of a walking pedestrian. The maximum errors from the measured path and the maximum variances along the axes that have been

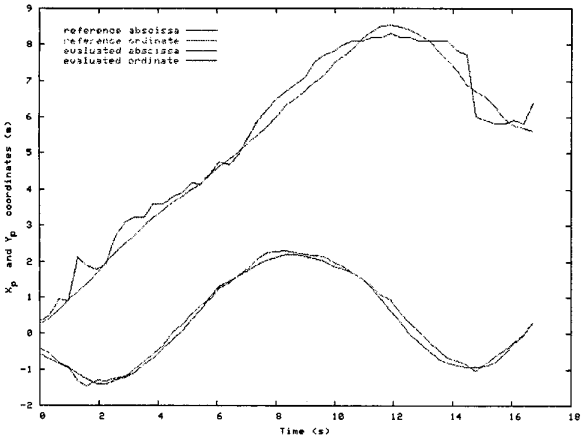
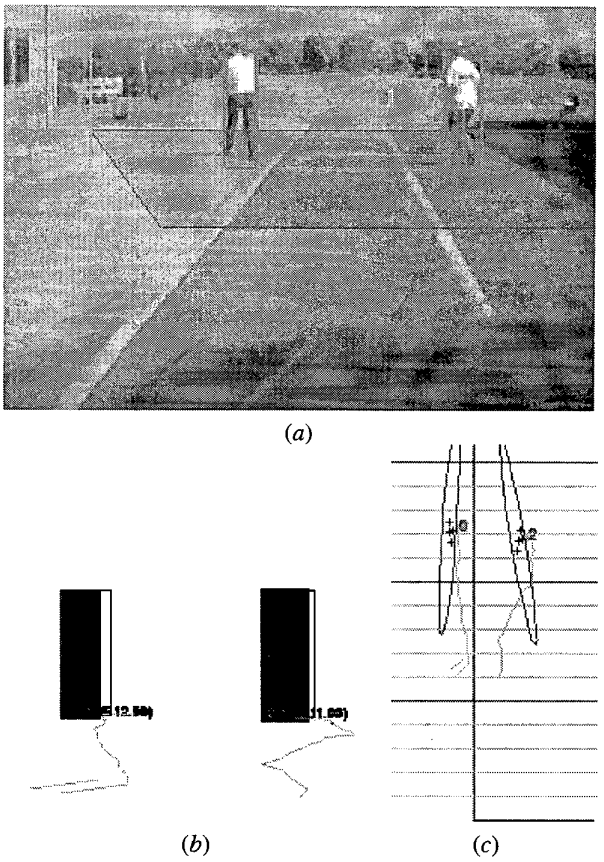


Fig. 8. Temporal comparison example of ground plane coordinates between the imposed trajectory and the evaluated pedestrian path.

observed during the system activity on the indoor prerecorded image sequences, have proved a high reliability of the new approach. It has been possible therefore to adopt the new tracker module for the outdoor vehicular system activity and the multi trace results so obtained are illustrated in fig. 9. Integration of observations obtained from other different types of sensors can be easily achieved with the current system structure and can lead to more significant results in the form of the *data fusion* paradigm.

VI. ACKNOWLEDGMENTS

The authors gratefully thank Dr. C. Adams and Dr. M. Del Rose from U. S. Army TACOM and Dr. S. Sampath from USARDSG-UK for their support in the research, as also the Italian - French Galileo program for cultural exchange and researchers mobility.



(a) (b) (c)

Fig. 9. Outdoor vehicular stereo results in *multi tracking* mode: (*a*) the vision algorithm recognizes the pedestrians (the stereo localization area is shown in transparent blue on the ground); (*b*) perspective view of the results, trajectories provided by the tracker are also shown; (*c*) pedestrian trajectories and error ellipsis of the current estimated pedestrian positions are represented on the road plane and marked with the corresponding pedestrian `id` (there is no correspondence between the reference grid of the graphical representation that represents the camera reference system and the grid painted on the asphalt).

REFERENCES

[1] C. Curio, J. Edelbrunner, T. Kalinke, C. Tzomakas, and W. von Seelen, "Walking Pedestrian Recognition," *IEEE Trans. on Intelligent Transportation Systems*, vol. 1, no. 3, pp. 155–163, Sept. 2000.

[2] D. M. Gavrila, "Pedestrian Detection from a Moving Vehicle," in *Procs. of European Conference on Computer Vision*, vol. 2, June–July 2000, pp. 37–49.

[3] C. Papageorgiou, T. Evgeniou, and T. Poggio, "A Trainable Pedestrian Detection System," in *Procs. IEEE Intelligent Vehicles Symposium'98*, Stuttgart, Germany, Oct. 1998, pp. 241–246.

[4] A. Broggi, M. D. Rose, A. Fascioli, I. Fedriga, and A. Tibaldi, "Stereo-based Preprocessing for Human Shape Localization in Unstructured Environments," in *Procs. IEEE Intelligent Vehicles Symposium 2003*, Columbus, USA, June 2003, pp. 410–415.

[5] A. Broggi, A. Fascioli, M. Carletti, T. Graf, and M. Meinecke, "A Multi-resolution Approach for Infrared Vision-based Pedestrian Detection," in *Procs. IEEE Intelligent Vehicles Symposium 2004*, Parma, Italy, June 2004, in press.

[6] D. M. Gavrila and J. Geibel, "Shape-Based Pedestrian Detection and Tracking," in *Procs. IEEE Intelligent Vehicles Symposium 2002*, Paris, France, June 2002.

[7] H. Elzein, S. Lakshmanan, and P. Watta, "A Motion and Shape-Based Pedestrian Detection Algorithm," in *Procs. IEEE Intelligent Vehicles Symposium 2003*, Columbus, USA, June 2003, pp. 500–504.

[8] H. Nanda, C. Benabdelkedar, and L. Davis, "Modelling Pedestrian Shapes for Outlier Detection: a Neural Net based Approach," in *Procs. IEEE Intelligent Vehicles Symposium 2003*, Columbus, USA, June 2003, pp. 428–433.

[9] M. Bertozzi, A. Broggi, R. Chapuis, F. Chausse, A. Fascioli, and A. Tibaldi, "Shape-based pedestrian detection and localization," in *Procs. IEEE Intl. Conf. on Intelligent Transportation Systems 2003*, Shangai, China, Oct. 2003, pp. 328–333.

[10] R. E. Kalman, "A New Approach to Linear Filtering and Prediction Problems," *Trans. ASME Journal of Basic Engineering*, vol. 82, no. 1, pp. 35–45, Mar. 1960.

[11] D. Koller, K. Daniilidis, and H. Nagel, "Model-based object tracking in monocular image sequences of road traffic scene," *International Journal of Computer Vision*, vol. 3, no. 10, pp. 257–281, 1993.

[12] R. Aufrère, F. Marmoiton, R. Chapuis, J. P. Derutin, and F. Collange, "Road detection and vehicle tracking by vision for Adaptative Cruise Control," *International Journal of Robotics Research*, vol. 20, no. 4, pp. 267–286, April 2001.

[13] D. Reisfeld, H. Wolfson, and Y. Yeshurun, "Context Free Attentional Operators: the Generalized Symmetry Transform," *Intl. Journal of Computer Vision, Special Issue on Qualitative Vision*, vol. 14, pp. 119–130, 1994.