2015 IEEE 18th International Conference on Intelligent Transportation Systems

Assistive pedestrian crossings by means of stereo localization and RFID anonymous disability identification

D. F. Llorca, R. Quintero, I. Parra, R. Izquierdo, C. Fernández, M. A. Sotelo

***Abstract*— Assistive technology usually refers to *systems used to increase, maintain, or improve functional capabilities of individuals with disabilities*. This idea is here extended to transportation infrastructures, using pedestrian crossings as a specific case study. We define an Assistive Pedestrian Crossing as a *pedestrian crossing able to interact with users with disabilities and provide an adaptive response to increase, maintain or improve their functional capabilities while crossing*. Thus, the infrastructure should be able to locate the pedestrians with special needs as well as to identify their specific disability. In this paper, user location is obtained by means of a stereo-based pedestrian detection system. Disability identification is proposed by means of a RFID-based anonymous procedure from which pedestrians are only required to wear a portable and passive RFID tag. Global nearest neighbor is applied to solve data association between stereo targets and RFID measurements. The proposed assistive technology is validated in a real crosswalk, including different complex scenarios with multiple RFID tags.**

***Index Terms*— Assistive technology, pedestrian crossings; RFID identification and location; stereo-based pedestrian detection; RFID and stereo association.**

I. INTRODUCTION

The commitment of the European Commission of reducing road fatalities by 50% in the period 2000 – 2010 was almost achieved with an estimated reduction of 44% [1], and the same target was adopted for the period 2010 – 2020. Pedestrians - including people with disabilities - account for 20% of the fatalities in EU-24. More than 6.000 pedestrians died in road traffic accidents in 2010. The 75% of pedestrians fatalities takes place in urban environments and around the 25% of them occurs in on or close to a pedestrian crossing [1]. Accordingly, new solutions specifically devised to increase the safety of pedestrians, including users with disabilities, at urban crosswalks can be of great help to reach this ambitious mid-term goal.

Assistive technology is usually defined as *any item, piece of equipment, or product system, whether acquired commercially, modified, or customized, that is used to increase, maintain, or improve functional capabilities of individuals with disabilities* [2]. In this paper, this concept is extended to *Assistive Intelligent Transportation Infrastructures*, using intelligent pedestrian crossings as a specific case study. We define an *Assistive Pedestrian Crossing* as a *intelligent pedestrian crossing able to interact with users with disabilities and provide an adaptive response to each type of disability to increase, maintain or improve their functional capabilities while crossing*. This definition involves several technologies such as pedestrian detection and tracking, disability identification, and infrastructure adaptation.

D. F. Llorca, R. Quintero, I. Parra, R. Izquierdo, C. Fernández, M. A. Sotelo are with the Computer Engineering Department, Polytechnic School, University of Alcalá, Madrid, Spain. email: llorca@aut.uah.es.

In this paper pedestrian location is implemented by means of a wide-angle stereo-based pedestrian detection and tracking system, that has been developed to be used in both daytime and nighttime conditions. Pedestrian detection is performed by region growing over temporal 3D density maps, which are obtained by means of stereo reconstruction and background modelling. 3D tracking allows to correlate the pedestrian position with the different pedestrian crossing regions (waiting and crossing areas). Pedestrian disability identification is carried out by means of an RFID-based approach that ensures anonymous identification. Pedestrians with disabilities are only required to wear a portable and passive RFID tag, whereas the infrastructure is equipped with a RFID reader and two antennas. An automatic stereo-RFID calibration procedure is defined to transform from RSSI (Received Signal Strength Indicator) signal to 3D distances. In order to associate detected RFID tags with pedestrians in the scene, a global nearest neighbor algorithm is applied to solve data association between stereo targets and RFID measurements. Extensive experimental results are provided to validate the stereo-based pedestrian location. Finally, the proposed assistive technology combining stereo location and RFID anonymous identification is validated in a real crosswalks, including different complex scenarios with multiple pedestrians and RFID tags.

The rest of the paper is organised as follows: Section II describes the state-of-the-art concerning stereo-based pedestrian detection as well as RFID-vision localization approaches. The system description including the stereo-based pedestrian location, the RFID localization and the stereo-RFID data association procedures are given in Section III. Experimental results are provided in Section IV. Finally, conclusions and future works are addressed in Section V.

II. RELATED WORK

Stereo-based pedestrian location and tracking is a well-known topic in the context of ITS. Vehicle-based pedestrian protection systems have been widely surveyed in [3]. Stereo cues are highly relevant since they enhance both the region of interest selection [4] and the classification [5] stages, providing relative distance values that are essential for collision avoidance manoeuvres such as automatic steering [6] or emergency braking [7]. Monocular approaches have been widely proposed in the context of infrastructure-based applications since background subtraction or motion history

978-1-4673-6596-3/15 $31.00 © 2015 IEEE
DOI 10.1109/ITSC.2015.223

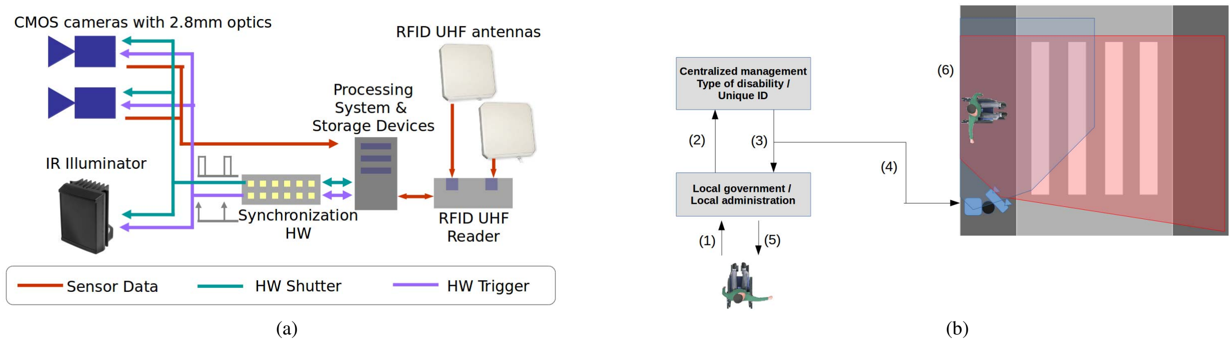


Fig. 1. (a) Sensor architecture; (b) General procedure for anonymous identification.

techniques can be directly applicable [8], [9]. However, accurate depth measurements are still needed to allow the applicability of infrastructure-based safety measurements. Thus, in [10] a multi-sensor platform is proposed to detect pedestrians at intersections, with laser scanners and FIR cameras. We also remark the SafeWalk commercial system [11] which was the first stereo pedestrian detection platform available for its use at urban crosswalks. The main drawbacks of this system are its narrow field of view and its close range, which limits its use to the pedestrian waiting area at sidewalks. Accordingly, monitoring a multiple lane crosswalk would require at least two SafeWalk systems for the pedestrian waiting areas and one C-Walk (monocular) for the crosswalk. Thus, stereo measurements are only available at pedestrian waiting zones. Our approach allows monitoring a two-lane crosswalk including pedestrian waiting areas with only one stereo platform. This approach was previously presented to automatically assess relevant variables related with the pedestrian's intent to cross or wait [12].

A considerable number of works have been proposed for the localization of passive or active RFID tags with fixed RFID readers as well as the localization of readers using a fixed set of tags [13]. However, for the course of this paper, we focus on the localization of moving passive RFID tags using RFID readers in combination with vision-based approaches. In most cases, the combination of RFID- and vision-based localization approaches is used to increase the global localization accuracy. Thus, in [14], eight directive antennas, a RFID reader and a camera are embedded on a mobile robot to detect passive tags worn on the user's clothes. A particle filter (PF) is then used to fuse data from RFID and vision. In [15], RFID-based localization in a small area of interest is carried out via RSSI measurements and combined with a camera-based localization by means of an Unscented Kalman Filter (UKF). There is an obvious improvement in the RFID-based localization accuracy thanks to the use of the monocular vision system. A similar fusion scheme using a Particle Filter (PF) to combine RSSI data with stereo measurements is proposed in [16]. However, as suggested by several studies [17], [18] there are intrinsic limitations when using RSSI as a distance metric in terms of accuracy and stability for localization purposes. Thus, as in [19], we propose to use the RFID system as an identification tool (type of disability), and the vision system (stereo) for localization. Thus the data fusion problem can be translated into a pure data association problem. A global nearest neighbor algorithm with a novel distance metric is proposed to link RFID tags with stereo objects (pedestrians). The system is devised to be used in outdoor scenarios, in medium sized areas with a measurement range up to 15m.

III. SYSTEM DESCRIPTION

A. System Layout

The sensor architecture is depicted in Fig. 1(a). The stereo platform is composed of two CMOS USB cameras, with a focal length of 2.8mm and a baseline of 30cm, synchronized with an IR illumination device controlled by a photocell. A specific synchronization HW controls the external trigger of the cameras and the IR illumination. An UHF RFID reader with two UHF antennas is connected to the Ethernet card of the PC. Stereo and RFID data are synchronized via SW.

The general approach to perform anonymous identification of the type of disability is described in Fig. 1(b). The person with disability applies (1) to the local administration for a disability identification card (RFID tag). The local administration certifies the type of disability and (2) asks for a RFID tag to the central management unit (CMU). The CMU assigns the RFID identifier to the specific disability, (3) sends the RFID tag to the local administration, and (4) remotely updates this information in all the assistive systems. Once the local administration (5) provides the user with the RFID tag, the disability identification process (6) will be fully anonymous.

In order to apply effective assistive actions from the infrastructure, knowledge about the location of the user with disability is critical. A robust stereo-RFID data association procedure is needed to met this requirement. This paper focuses in stereo localization, RFID identification and data association. Experiments including sophisticated assistive actions from the infrastructure are out of the scope and considered as future works. Fig. 2 depicts the overall layout of the proposed approach.

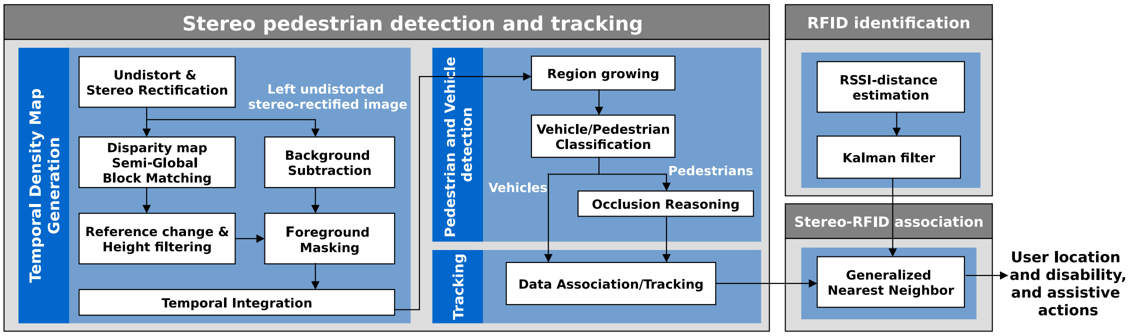


Fig. 2. Flowchart representation of the overall approach.

B. Stereo-based pedestrian detection and tracking

Pedestrian location is performed by using a *XZ* density map which can be seen as an accumulated buffer [12]. First, the origin of coordinates is translated from left camera to the road plane. Then, 3D-points are projected on a *XZ*-map and points within the range $0.2m < Y < 2m$ are removed. Each point adds a value to the accumulation buffer which follows a Gaussian distribution with the maximum at the centre pixel and decreasing in the neighbouring pixels. Since the influence of each 3D-point on the density map is cumulative, the resulting map will contain large values in regions with a high density of 3D points. In order to reduce the effect of stereo matching errors, a temporal density map is finally obtained by integrating the projected points during the last N frames. This map includes information related with pedestrians and vehicles, but also with static objects such as poles, trees, etc. In order to remove static objects a dynamical background subtraction algorithm [20] is applied to mask the disparity map with the foreground objects (see Fig. 3). A learning rate of 0.1 is used to adapt the background, minimising the probabilities of incorporating pedestrians or vehicles inside the background model.

Object segmentation is carried out by means of region-growing over the masked-temporal density map, providing a list of object hypotheses. Each candidate is firstly classified as pedestrian or vehicle analysing their velocities, size and the image location of their first appearances in the scene. The shape of each candidate with a size that may correspond to multiple pedestrians is managed using an occlusion reasoning algorithm [9] which uses compactness, convexity and convex hull to divide the blob in multiple blobs. Tracking is finally carried out using a linear Kalman filter. The motion of both pedestrians and vehicles are modelled using a constant velocity model, allowing accelerations by means of process noise. The state variables are pixel and 3D positions and their corresponding velocities. The measurement vector includes pixel and 3D positions. Data association problem is solved using the Hungarian assignment with a metric that combines the 3D Mahalanobis distance and blob appearance (normalised cross-correlation) [21]. See Fig. 2 for a detailed description of the algorithm.

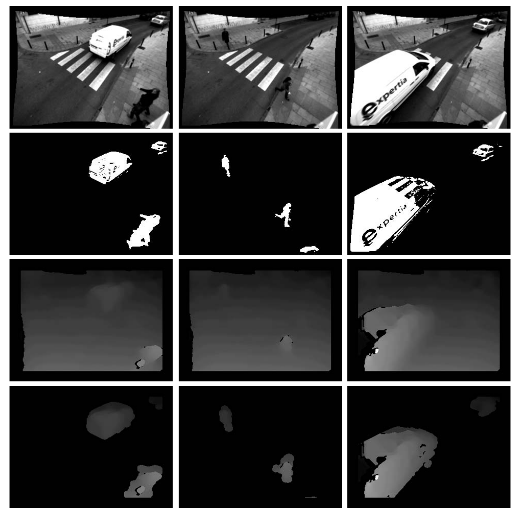


Fig. 3. First row: rectified left images ; Second row: foreground map after background subtraction; Third row: disparity map; Fourth row: disparity map masked with foreground map.

C. RFID-based localization

Generally, the signal strength received by a sensor from another one is considered as a monotonically decreasing function of their distance, which includes the transmission and reception antennas power and their gains. In practice, a simplified form of the relation between distance and receive power has been mostly used [18]:

$$P_r(dBm) = P_{r1}(dBm) - K.log_{10}(D(m)) \quad (1)$$

where P_{r1} is the received power in dBm at one meter, K is the loss parameter and D is the distance between the transmitter and the receiver. The values of P_{r1} are K are determined by minimizing the root mean square error using calibration data, i. e., RSSI and distance values. One of the main contributions of our approach is that this calibration data can be automatically obtained. A sequence of one person moving around the stereo region and wearing one tag in a fixed position and orientation has to be recorded. Then, the stereo-based pedestrian location system is used to get 3D measurements w.r.t. the left camera, which are directly associated with the RSSI values given by the antennas (note that data association is not needed at this point). We approximate the 3D tag position to the center of the blob in the *XZ*-map, assuming a fixed tag height w.r.t. the road plane. Although this approach provides distance measurements that suffer from both the stereo accuracy and the simplification of considering the tag at the center of the blob, it also automatically provides thousands of measurements in a really short period of time.

In order to have all sensors located at the same point, we have devised a two antennas configuration like the one depicted in Fig. 4, that is integrated at the same stereo baseline. Point w.r.t. the left camera (LC) P^1_{LC} is directly given by 3D stereo reconstruction. The relative positions of both the left and the right antennas (LA; RA) w.r.t. the LC are approximated by using an identity rotation matrix and translation vectors with only the X component. Thus, points P^1_{LA} and P^1_{RA} are directly obtained. After applying the automatic calibration procedure, we obtain the parameters of Eq. 1 and the RSSI-distance curves depicted in Figs. 5(a) and 5(b) for left and right antennas respectively.

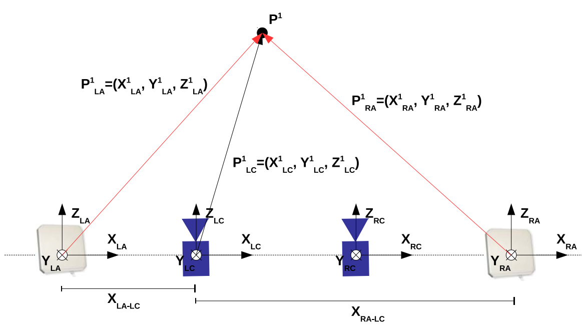
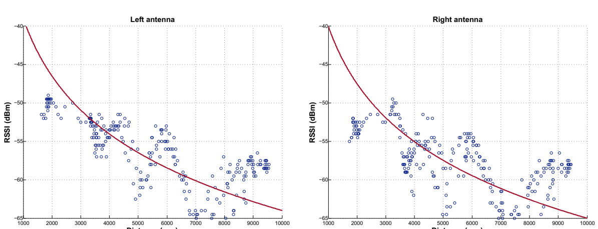


Fig. 4. Relative position between cameras and RFID antennas.



(a) (b)

Fig. 5. Obtained relationship between RSSI and distance for (a) left ($P_{r1} = -36dBm$; $K = 25$) and (b) right ($P_{r1} = -35dBm$; $K = 25$) antennas.

Finally, a Kalman filter is used to get steadier distance measurements, for each tag and antenna. The state vector includes the RSSI value and its variation, and the measurement vector only the RSSI value. The variance is modeled using the calibration measurements and the obtained RSSI-distance curves (see Fig. 2 for details).

D. Stereo-RFID data association

Ideally, a single RSSI value yields a sphere with the antenna position at its center as possible tag locations. One common assumption is to consider a fixed and known tag height to reduce the sphere to a circle. Then, the tag position can be determined by intersecting the circles from different antennas. A minimum of three antennas are required to obtain one location. In our case, the radiation pattern of both antennas covers 180° degrees, so one of the intersection locations can be discarded, and thus two antennas are enough for providing a unique solution. However, as suggested by previous studies [17], [18] there are intrinsic limitations when using RSSI as a distance metric in terms of accuracy and stability for localization purposes (see Figs. 5(a) and 5(b)). On top of that, our antennas are located very close to each other, so in most cases the intersection point (or area, including the uncertainties) is not enough for a robust association. Accordingly, a new data association procedure is proposed by means of general nearest neighbor and a new distance metric that models the probability of association between a 3D blob and the detected tags.

The distance from the tag to the circumference of one of the antennas (assuming fixed height of the tag) is modeled using a univariate normal distribution where the mean value is the RFID distance, the variance σ_R^2 is the one computed after RSSI-distance calibration, and the independent variable is the difference given by the stereo system w.r.t. the antenna. Following the example with the right antenna depicted in Fig. 6, lets consider a blob i detected by the stereo system. After applying proper translation, we obtain the 3D coordinates of blob i w.r.t. the right antenna $P_i^{stereoR} = (X_i^{stereoR}, Y_i, Z_i^{stereoR})$, and the distance as $d_i^{stereoR} = ||P_i^{stereoR}||$. On the other hand, a tag j has been detected, and the RSSI value has been transformed into the distance value d_j^{rfidR}. Thus, the distance metric that represents the probability of tag j being worn by person i for the right antenna is:

$$d_{ij}^R = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(d_i^{stereoR} - d_j^{rfidR})^2}{2\sigma^2}} \quad (2)$$

The same procedure is applied to left antenna to obtain d_{ij}^L. If one of the antennas does not receive signal, the metric would be set to zero. In order to compute the global metric that represents the probability that tag j is being worn by person i, the following equation would be applied:

$$d_{ij} = d_{ij}^L . d_{ij}^R \quad (3)$$

This approach can be easily extended to N antennas by applying the following equation:

$$d_{ij} = \prod_{k=1}^{N} d_{ij}^k \quad (4)$$

The distance metrics are computed between the predicted position of all pedestrians and all the RFID tags. The Hungarian method is applied to solve the nearest neighbor problem, with the following modification: once the metric provided by one tag and one pedestrian is assigned with a distance metric greater than 0.8, this association will be maintained until that pedestrian or tag are not detected.

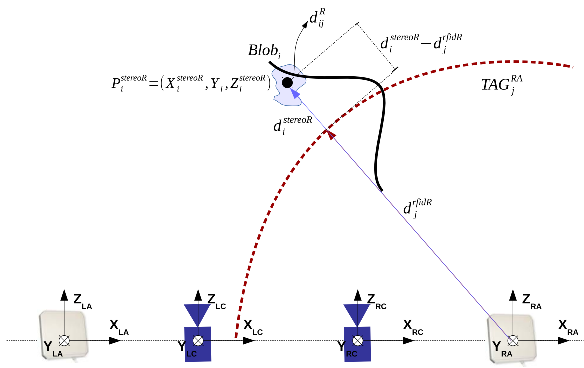


Fig. 6. Example of the proposed distance metric between a blob and a tag for the right antenna.

TABLE I
PEDESTRIAN DETECTION RESULTS.

Dataset Id.	**Lighting conditions**	**Duration (minutes)**	**DR**	**FPR**
1	Daytime	4.1	99.79%	0.20%
2	Daytime	9.4	97.01%	1.10%
3	Nighttime	15.7	97.17%	0.00%
4	Daytime	24.2	98.73%	1.34%
5	Daytime/Sunset	612	99.01%	1.63%

IV. EXPERIMENTS

A. Stereo pedestrian detection results

The stereo-based pedestrian location system has been tested in different scenarios with different lighting conditions (see Fig. 7) and different platforms (see Fig. 8). Extensive results were obtained using a prototype fully operative in the city of Guadalajara (Spain). Detection rates (DR) and false positive rates (FPR) are given in Table I for a set of sequences corresponding to different locations and illumination conditions. As can be observed the system provided very accurate results, even for nighttime conditions, with an average DR of 99% at a FPR of 1.5%. In addition, we have analyzed the detection delay. On average the 90% of the pedestrians detected by the system were tracked in less than 10 frames after they were fully visible (around 0.33 seconds).

Fig. 7. Different scenarios covering two-lane crossing and two pedestrian waiting areas.

B. Stereo-RFID data association

The stereo-RFID setup is depicted in Fig. 9. A set of short sequences (around one minute long) has been recorded including different number of people and tags. The users were required to wear the tag at a fixed height, and pointing to the antennas. Correct and false/none associations between pedestrians and tags are evaluated, and global results are

Fig. 8. Different stereo platforms with IR illuminator.

TABLE II
STEREO-RFID ASSOCIATION RESULTS.

Type of sequence	**Duration (minutes)**	**Correct associations**
One tag/ One pedestrian	9.6	6/6 (100.00%)
Two tags / Two pedestrians	5.4	13/16 (81.25%)
N tags / M pedestrians	3.3	3/5 (60.00%)

presented in Table II. The overall performance is reasonably high considering the RSSI-distance accuracy and stability. Fig. 10 depicts some examples, including 3D reconstruction, and final association.

Fig. 9. Left: RFID TAGs and reader; Right: platform with RFID antennas and cameras.

V. CONCLUSION

We presented a novel stereo-based pedestrian detection and tracking system combined with a RFID-based identification approach to anonymously identify the disability of the pedestrian in the context of assistive pedestrian crossings. The accuracy of the stereo detection system provides the infrastructure with knowledge about each user location. We show that the proposed stereo-RFID combination allows an automatic RSSI-distance calibration procedure including thousand of measurements. A novel distance metric was proposed to solve the data association problem between RFID tags and stereo blobs. We obtained encouraging results that validate the proposed approach. Future work involves the study of new identification technologies (active RFID, active bluetooth, etc.) as well as new antennas location to improve the data association problem. Finally, a set of assistive and adaptive measures taken from the infrastructure will be designed and tested with different types of disabilities.

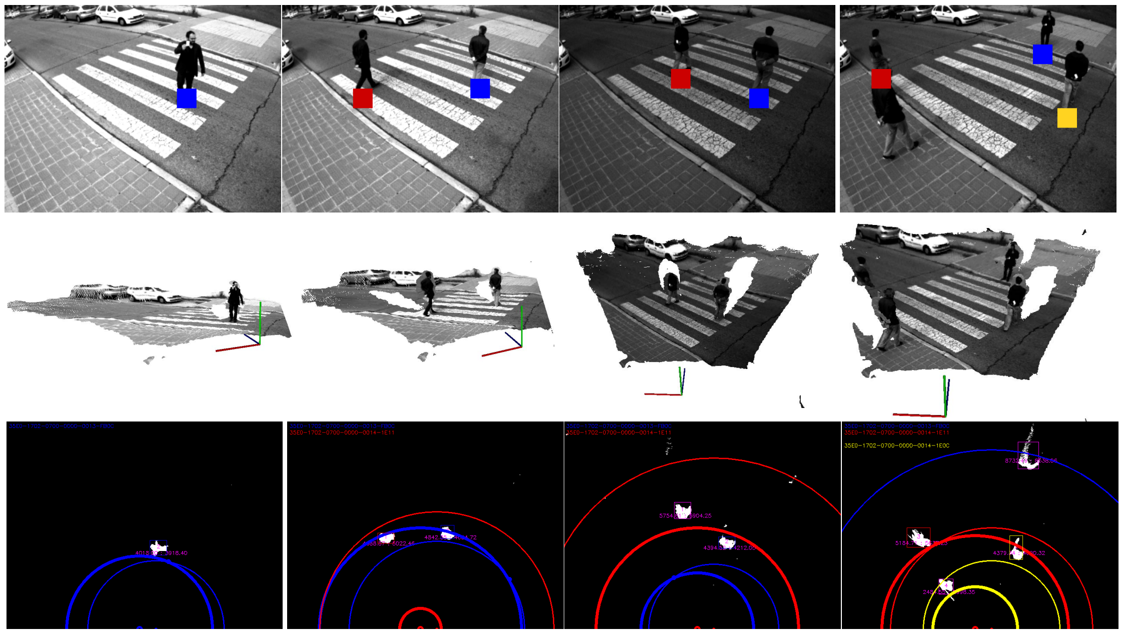


Fig. 10. Upper row: left image with color-coded identification. Middle row: 3D reconstruction. Lower row: *XZ*-map, detected blobs and RSSI circumferences. Each tag is labeled with a different color. The line width is different for each antenna.

VI. ACKNOWLEDGMENTS

This work was supported by the Research Grants DISADAPT SPIP2014-1300 (General Traffic Division of Spain), IMPROVE DPI2014-59276-R (Spanish Ministry of Economy), and SEGVAUTO-TRIES-CM S2013/MIT-2713 (Community of Madrid).

REFERENCES

[1] EC, "Basic fact sheets 2012. mobility and transport, road safety." 2012, site: http://ec.europa.eu/transport/road_safety/specialist/statistics/index_en.htm.

[2] J. Relton, "The assistive technology act of 2004," *Access World Magazine*, vol. 6, no. 1, 2004, site: http://www.afb.org/afbpress/pub.asp?DocID=aw060109.

[3] D. Gerónimo and A. M. López, *Vision-based Pedestrian Protection Systems for Intelligent Vehicles*. Springer, 2014.

[4] D. F. Llorca, M. A. Sotelo, A. M. Hellín, M. Gavilán, I. G. Daza, and A. G. Lorente, "Stereo regions-of-interest selection for pedestrian protection: a survey," *Transportation Research Part C*, vol. 25, pp. 226–237, 2012.

[5] C. G. Keller, M. Enzweiler, M. Rohrbach, D. F. Llorca, C. Schnorr, and D. M. Gavrila, "The benefits of dense stereo for pedestrian detection," *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 4, pp. 1096–1106, 2011.

[6] D. F. Llorca, V. Milanés, I. Parra, M. Gavilán, I. G. Daza, J. Pérez, and M. A. Sotelo, "Autonomous pedestrian collision avoidance using a fuzzy steering controller," *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 2, pp. 390–401, 2011.

[7] V. Milanes, D. F. Llorca, J. Villagrá, J. Pérez, I. Parra, C. González, and M. A. Sotelo, "Vision-based active safety system for automatic stopping," *Expert Systems with Applications*, vol. 39, no. 12, pp. 11 234–11 242, 2012.

[8] S. Kohler, M. Goldhammer, S. Bauer, S. Zecha, K. Doll, U. Brunsmann, and K. Dietmayer, "Stationary detection of the pedestrian's intention at intersections." *IEEE Intelligent Transportation Systems Magazine*, vol. 5, no. 4, pp. 87–99, 2013.

[9] S. Álvarez, D. F. Llorca, and M. A. Sotelo, "Hierarchical camera auto-calibration for traffic surveillance systems," *Expert Systems with Applications*, vol. 41, pp. 1532–1542, 2014.

[10] D. Weimer, S. Kohler, C. Hellert, K. Doll, U. Brunsmann, and R. Krzikalla, "Gpu architecture for stationary multisensor pedestrian detection at smart intersections," in *IV2011, IEEE Intelligent Vehicle Symposium*, 2011.

[11] W. Favoreel, "Pedestrian sensing for increased traffic safety and efficiency at signalized intersections," in *8th IEEE International Conference on Advanced Video and Signal-Based Surveillance*, 2011.

[12] D. F. Llorca, I. Parra, R. Quintero, C. Fernández, R. Izquierdo, and M. A. Sotelo, "Stereo-based pedestrian detection in crosswalks for pedestrian behavioural modelling assessment," in *ICINCO2014, Int. Conf. on Inf. in Control, Automation and Robotics*, 2014.

[13] T. Sanpechuda and L. Kovavisaruch, "A review of rfid localization: Applications and techniques," in *ETI-CON 2008*, 2008.

[14] T. Germa, F. Lerasle, N. Ouadah, and V. Cadenat, "Vision and rfid data fusion for tracking people in crows by a mobile robot," *Computer Vision and Image Understanding*, vol. 114, pp. 641–651, 2010.

[15] T. Nick, S. Cordes, J. Gotze, and W. John, "Camera-assisted localization of passive rfid labels," in *International Conference on Indoor Positioning and Indoor Navigation*, 2012.

[16] F. Schwiegelshohn, T. Nick, and J. Gotze, "Localization based on fusion of rfid and stereo image data," in *10th Workshop on Positioning Navigation and Communication (WPNC)*, 2013.

[17] A. T. Parameswaran, M. I. Husa, and S. Upadhyaya, "Is rssi a reliable parameter in sensor localization algorithms an experimental study," in *Field Failure Data Analysis Workshop*, 2009.

[18] K. Heurtefeux and F. Valois, "Is rssi a good choice for localization in wireless sensor network?" in *IEEE 26th International Conference on Advanced Information Networking and Applications (AINA)*, 2012.

[19] A. Isasi, S. Rodriguez, J. L. D. Armentia, and A. Villodas, "Location, tracking and identification with rfid and vision data fusion," in *2010 European Workshop on Smart Objects: Systems, Technologies and Applications (RFID Sys Tech)*, 2010.

[20] S. Álvarez, M. A. Sotelo, D. F. Llorca, R. Quintero, and O. Marcos, "Monocular vision-based target detection on dynamic transport infrastructures," *Lecture Notes in Computer Science*, vol. 6927, pp. 576–583, 2012.

[21] I. Parra, D. F. Llorca, M. A. Sotelo, L. M. Bergasa, P. R. de Toro, J. Nuevo, M. Ocana, and M. A. García-Garrido., "Combination of feature extraction methods for svm pedestrian detection," *IEEE Transactions on Intelligent Transportation Systems*, vol. 8, no. 2, pp. 292–307, 2007.