



# SAVE-U

## ***SENSORS AND SYSTEM ARCHITECTURE FOR VULNERABLE ROAD USERS PROTECTION***



Information Society Technologies: Systems and Services for the Citizen

### Project IST – 2001 – 34040

Project funded by the European Community under  
the "Information Society Technology" Programme (1998-2002)

## **Deliverable D9: Evaluation of the embedded image processing platform and associated sensors**

Authors	D M. Gavrilă (DaimlerChrysler) S. Munder (DaimlerChrysler)
Co-Author	M. Vinciguerra (CEA LIST)
Dissemination Level:	PU
Version:	3.0
Contract Date:	1/3/2002
Document Date:	19/11/2004

Partners: FAURECIA Industries (FR), Siemens VDO Automotive AG (DE),  
MIRA Ltd (GB), Commissariat à l'Énergie Atomique (FR),  
DaimlerChrysler (DE), Volkswagen AG (DE)

Commission of the European Communities – Directorate-General Information Society

## *Revision chart and history log*

Version	Date	Reason
1.0	06.09.2004	First draft from DC to partners for feedback
2.0	19.11.2004	Second draft from DC to partners for feedback
3.0	19.11.2004	Third draft from DC to partners for feedback

## Credits

This document was prepared by partners of the SAVE-U consortium. Please contact one of the following project partners for further information regarding this document or the project.

**Commissariat à l’Energie Atomique**

**DaimlerChrysler AG**

**Faurecia Industries**

**MIRA Ltd**

**SiemensVDO Automotive AG**

**Volkswagen AG**

Work Package Leader:

Name:	Mr. Dariu M. GAVRILA
Company:	DaimlerChrysler AG
Phone:	+49 731 505 2388
Fax:	+49 731 505 2388
E-mail:	dariu.gavrila@DaimlerChrysler.com

Project Coordinator:

Name:	Mr. Philippe MARCHAL
Company:	FAURECIA Industries
Phone:	+33 (0)3 81 36 48 10
Fax:	+33 (0)3 81 36 45 20
E-mail:	pmarchal@seloncourt.faurecia.com

## 1. Table of Contents

<b>REVISION CHART AND HISTORY LOG.....</b>	<b>2</b>
<b>CREDITS.....</b>	<b>3</b>
<b>1. TABLE OF CONTENTS .....</b>	<b>4</b>
<b>2. EXECUTIVE SUMMARY .....</b>	<b>5</b>
<b>3. INTRODUCTION .....</b>	<b>7</b>
<b>4. TEST METHODOLOGY .....</b>	<b>8</b>
<b>5. DATA SETS .....</b>	<b>11</b>
5.1 Configuration A .....	12
5.2 Configuration B .....	12
<b>6. RESULTS .....</b>	<b>13</b>
6.1 Sensor Configuration A.....	13
6.2 Sensor Configuration B.....	19
<b>7. DISCUSSION .....</b>	<b>21</b>
<b>8. CONCLUSIONS .....</b>	<b>24</b>
<b>9. REFERENCES .....</b>	<b>25</b>

## 2. Executive Summary

This Deliverable describes the validation of the SAVE-U vision sub-system developed in Work Package 6. Two sensor configurations were considered: Configuration A, the focus of the project, consisted of a FIR and color camera and, for comparison, Configuration B consisted of a grayscale stereo vision system, an (in SAVE-U) improved version of a pedestrian system previously used in EU Project PROTECTOR. For a description of the respective hardware/software architecture, see Deliverable 10. For a description of the corresponding algorithms, see Deliverable 11.

This document first describes the validation methodology used. At the core, the aim was to compare entries from ground truth and from system output, related to 3D object (pedestrian) position relative to the vehicle. This involved a human operator diligently label objects in monocular images and using scene knowledge to back-project into 3D. For the case of pedestrians, the latter means making the “flat world” assumption coupled with the reasonable conjecture that the pedestrian feet stand on the ground plane. We state the specifics of the sensor coverage area (depth range 5-25m, lateral range 1.4-4m on each side of vehicle), the localization tolerance used (10% and 30% error in pedestrian localization laterally and in depth), and ways how multiple matches between system output and ground truth data were handled (“group rule”).

Thereafter, we provide a description of the two image data sets used for validation. Both data sets are quite large involving several hundreds, if not thousands of pedestrians in real urban traffic, allowing for a meaningful validation of the two configurations.

To illustrate the strengths and remaining weaknesses of the chosen approaches, we provide a variety of results obtained with Configuration A. They involve cases where pedestrians were successfully recognized, as well as recognition failures. The latter can be subdivided in two classes: pedestrians who were not recognized and pedestrians who were found when there were none (so-called “false positives”). As described in Deliverable 11, Configuration A consists conceptually of three parts: a detection component which identifies probable pedestrian instances in images, a classification component which verifies these candidate solutions, and finally, a tracking component which integrates results over time. We provide illustrations of the results obtained with these components. Finally, we provide “hard” quantitative

performance data on the entire data set. This involves listing the correct recognition vs. the false positive rate; we distinguish both per-frame and per-trajectory cases (i.e. before/after tracking). A similar analysis is performed for Configuration B.

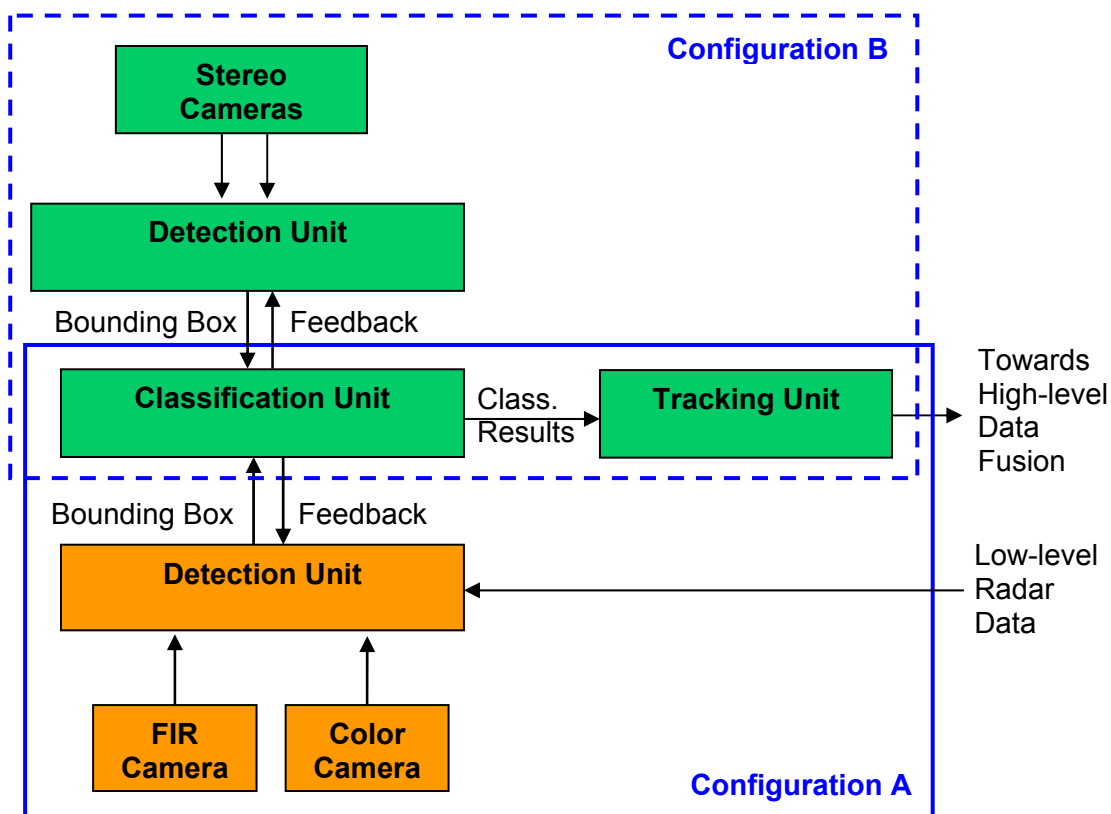
Experimental results show that Configurations A and B have roughly the same performance. Configuration A has a correct recognition rate of 74-95% with a false positive rate of 1 each 17.7-19.7s. Configuration B scores a correct recognition rate of 78-88% with a false positive rate of 1 each 17s. Both configurations comply with the boundary conditions set within the SAVE-U project listed in Deliverable 7 as to sensor coverage area, vehicle speed, processing rate, etc. Related to recognition performance, we note that obtained results are *above* the targeted 70% correct recognition performance for the overall SAVE-U system. The false positive rates are still up to a factor of 3 *below* the targeted values, but this performance gap is relatively minor and can be expected to be closed by the addition of the SAVE-U radar component and sensor fusion (Work Packages 8 and 9, a main focus within SAVE-U) while still keeping the correct recognition rate above 70%.

We conclude that extensive testing in realistic traffic situations has proved the vision component successful in bringing recognition performance within reach of specified overall SAVE-U targets.

Our final comment is to emphasize that an integrated SAVE-U application will in practice have significantly lower false “alarms” (i.e. final system reactions such as driver warning or vehicle braking) than the false positive rates mentioned above. Above performance numbers were obtained by treating all pedestrian trajectories within the sensor coverage area equally, i.e. making no distinction between the case of a pedestrian briefly entering the coverage area at 25m, 4m aside of the vehicle and the case of the pedestrian right in front of the vehicle at 5m). An integrated SAVE-U application will involve an additional Collision Risk Assessment module which will filter out any “non-risky” pedestrian trajectories. Focusing only on “risky” pedestrian trajectories (defined here as within 1.5m lateral offset to the vehicle medial axis) results in correct recognition rates as high as 100% and false “alarms” as low as 1 in 5 min (Configuration B), *before* sensor fusion with radar.

### 3. Introduction

This Deliverable deals with the validation of the SAVE-U vision sub-system. Two sensor configurations were considered, see Figure below. Configuration A, the focus of the project, consisted of a FIR and color camera (solid box). For comparison, Configuration B consisted of a grayscale stereo vision system (dotted box), an improved version of a pedestrian system previously used in EU Project PROTECTOR. For a description of the respective hardware/software architecture, see Deliverable 10. For a description of the corresponding algorithms, see Deliverable 11. In the Figure below, the orange diagram blocks were developed by project partner CEA, whereas the green diagram blocks were covered by project partner DaimlerChrysler.



**Figure 1. The two configurations for the vision sub-component used in SAVE-U: Configuration A involves mono color-video and FIR, Configuration B involves stereo vision (monochrome)**

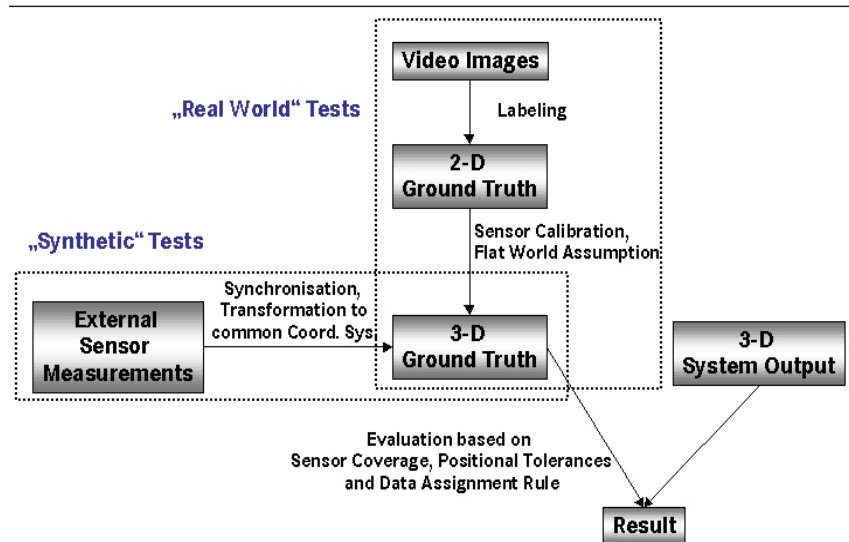
The outline of this Deliverable is as follows. Section 4 describes the test methodology used to evaluate both (vision sub-system) Configurations A and B. A description of

the image data sets used for validation is given in Section 5. The corresponding results are given in Section 6. Section 7 places the obtained results in context by comparing them to published quantitative data in the literature (i.e. state-of-the-art). Finally, Section 8 presents the conclusions.

## 4. Test Methodology

The followed test methodology is illustrated in Figure 2. At the core, our aim is to compare entries from ground truth and from system output, related to 3D object position relative to the vehicle (we prefer to evaluate the system in 3D rather than in image space, because application specific considerations can more easily be described in 3D terms). There are two possibilities for obtaining 3D ground truth data. The first involves designing a test set-up where, by means of auxiliary measurement equipment, vehicle- and object position over time is determined in a world coordinate system. Synchronization and transformation into the vehicle coordinate system leads to the desired 3D ground truth data. This procedure will be followed later in the project as part of the MIRA WP11 experiments with the test rig and on the test track. For the case that controlled environment is not available, the alternative is for a human operator to diligently label objects in monocular images and using scene knowledge to back-project into 3D. In case of pedestrians, the latter means making the “flat world” assumption coupled with the reasonable conjecture that the pedestrian feet stand on the ground plane. This option had to be taken for the Work Package (WP) 6 tests in real traffic.

Table 1 lists the terminology used in the remainder of this paper. We consider established performance ratios such as sensitivity and precision. Evaluation is made on the frame- and trajectory-level. For the latter, we further distinguish two types of trajectories: “class-B” and the higher quality “class-A” trajectories that have at least one entry or at least 50% of their entries matched, respectively. When comparing ground truth and system entries, the following items need to be specified.



**Figure 2. Test methodology.** Ground truth data is determined either by auxiliary measurement equipment or by manually labeling images. This data is then compared with system output.

**Sensor Coverage Area.** The sensor coverage area represents the space surrounding the vehicle where the defined object detection capability is required. Outside the sensor coverage area, we consider detection capability optional in the sense that the system is not rewarded or penalized for correct/false/missing detections. The SAVE-U vision sensor coverage area for Configuration A and B is shown in Figure 3.

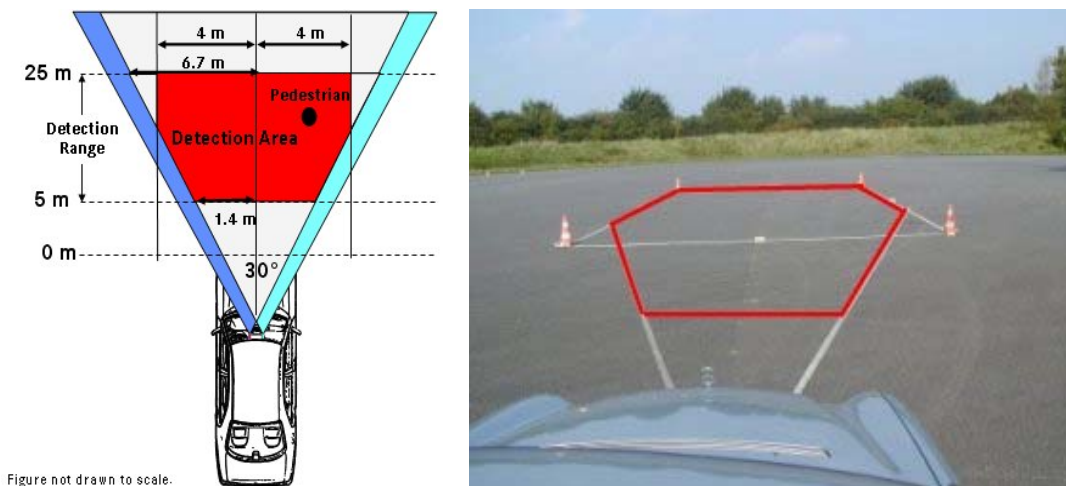


Figure not drawn to scale.

**Figure 3. SAVE-U coverage area**

**Localization Tolerance.** Given an object detected by the system at a certain location (“alarm”), and given a true object location (“event”), the localization tolerance is the maximum positional deviation that still allows us to count the alarm as a match.

This localization tolerance is the sum of an application-specific component (how precise does the object localization have to be for the application) and a component related to measurement error (how exact can we determine true object location). For the SAVE-U WP6 tests, we define object localization tolerance as percentage of distance, for lateral and longitudinal direction (X and Z), with respect to the vehicle. Regarding the application-specific component, values of  $X_a = 5\%$  and  $Z_a = 15\%$  appear reasonable; for example, this means that, at 20m distance, we tolerate a localization error of  $\pm 1\text{m}$  and  $\pm 3\text{m}$  in the position of the pedestrian, lateral and longitudinal to the vehicle driving direction, respectively. Regarding the measurement-specific component,  $X_m = 5\%$  and  $Z_m = 15\%$  appear necessary (with the larger  $Z_m$  value to account for non-flat road surface and/or vehicle pitch in case of ground truth by monocular image labelling). For the SAVE-U WP6 tests, we then use overall tolerances of  $X = 10\%$  and  $Z = 30\%$ .

**Data Assignment (“Group Rule”).** For the SAVE-U application we allow many-to-many correspondences. An event is considered matched if there is at least one alarm matching it. In practice, this means that in the case a group of pedestrians walking sufficiently close together in front of the vehicle, the system does not necessarily have to detect all of them in isolation; it suffices if each true pedestrian is within the localization tolerance of a detected pedestrian.

Finally, having established rules for matching ground truth and system entries, we need to specify what statistics to collect to describe detection performance. We consider performance at two levels, at individual frame level and at the trajectory level. Among the latter, we distinguish two trajectory types: “class-B” and the higher quality “class-A” trajectories that have at least one entry or at least 50% of their entries matched, respectively. We consider established performance ratios such as sensitivity and precision. See Table I for the terminology used in the remainder of this Deliverable.

<b>event</b>	a ground truth object, i.e. the true position of a pedestrian at time $t$ (frame-level)
<b>recognition</b>	a system output, i.e. a recognition made by the system at time $t$ , usually after processing a single video frame
<b>required event</b>	an event within the defined sensor coverage area and of higher visibility than specified as system minimum.
<b>optional event</b>	an event outside the defined sensor coverage area or of lower visibility than specified as system minimum
<b>good event</b>	a required event with at least one matching detection
<b>correct recognition</b>	a recognition with at least one matching event (either required or optional)
<b>false recognition, false positive</b>	a recognition with no matching event (either required or optional)
<b>event trajectory</b>	a sequence of events, i.e. the true locations of a pedestrian at each point in time
<b>required event trajectory</b>	an event trajectory with at least one required event
<b>recognized trajectory</b>	a sequence of recognitions tracked by the system
<b>class-B event/recognized trajectory</b>	an event/detected trajectory with at least one good event/correct recognition
<b>class-A event/recognized trajectory</b>	an event/detected trajectory with at least 50% good events/correct recognitions
<b>frame-level sensitivity</b>	the number of good events divided by the total number of required events
<b>frame-level precision</b>	the number of correct recognitions divided by the total number of recognition
<b>class-A/-B trajectory sensitivity</b>	the number of class-A/-B event trajectories divided by the number of required event trajectories
<b>class-A/-B trajectory precision</b>	the number of class-A/-B recognized trajectories divided by the number of all recognized trajectories

**Table 1. Terminology used for the validation of the SAVE-U system**

## 5. Data Sets

Both Configurations A and B were tested on large image databases derived from real urban traffic scenes. The data sets are quite large involving several hundreds, if not thousands of pedestrians in real urban traffic, allowing for a meaningful validation of the two configurations.

---

## 5.1 Configuration A

---

	<i>Training set</i>	<i>Test set</i>
Sequences	3	7
Images	2188	2727
Duration	4 min 11s	2 min 57s
VRU trajectories (all)	20	38
VRU instances (all)	4076	1134
VRU trajectories (“risky”)		17
VRU instances (“risky”)		399

**Table 2: Training and test data set statistics. The training sequences were used to optimize the system parameters, while the performance was evaluated on the test set.**

---

## 5.2 Configuration B

---

	<i>Training set</i>	<i>Test set</i>
Sequences	1	1
Images	21239	17390
Duration	27min	24min
VRU trajectories (all)	28	32
VRU instances (all)	467	353
VRU trajectories (“risky”)	13	10
VRU instances (“risky”)	71	47

**Table 3: Configuration B training and test sets.**

---

## 6. Results

---

To illustrate the strengths and remaining weaknesses of the chosen approaches, we first provide a variety of example results obtained with Configurations A and B. They involve cases where pedestrians were successfully recognized, as well as recognition failures. The latter can be subdivided in two classes: pedestrians who were not recognized and pedestrians who were found when there were none (so-called “false positives”).

Subsequently, we list the obtained quantitative performance statistics on the respective data sets (see previous Section). This involves listing the correct recognition vs. the false positive rate; we distinguish both per-frame and per-trajectory cases (i.e. before/after tracking). The per-trajectory cases are further subdivided in the “class-B” and “class-A” cases, recall discussion in Section 4.

A separate evaluation was made for pedestrians directly in front of the car, i.e., which are in particular risk, by restricting the sensor coverage area to a maximum lateral offset from the vehicle medial axis of 1.5 m (instead of 4 m).

We recall from Deliverable 7 that for the purpose of performance evaluation, we only considered the scenario of a vehicle driving on an approximately straight road (i.e. no turn at intersection) and the cases where pedestrians are at least 90% visible (i.e. cases where only pedestrian head is visible above parked car is outside scope).

---

### 6.1 Sensor Configuration A

---

As described in Deliverable 11 and seen in Figure 1, Configuration A consists conceptually of three parts: a detection component which identifies probable pedestrian instances in images, a classification component which verifies these candidate solutions, and finally, a tracking component which integrates results over time. Figure 4 - Figure 10 provide some illustrative examples of the functioning of the detection and classification component. Output of the detection component is indicated by bounding boxes superimposed onto the right images (FIR modality). The boxes are coloured “red” or “green” based on whether the classification component accepted or rejected the candidate solutions, respectively. The classification result is furthermore superimposed on the left images (colour modality). Note the accurate registration between colour and FIR images.

Figure 4 - Figure 6 illustrate correct operation of the system at various distances of the pedestrian to the vehicle. As can be seen, the detection component is quite able to restrict image content to a few candidate locations (bounding boxes). Furthermore, the classification component successfully distinguishes pedestrian vs. non-pedestrian candidate regions. Figure 7 and Figure 8 illustrate examples of false positives: non-pedestrian regions incorrectly classified as pedestrians. Typically, these involve heat regions with pronounced left/right vertical borders. Figure 9 and Figure 10 illustrate the other system fault mode: pedestrians not recognized by the system. Figure 9 involves a missed detection and is a clear example where radar-based detection could help. Figure 10 involves a missed classification.



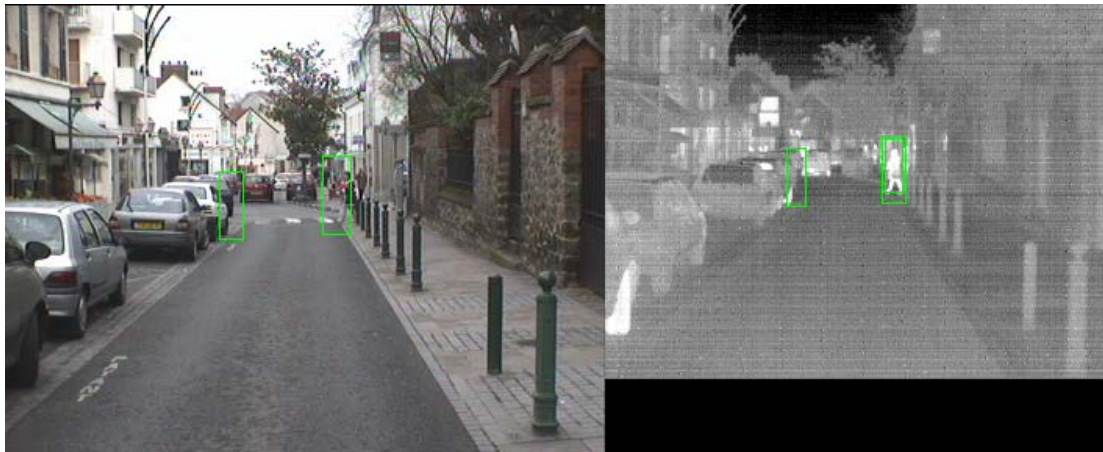
**Figure 4:** Example of a VRU correctly recognized by the vision system. The right sub-image shows the classification results superimposed on the FIR image. Green boxes denote ROIs classified as VRUs, ROIs rejected as non-VRUs are shown in red. The left sub-image displays the output of the tracking module superimposed on the color image.



**Figure 5:** Same as Figure 4. In addition to the correctly recognized VRU, a false ROI from the detection module was correctly rejected by the classifier.



**Figure 6:** Same as Figure 5: Correct recognition plus two correct rejections by the classifier.



**Figure 7:** Example of a false positive: The side of a vehicle was erroneously classified as a VRU.



**Figure 8:** Another example of a false positive: Although the image ROI contains a VRU, the position of the box is incorrect.

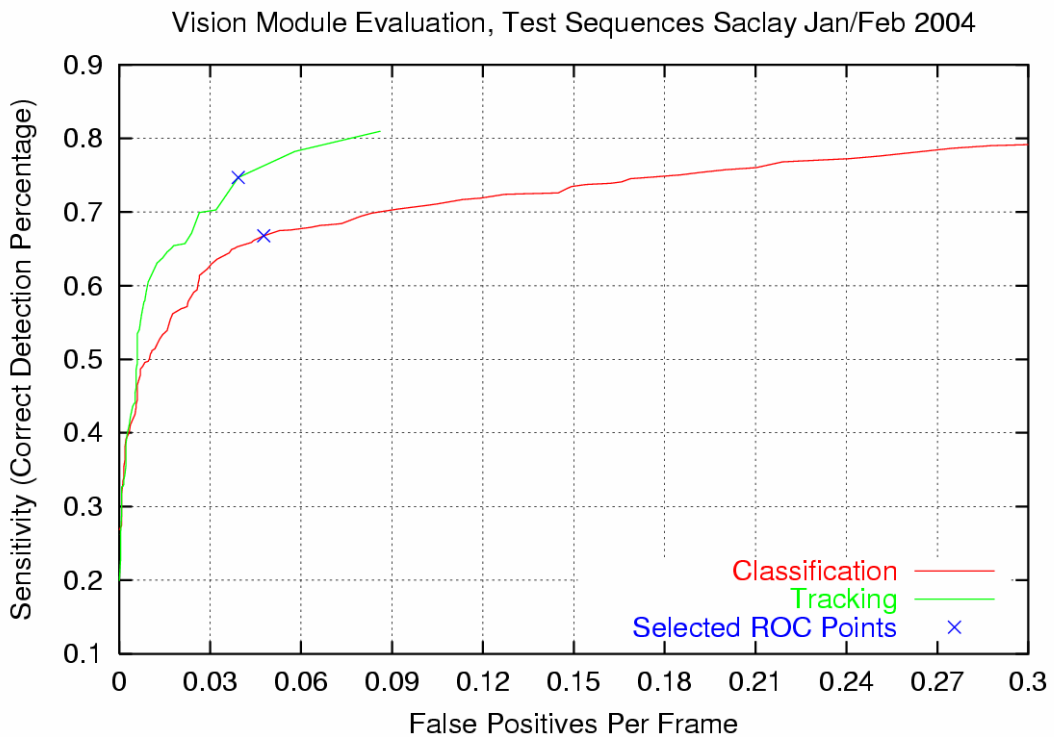


**Figure 9:** Example of a missing recognition: No ROI generated by the detection module.



**Figure 10: Another example of a missing recognition: The second more distant pedestrian was misclassified as a non-VRU.**

The trade-off between correct recognition and false positives is determined by the values of relevant system parameters and is illustrated by the so-called Receiver Operating Characteristic (ROC) curve. Figure 11 shows the ROC curve for Configuration A related to single-frame (red line) and multi-frame processing (green line), i.e. the latter after tracking. The parameter settings associated with the two ROC points shown as blue crosses, resulted in a system performance as indicated in Table 4.



**Figure 11. Receiver Operating Characteristic (ROC) Curve for single-frame processing (red line) and multi-frame processing (green line) i.e. after tracking, for SAVE-U Configuration A.**

	<b>Frame-based</b>	<b>Trajectory-based (Class A)</b>	<b>Trajectory-based (Class B)</b>
<b>Sensitivity (all)</b>	74.7%	73.7%	94.7%
<b>Precision (all)</b>	92.4%	84.5%	86.6%
<b>False Recognition per Frame / Minute (all)</b>	0.039 (1 every 25 frames)	5.8 (1 every 10.4s)	5.8 (1 every 13.6s)
<b>Sensitivity (“risky”)</b>	96.0%	94.1%	94.1%
<b>Precision (“risky”)</b>	92.2%	79.2%	81.2%
<b>False Recognition per Frame / Minute (“risky”)</b>	0.025 (1 every 40 frames)	3.4 (1 every 17.7s)	3.1 (1 every 19.7s)

**Table 4: SAVE-U Configuration A. Performance per frame and per trajectory (A-Class and B-Class). The number of false recognitions is given per frame for frame-level performance and per driving minute for trajectory performance.**

It is interesting to examine at the chosen ROC point (i.e. system parameter setting resulting in Table 4) what the contributions are of the three sub-components: detection, classification and tracking. The successive improvement one obtains is nicely shown in the Table below.

	<b>False Positives (Per Frame)</b>	<b>Sensitivity (Per Frame)</b>	<b>Precision (Per Frame)</b>
<b>After Detection</b>	2.37	85.9%	39.0%
<b>After Classification</b>	0.048	66.7%	92.4%
<b>After Tracking</b>	0.039	74.7%	92.4%

**Table 5. Successive recognition performance improvement obtained by the three logical components of SAVE-U configuration A: detection, classification and tracking.**

## 6.2 Sensor Configuration B

SAVE-U Configuration B also consists conceptually of three parts: a (stereo-based) detection component which identifies obstacles in front of the vehicle, a classification component which verifies if these obstacle regions involve pedestrians, and finally, a tracking component which integrates results over time, see [Gavrila2004]. Figure 12- Figure 13 provides some illustrative output examples of this Configuration. The grey bounding boxes in the left windows are the output of stereo-based obstacle detection and denote regions where the classification module is activated. System output (i.e. after classification and tracking) is shown in the middle and right windows. The middle windows show recognized pedestrians superimposed on one of the two stereo images, whereas the right windows show a top view of the situation in front of the vehicle. Currently and previously detected pedestrians are indicated in the top view by red and blue dots, respectively. See for example the correctly recognized pedestrian trajectory in the right window of Figure 12: the trajectory is initiated soon after the pedestrian enters the coverage range at 25m. Figure 13 shows an instance of a false positive,



**Figure 12: Output of SAVE-U Configuration B. Left window shows results of stereo-based object detection (grey bounding boxes). Middle window shows correctly recognized pedestrian (red shape). Right window shows a top view of situation in front of vehicle: pedestrian at current/previous time instances shown by red/blue dot. Right hand scale is given in meters.**

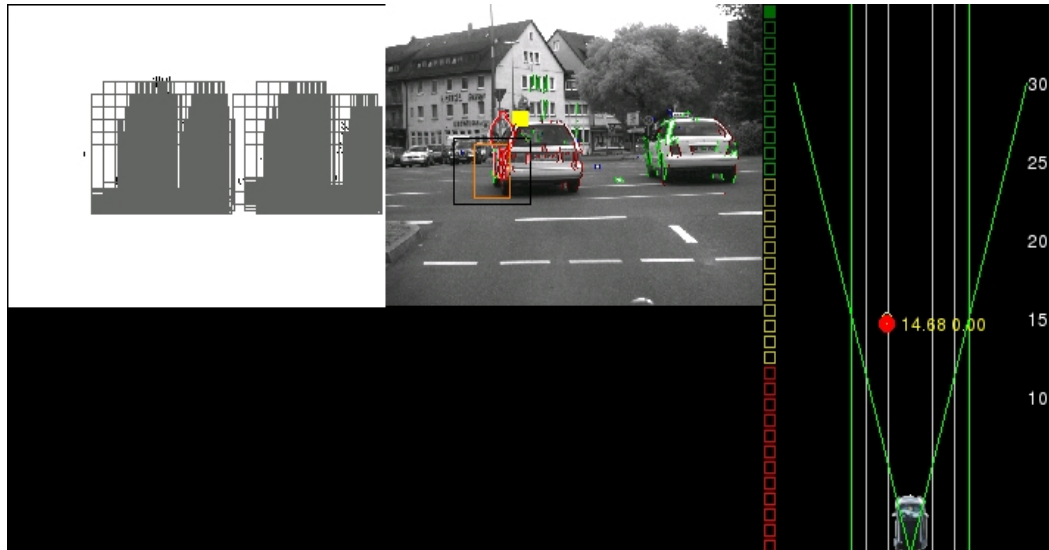


Figure 13. Same as Figure 12, but here with example of false positive on side of vehicle.

	Frame-based	Trajectory-based (Class A)	Trajectory-based (Class B)
<b>Sensitivity (all)</b>	61.4%	78.1%	87.5%
<b>Precision (all)</b>	42.5%	34.8%	35.6%
<b>False Recognition per Frame / Minute (all)</b>	0.031	3.58	3.54
<b>Sensitivity (“risky”)</b>	76.6%	100%	100%
<b>Precision (“risky”)</b>	73.7%	75.0%	75.0%
<b>False Recognition per Frame / Minute (“risky”)</b>	0.001	0.21	0.21

Table 6: SAVE-U Configuration B. Performance per frame and per trajectory (A-Class and B-Class), optimized under the run time constraint of average processing of max 100ms/frame. The number of false recognitions is given per frame for frame-level performance and per driving minute for trajectory performance.

Average processing time was 10 Hz over all images of data set, when considering only images which contained one or more pedestrian, processing rate decreased to 7-9 Hz. All figures on a 3.2 GHz Pentium IV processor (standard PC).

---

## 7. Discussion

---

Let us put the obtained performance of the SAVE-U vision component into context by comparing it to

- the performance of the previous state-of-the-art vision stereo vision system by DaimlerChrysler from the 5-th Framework EU Project PROTECTOR (see [Gavrila2004]), and
- the performance target for the *overall* SAVE-U system (i.e. including the radar component) as described in Deliverable 7, “Specifications of SAVE-U sensor platform and ECUs”. Recall that the performance target was rather ambitious: compared to the predecessor PROTECTOR system it involves an enlargement of the sensor coverage area, an increase in vehicle speed combined with an considerable increase in correct detection rate, at the same time as the false detection rate is decreased by *one order of magnitude*.

See Table 7. The comparison between SAVE-U Configuration B and EU PROTECTOR system is particularly revealing since both involve the identical test set.

	<b>Protector DC System</b>	<b>SAVE-U Config. A</b>	<b>SAVE-U Config. B</b>	<b>SAVE-U Target (D7)</b>
<b>Date</b>	March 2003	Nov. 2004	Nov. 2004	2005
<b>Range</b>	10-25m	5-25m	5-25m	5-25m
<b>Vehicle Speed</b>	up to 30 km/h	up to 40 km/h	up to 40 km/h	up to 40 km/h
<b>Correct Recognition<sup>1</sup></b>	28 - 45%	74 - 95%	78 - 88%	>70%
<b>False Recognition<sup>1</sup></b>	each 3-6s	each 17.7-19.7s	each 17s	each 30-60s
<b>Processing Rate</b>	4-7 Hz	9-11 Hz <sup>2</sup>	7-9 Hz <sup>2</sup>	5 Hz

**Table 7. Performance comparison between previous state-of-the-art PROTECTOR system, SAVE-U Configurations A and B (vision component only) and SAVE-U target (vision and radar combined).**

Table 7 shows that Configurations A and B have roughly the same performance. Configuration A has a correct recognition rate of 74-95% with a false positive rate of 1 each 17.7-19.7s. Configuration B scores a correct recognition rate of 78-88% with a false positive rate of 1 each 17s. Both configurations comply with the boundary conditions set within the SAVE-U project listed in Deliverable 7 as to sensor coverage area, vehicle speed, processing rate, etc. Related to recognition performance, we note that obtained results are *above* the targeted 70% correct recognition performance for the overall SAVE-U system. The false positive rates are still up to a factor of 3 *below* the targeted values, but this performance gap is relatively minor and can be expected to be closed by the addition of the SAVE-U radar component and sensor fusion (Work Packages 8 and 9, a main focus within SAVE-U) while still keeping the correct recognition rate above 70%.

Note that an integrated SAVE-U application will in practice have significantly lower false “alarms” (i.e. final system reactions such as driver warning or vehicle braking)

<sup>1</sup> Lower and higher bounds refer to class-A and class-B trajectory performance, respectively.

<sup>2</sup> For scenes containing at least one pedestrian in sensor coverage area. Higher processing rates for scenes not containing pedestrians.

than the false positive rates mentioned above. Above performance numbers were obtained by treating all pedestrian trajectories within the sensor coverage area equally, i.e. making no distinction between the case of a pedestrian briefly entering the coverage area at 25m, 4m aside of the vehicle and the case of the pedestrian right in front of the vehicle at 5m). An integrated SAVE-U application will involve an additional Collision Risk Assessment module which will filter out any “non-risky” pedestrian trajectories. Focusing only on “risky” pedestrian trajectories (defined here as within 1.5m lateral offset to the vehicle medial axis) results in correct recognition rates as high as 100% and false “alarms” as low as 1 in 5min (Configuration B), *before* sensor fusion with radar.

## 8. Conclusions

This document first described the validation methodology used for the vision component of the SAVE-U system (Work Package 6). At the core, the aim was to compare entries from ground truth and from system output, related to 3D object (pedestrian) position relative to the vehicle. This involved a human operator diligently label objects in monocular images and using scene knowledge to back-project into 3D. For the case of pedestrians, the latter means making the “flat world” assumption coupled with the reasonable conjecture that the pedestrian feet stand on the ground plane. We stated the specifics of the sensor coverage area (depth range 5-25m, lateral range 1.4-4m on each side of vehicle), the localization tolerance used (10% and 30% error in pedestrian localization laterally and in depth), and ways how multiple matches between system output and ground truth data were handled (“group rule”).

Thereafter, we provided a description of the two image data sets used for validation. Both data sets are quite large involving several hundreds, if not thousands of pedestrians in real urban traffic, allowing for a meaningful validation of the two configurations.

To illustrate the strengths and remaining weaknesses of the chosen approaches, we provided a variety of results obtained with Configuration A. They involve cases where pedestrians were successfully recognized, as well as recognition failures. As described in Deliverable 11, Configuration A consists conceptually of three parts: a detection component which identifies probable pedestrian instances in images, a classification component which verifies these candidate solutions, and finally, a tracking component which integrates results over time. We provided illustrations of the results obtained with these components. Finally, we provided “hard” quantitative performance data on the entire data set. This involves listing the correct recognition vs. the false positive rate; we distinguish both per-frame and per-trajectory cases (i.e. before/after tracking). A similar analysis was performed for Configuration B.

Experimental results showed that Configurations A and B have roughly the same performance. Configuration A has a correct recognition rate of 74-95% with a false positive rate of 1 each 17.7-19.7s. Configuration B scores a correct recognition rate of 78-88% with a false positive rate of 1 each 17s. Both configurations comply with the boundary conditions set within the SAVE-U project listed in Deliverable 7 as to

sensor coverage area, vehicle speed, processing rate, etc. Related to recognition performance, we note that obtained results are *above* the targeted 70% correct recognition performance for the overall SAVE-U system. The false positive rates are still up to a factor of 3 *below* the targeted values, but this performance gap is relatively minor and can be expected to be closed by the addition of the SAVE-U radar component and sensor fusion (Work Packages 8 and 9, a main focus within SAVE-U) while still keeping the correct recognition rate above 70%.

We conclude that extensive testing in realistic traffic situations has proved the vision component successful in bringing recognition performance within reach of specified overall SAVE-U targets.

Our final comment is to emphasize that an integrated SAVE-U application will in practice have significantly lower false “alarms” (i.e. final system reactions such as driver warning or vehicle braking) than the false positive rates mentioned above. Above performance numbers were obtained by treating all pedestrian trajectories within the sensor coverage area equally, i.e. making no distinction between the case of a pedestrian briefly entering the coverage area at 25m, 4m aside of the vehicle and the case of the pedestrian right in front of the vehicle at 5m). An integrated SAVE-U application will involve an additional Collision Risk Assessment module which will filter out any “non-risky” pedestrian trajectories. Focusing only on “risky” pedestrian trajectories (defined here as within 1.5m lateral offset to the vehicle medial axis) results in correct recognition rates as high as 100% and false “alarms” as low as 1 in 5 min (Configuration B), *before* sensor fusion with radar.

---

## 9. References

---

[Gavrila2004] D. M. Gavrila, J. Giebel and S. Munder, „Vision-based Pedestrian Detection: the PROTECTOR+ System“, *Proc. of the IEEE Intelligent Vehicles Symposium*, Parma, Italy, 2004.

Partners owning:

Partners contributed:

DaimlerChrysler

Commissariat à l'Energie Atomique

Made available to:

Commission, Consortium