Received December 18, 2019, accepted January 9, 2020, date of publication January 20, 2020, date of current version January 30, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2968175

Towards Intelligent Retail: Automated On-Shelf Availability Estimation Using a Depth Camera

ANNALISA MILELLA, ANTONIO PETITTI, ROBERTO MARANI, GRAZIA CICIRELLI, AND TIZIANA D'ORAZIO

Institute of Intelligent Industrial Systems and Technologies for Advanced Manufacturing, National Research Council of Italy, 70126 Bari, Italy

Corresponding author: Annalisa Milella (annalisa.milella@stiima.cnr.it)

This work was supported by the grant Electronic Shopping & Home delivery of Edible goods with Low environmental Footprint (E-SHELF), POR Puglia FESR-FSE 2014–2020 (Grant Id. OSW3NO1).

ABSTRACT Efficient management of on-shelf availability and inventory is a key issue to achieve customer satisfaction and reduce the risk of profit loss for both retailers and manufacturers. Conventional store audits based on physical inspection of shelves are labor-intensive and do not provide reliable assessment. This paper describes a novel framework for automated shelf monitoring, using a consumer-grade depth sensor. The aim is to develop a low-cost embedded system for early detection of out-of-stock situations with particular regard to perishable goods stored in countertop shelves, refrigerated counters, baskets or crates. The proposed solution exploits 3D point cloud reconstruction and modelling techniques, including surface fitting and occupancy grids, to estimate product availability, based on the comparison between a reference model of the shelf and its current status. No *a priori* knowledge about the product type is required, while the shelf reference model is automatically learnt based on an initial training stage. The output of the system can be used to generate alerts for store managers, as well as to continuously update product availability estimates for automated stock ordering and replenishment and for e-commerce apps. Experimental tests performed in a real retail environment show that the proposed system is able to estimate the on-shelf availability percentage of different fresh products with a maximum average discrepancy with respect to the actual one of about 5.0%.

INDEX TERMS Automated stock monitoring, intelligent retail, RGB-D sensors, 3D reconstruction and modeling.

I. INTRODUCTION

A product is Out-Of-Stock (OOS) when it is not available on shelf for customer purchase for some contiguous time [1]. Efficient Consumer Response (ECR) [2] reports that stock-outs are a central concern for consumers, being the third most important issue for shoppers, after the desire for shorter lines at the cash register and more promotions. If OOS conditions occur repeatedly, customer satisfaction is reduced with potentially negative effects for both retailers and manufacturers. Significant research efforts have been devoted, so far, by academic and industrial bodies to address the OOS problem, mainly focusing on supply chain management issues. Nevertheless, the overall OOS rate remains high in all retail sectors [3], [4], with an estimated loss of profit for the retail industry of billions of euros per year.

The associate editor coordinating the review of this manuscript and approving it for publication was Okyay Kaynak.

Investigations on root cause analysis of OOS worldwide revealed that most of responsibility for lowering OOSs rests in the retail store with, on average, 25% of OOSs being shelf-OOSs (SOOSs), i.e., the product is available but the shelf has not been refilled yet [3]. SOOSs are particularly frustrating, since there is still stock in the store, meaning that forecast may have been correct and the supply and delivery functions were executed appropriately, but due to replenishment delays or errors, the product could not land in the shopper's cart. High checking frequency of shelves can help reducing SOOS events. However, current human-based survey practices are labor-intensive and do not provide reliable assessment, especially for Fast-Moving Consumer Goods (FMCG), fresh food and products with a due date.

In the last decade, advanced sensor-based technologies mainly using Auto-ID systems, weight sensors and imaging devices have been proposed for automatic stock monitoring and inventory [5], [6]. While these systems were shown to improve on-shelf availability to some extent, further research efforts are still needed towards the development of more robust and efficient technological solutions, able to deal with the large variability of store environments and goods.

In this work, a novel framework for online shelf monitoring using a depth sensor is proposed. It can generate a 3D point cloud of the shelf and products therein, while recovering information on the actual product availability, based on surface reconstruction and modelling techniques. No *a priori* knowledge of the product characteristics is required. The only assumption is that the shelf can be represented as a planar surface, whose model is built via an initial automated calibration stage. Then, points sticking out of the plane are segmented as belonging to items on the shelf and are used to estimate the current product On-Shelf Availability (OSA). The latter is defined in terms of percentage of the volume occupied by the items on the shelf with respect to the maximum available space. The module is integrated into an eco-sustainable electronic commerce system, in order to provide updated stock inventory data to an e-commerce app.[1]

Experimental validation was carried out in a real retail environment with an Intel RealSense D435[2] sensor. Different types of products were considered with specific attention to critical items such as perishable and fresh goods stored in countertop shelves, refrigerated counters, baskets and crates. Results show that, despite the low quality of the sensor data, it is possible to extract useful information about the shelf status in a completely automated and non-invasive way, thus overcoming time and cost problems related to traditional man-made surveys.

The rest of the paper is structured as follows. Section II reports related work. The data acquisition system and the proposed algorithms for online OSA estimation are described in Section III. Experimental results are presented in Section IV. Conclusions are drawn in Section V.

II. RELATED WORK

To ensure high on-shelf availability of products, human-based audits have been traditionally adopted by retailers. However, such a method is labor-intensive and subject to human error. In addition, its effectiveness highly depends on the frequency of the surveys. As a result, sensor-based technologies have emerged in the last decade to automatically monitor on-shelf availability and detect OOS on a rapid basis, without having to conduct physical store audits.

In this respect, pilot studies carried out by some well-established retailers have shown that Radio-Frequency Identification (RFID) solutions are helpful to improve inventory tracking and reduce stockouts [7], [5]. However, RFID technology is still expensive and requires time-consuming item-level tagging, which prevents its widespread use. As an alternative, in [8], a wireless sensor network is proposed for real-time shelf monitoring, using IR-based sensors, that can be installed and removed through bars attached to the shelves. Smart shelves using weight sensors have been developed in [9] to detect weight changes and determine accordingly the number of products available on the shelf. However, weight sensors entail high installation costs. In addition, they can only determine the number of products stacked on the shelf without accounting for possible product misplacements, as they do not allow for product identification and tracking [10].

More recently, computer vision systems have been proposed by many as a promising solution for smart retail applications including detection of misplaced products [11], verification of planogram compliance [6], and stock assessment [12]. Cameras for data gathering can be installed at fixed locations in the store [6], or integrated into smartphones [13], manually-driven carts [12] or on-board mobile robotic platforms [14]. Image processing techniques mainly using state-of-the-art feature-based or template-based matching algorithms have been adopted to look for anomalies and trigger alarms for store managers. For instance, in [11], an image matching technique using SURF features is developed to match the current image of a shelf with a reference image of the same shelf to estimate on-shelf availability and detect product misplacements. The method is demonstrated for a limited number of test cases with no performance metrics. Exhaustive template matching is employed in [15] to look for products in an image of the shelf to count products and raise alarms in case of stock-out events or product misplacements. This technique is computationally expensive as it requires several templates to be created for a product under various lighting conditions, angles and distances. Furthermore, the authors only present qualitative results for a single test case.

In order to cope with the high variability of store environments, the use of machine learning techniques has been proposed, to ensure higher robustness and accuracy, although at the cost of labor-intensive manual image annotation for training. An example can be found in [12], where a supervised learning approach based on Support Vector Machines (SVM) is developed for OOS detection in panoramic images of retail shelves. The camera image setup includes a camera cart that is moved parallel to the shelf to acquire images from multiple viewpoints. Results are presented from supermarket shelves containing labels near the ruptures showing a detection accuracy of 84.5% in detecting OOS events. In [14], Deep Convolutional Neural Networks (DCNNs) are proposed to classify three types of shelves, namely shelves with special offers, shelves with a standard layout and shelves with SOOSs situations, based on visual and textual clues of monocular images acquired by a camera mounted on-board an autonomous mobile robot.

While 2D images convey detailed information about the surveyed scene and allow for accurate detection of planogram anomalies or ruptures, results are highly affected by occlusions due to perspective effects. As an alternative, RGB-D

[1] http://www.eshelf.it/
[2] https://www.intelrealsense.com/depth-camera-d435/

sensors, i.e., depth sensing devices that work in association with RGB cameras by augmenting conventional 2D images with distance information in a per pixel basis, have been demonstrated to be successful in different fields, such as indoor and outdoor environment mapping [16], [17] and gesture recognition [18]. An application of RGB-D sensors in the context of intelligent retail can be found in [19], where a top-view RGB-D camera is employed to monitor shoppers' behaviors. A Smart Shelf concept has been recently demonstrated by Intel [20] using Intel's RealSense depth cameras linked to stock sensors, to automatically recognize products and display associated information such as pricing, promotional offers or other appropriate content.

In this work, a 3D shelf monitoring system using an RGB-D sensor, namely an Intel RealSense D435 camera, is proposed. To the best of our knowledge, this is the first application of such a technology for online estimation of product on-shelf availability and early OOS detection. Compared to state-of-the-art vision-based methods, the main novelty of the proposed system is that it exploits depth information in order to not only provide an alert about the presence/absence of a product, but also to estimate the number of items actually available on the shelf, thus yielding a more precise and timely knowledge of the shelf status. In addition, no manual training is required, nor *a priori* information on the shelf or product characteristics is needed, whereas the only assumption is that the surveyed items lay on a planar shelf surface whose model is automatically built via an initial automated training phase.

III. MATERIALS AND METHODS

A shelf monitoring framework using 3D data from a depth sensor is proposed to automatically estimate the availability of products on shelves and early detect OOS in a retail environment. The system is intended to continuously survey critical products, such as perishable and fresh goods stored in countertop shelves, baskets or crates, although the proposed strategy is able to deal with any type of product available in a store. Some typical scenarios are shown in Fig. 1.

The overall processing pipeline is shown in Fig. 2. The sensor provides a 3D reconstruction of the scene, which is computed based on an infrared (IR) stereo processing algorithm running on-board the camera and is used for OSA estimation, as follows. First, a preliminary calibration stage is performed to built a reference model of the shelf in absence of products, using a plane fitting algorithm. Once the shelf model has been recovered, for each novel sensor acquisition, points sticking out of the shelf reference plane are segmented as belonging to products and are fed as input to a 2.5-dimensional occupancy grid algorithm to estimate the overall product volume. Such an estimate is retained as a measure of the product OSA.

In the rest of this section, first the acquisition platform and the experimental environment used for algorithm development and testing are introduced. Then, the single steps of the processing workflow are described.

(a)

(b)

FIGURE 1. Typical displays of fresh products: (a) cold cuts and cheese counter, (b) vegetable baskets and crates.

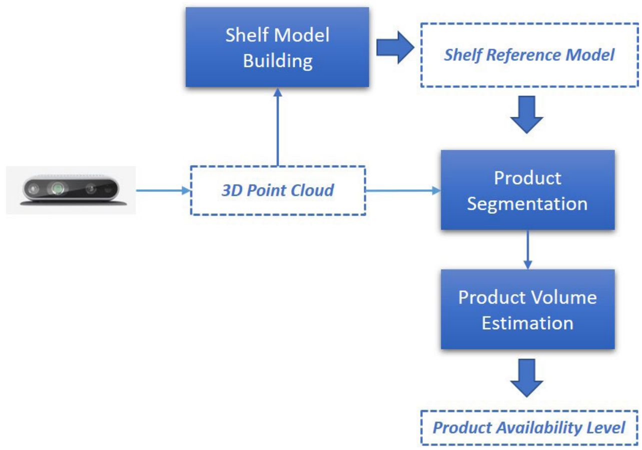
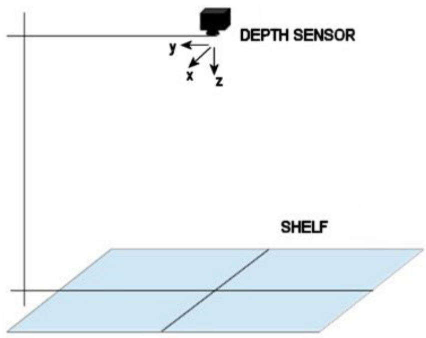


FIGURE 2. Schematic of the system workflow.

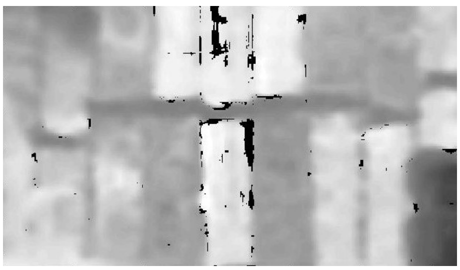
A. SYSTEM ARCHITECTURE AND SETUP

An Intel RealSense D435 imaging system is used. It features a left-right IR stereo pair and a color camera. The stereo pair consists of two global shutter CMOS monochrome cameras, with a nominal field of view of $87(H) \times 58(V) \times 95(D)$ deg, depth output resolution up to 1280×720 and frame rate up to 90 Hz, which were limited to 640×480 and a frame rate of 6.0 Hz. The reconstruction distance ranges from 0.105 m up to about 10.0 m. The color camera is a FullHD (1920×1080) Bayer-patterned, rolling shutter CMOS imager. The stream acquired by the color camera is time-synchronized with the stereo pair and is combined with stereo information to recover a color 3D point cloud of the scene. In this work, however, depth information is used for OSA estimation, while RGB images are employed for visualization purposes only. The sensor also includes



(a)

(b)



(c)

FIGURE 3. System setup and output: (a) schematic view of the sensor setup; (b) RGB image; (c) depth image. In (c), brighter points represent farther regions.

an infrared texture laser-based (Class 1) projector with a fixed pattern, which helps improving image matching in textureless surfaces, which is especially useful in indoor environments, where homogenous regions are more likely to occur.

Datasets were acquired for different shelf types and products in a retail point of sale in Bari, Italy, using the ROS wrapper for Intel RealSense.[3] A typical setup with the sensor downward looking at a cheese shelf is shown in Fig. 3 (a) along with the output of the sensor in Fig. 3 (b) and (c). It is worth to note that the proposed approach is independent of the orientation in space of the shelf support surface and, therefore, also applies to inclined or vertical shelves. In particular, in the latter case, the shelf back panel would provide the reference shelf plane in a forward-looking camera setup.

[3] https://github.com/IntelRealSense/realsense-ros

B. ON-SHELF AVAILABILITY ESTIMATION

A 3D reconstruction and modelling algorithm to estimate online the OSA of a product is developed. It consists of three main modules:

1) shelf model building;
2) product segmentation;
3) product OSA estimation.

Each module is described in detail in the following.

1) SHELF MODEL BUILDING

A preliminary calibration stage is performed to automatically build a model of the shelf in absence of products. Specifically, a planar model of the shelf support surface is assumed.

Without loss of generality, let us consider a top-view camera configuration, with the camera looking downward at the shelf as shown in the schematic of Fig. 3 (a). To fit the plane model, the relations correlating multiple views of the empty shelf by means of a maximum likelihood estimation algorithm, namely the M-estimator SAmple Consensus (MSAC) [21], are computed. In detail, N subsequent point clouds spanning a few seconds of acquisition, are used to solve a constrained nonlinear optimization problem with the aim of minimizing the Maximum Likehood Estimation (MLE) error, while filtering out measurement noise.

Once the shelf model has been built, it can be used in successive acquisitions for product segmentation and OSA estimation.

2) PRODUCT SEGMENTATION

A segmentation algorithm is applied to separate points belonging to items on the shelf from points pertaining to the shelf itself. Specifically, points whose distance from the shelf reference plane is higher than a threshold are retained as belonging to product instances and are further processed for OSA estimation. The segmentation threshold is automatically set, based on the sensor depth Root Mean Square (RMS) error at the given plane distance [22].

It is worth noticing that the segmentation process can be successfully performed only in absence of shelf occlusions, due, for instance, to the presence of shoppers or employees in close proximity of the shelf. In the current implementation, scenes whose minimum distance from the camera is lower than the sensor height relative to the shelf completely filled, are discarded.

3) PRODUCT OSA ESTIMATION

In the context of this work, the OSA level of a given product is defined in terms of percentage of occupied shelf volume.

Given a raw 3D point cloud, different modelling approaches can be adopted for volume estimation, such as convex hull, alpha-shape and occupancy grid [16]. Here, a 2.5-dimensional occupancy grid approach is employed. Specifically, first, a squared 2D grid with size s is fitted to the product point cloud; then, each cell $i = 1, 2, \dots n$ of the grid is assigned a depth value H_i, corresponding to the average distance of all points of the cell from the shelf reference plane. The volume V_i of each cell can be successively computed as:

$$V_i = s \times s \times H_i \tag{1}$$

The volume V_t occupied by the whole product at a certain observation time t can be estimated as:

$$V_t = \sum_{i=1}^{n} V_i \tag{2}$$

Finally, the OSA percentage of the product is recovered as:

$$OSA_{\%} = \frac{V_t}{V_{max}} \times 100 \tag{3}$$

where V_{max} is the maximum product volume, i.e., the volume occupied by the product when the shelf is fully replenished. V_{max} can be either calculated based on the knowledge of the shelf geometry or directly estimated by the proposed volume estimation approach after complete shelf refill. If the maximum number of products N_{max} corresponding to V_{max} is known and under the assumption that all items of the product occupy a similar volume, then the number of items per product at time t can also be easily recovered as the nearest integer of:

$$N_t = OSA_{\%} \times N_{max} \tag{4}$$

IV. EXPERIMENTAL RESULTS

Experimental tests were performed in a real retail environment with the depth sensor mounted in a top-view configuration as shown in Fig. 4.

FIGURE 4. Experimental setup in a real retail environment.

In order to evaluate the feasibility of the proposed method under various conditions, different products and shelf types were taken into account, including four varieties of cheese stored in countertops with different packaging and four kinds of fruit, stored in baskets or crates. For each product category, the accuracy of the proposed OSA estimation approach was evaluated by adding or removing a certain number of items from the shelf and comparing the result of the algorithm with the ground-truth obtained by visual inspection. Specifically, the actual availability level $OSA_{\%GT}$ was computed by manually counting the number of products available on the shelf N_{GT} and dividing this number by the maximum number of products that can be contained by the shelf N_{max}, i.e.:

$$OSA_{\%GT} = N_{GT}/N_{max} \times 100 \tag{5}$$

The discrepancy between the estimated and actual OSA level at a given observation time t is defined as:

$$d_t = |OSA^t_{\%GT} - OSA^t_{\%}| \tag{6}$$

Then, the average discrepancy and standard deviation between estimated and actual OSA over all the frames of an acquisition sequence is computed as

$$\bar{E} = \frac{1}{T}\sum_{t=1}^{T} d_t \tag{7}$$

$$\sigma = \sqrt{\frac{1}{T-1}\sum_{t=1}^{T} |d_t - \bar{E}|^2} \tag{8}$$

for $t = 1, 2, \ldots, T$ with T being the number of analysed frames. The average discrepancy can also be expressed in terms of absolute error on the estimation of the available number of products by multiplying its value by N_{max}. In addition, the coefficient of determination (r-squared) considering manual measurements and estimated OSA levels is computed as a further performance metric.

A. SHELF MODELLING

The shelf model building procedure for pre-packed slices of Silano cheese stored in a countertop and for a watermelon basket is shown in Fig. 5 and 6, respectively.

For both product types, a planar model of the shelf is built based on the plane fitting algorithm described in Section III-B1 (upper row of Fig. 5 for the cheese shelf and of Fig. 6 for the watermelon basket). To account for sensor noise, a sequence of frames spanning a few acquisition seconds is used for estimation of the reference shelf plane model. In this case, experiments have proven that six images, corresponding to a time interval of 1.0 s, are enough to filter outlier points from the input point clouds, giving more robust representations of the plane parameters.

Once the model of the blank shelf has been computed, it can be employed to estimate the available quantity of products based on the 2.5D occupancy grid approach, described

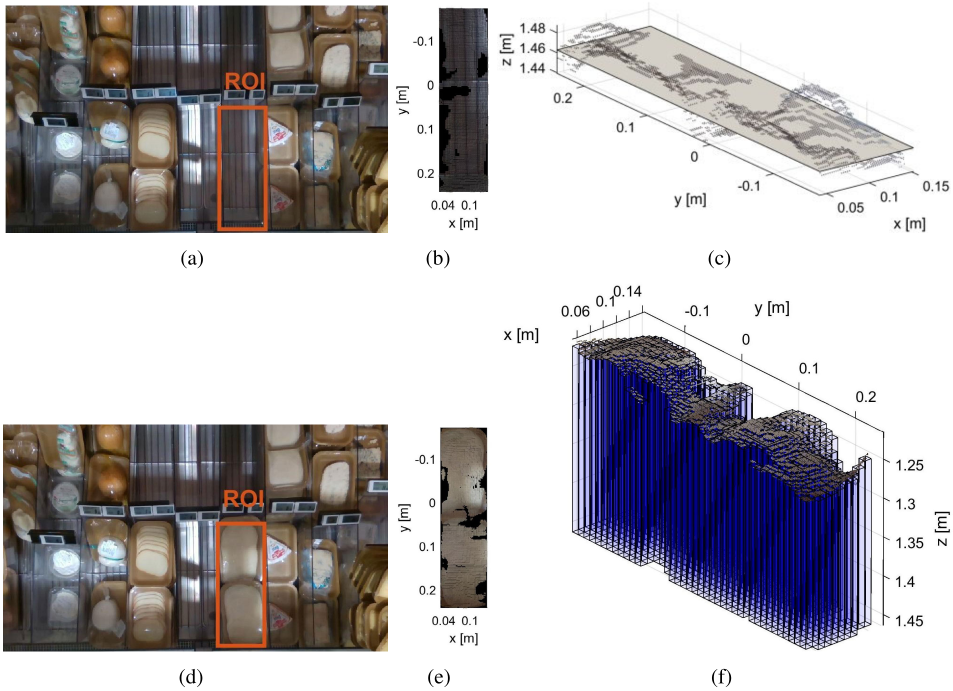


FIGURE 5. Shelf modelling (a)-(b)-(c) and maximum volume estimation (d)-(e)-(f) for a box of pre-packed Silano cheese on a cheese counter. (a)-(d) Color images with highlighted the shelf Region of Interest (ROI); (b)-(e) top view of the 3D reconstruction of the shelf ROI; (c) result of plane fitting for shelf modelling in absence of products; (f) product modelling using 2.5D occupancy grid for maximum volume estimation.

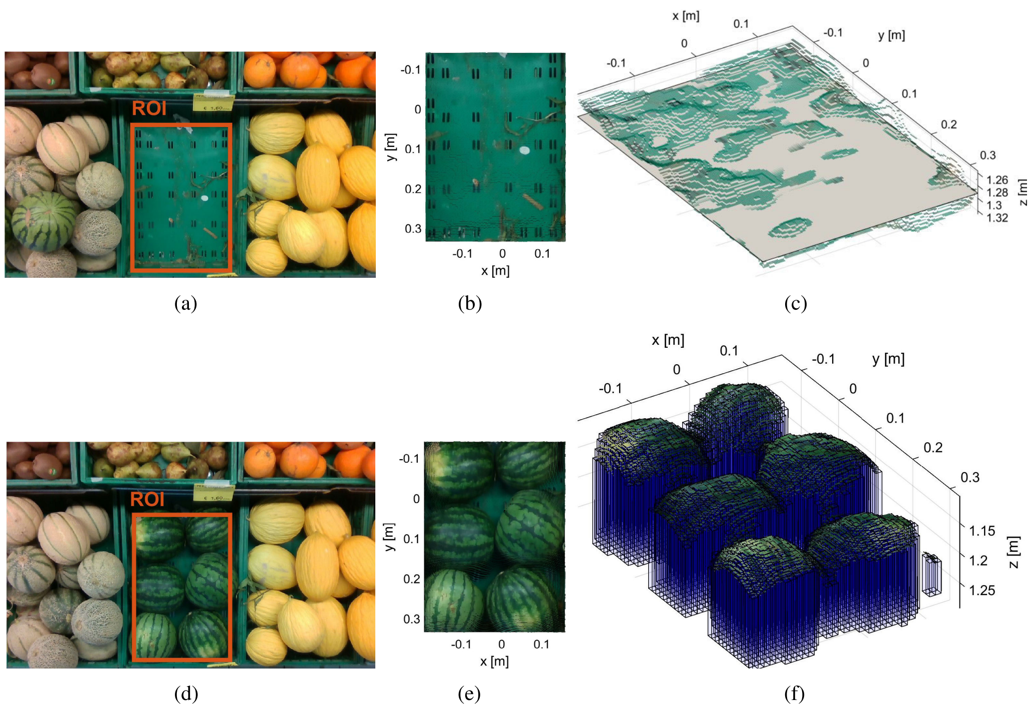


FIGURE 6. Shelf modelling (a)-(b)-(c) and maximum volume estimation (d)-(e)-(f) for a watermelon basket. (a)-(d) Color images with highlighted the shelf Region of Interest (ROI); (b)-(e) top view of the 3D reconstruction of the shelf ROI; (c) result of plane fitting for shelf modelling in absence of products; (f) product modelling using 2.5D occupancy grid for maximum volume estimation.

in Section III-B2. In particular, the maximum product volume can be computed after full replenishment of the shelf (lower row of Fig. 5 for the cheese shelf and of Fig. 6 for the watermelon basket) and is used as the reference volume V_{max} in Eq. 3. Fig. 5 (f) and Fig. 6 (f) show the result of the grid fitting algorithm for Silano cheese and watermelon,

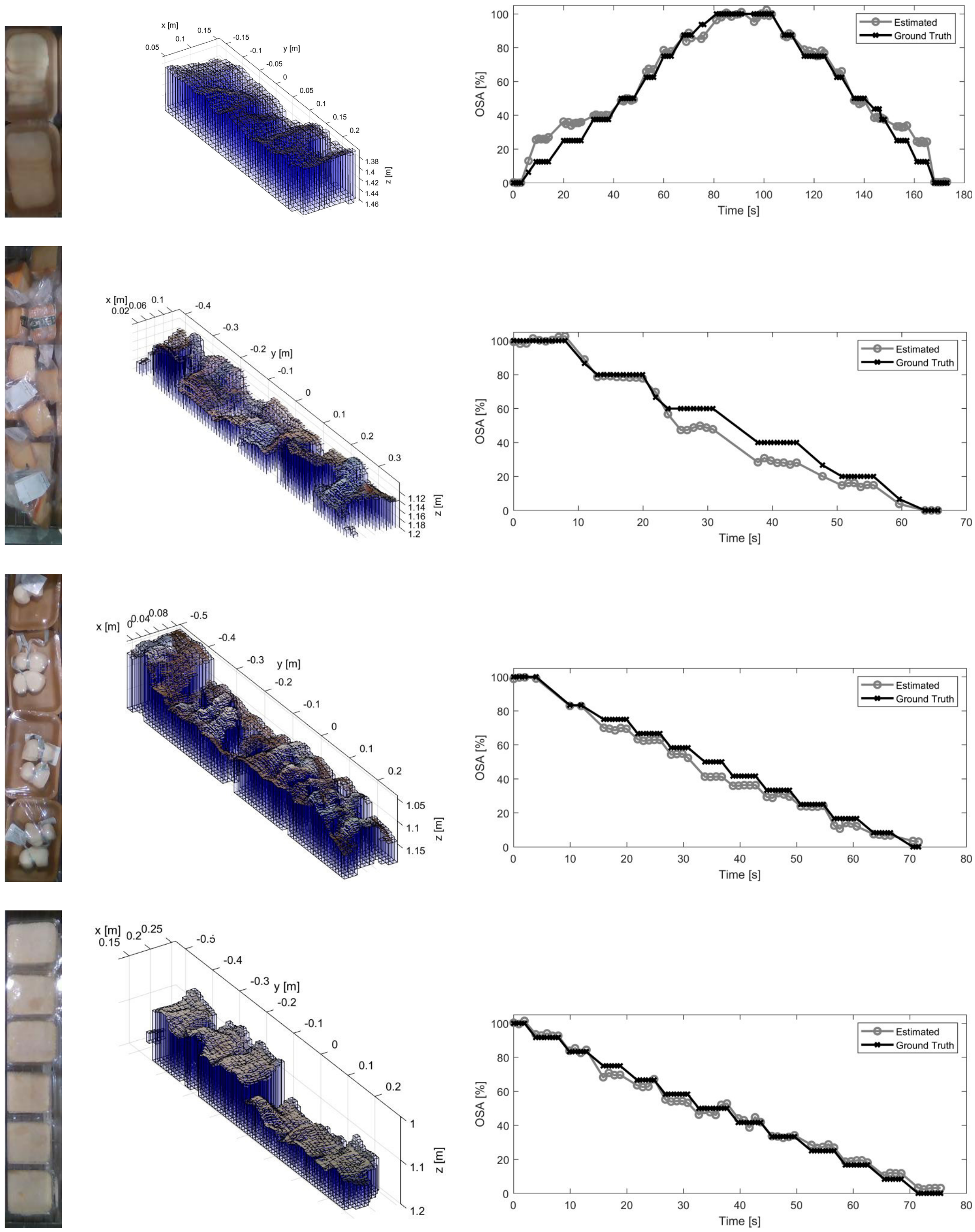


FIGURE 7. OSA estimation for different pre-packed cheese varieties. From top to bottom: Silano, Rodez, Scamorza Bianca, Grated Parmesan. From left to right: color images, occupancy grid-based product modelling and OSA estimates (gray) compared with ground-truth (black).

respectively. Similar results were obtained for all analysed product categories.

B. OSA ESTIMATION

Results of tests carried out on cheese counters are reported in Fig. 7. Four types of pre-packed cheese are analysed, namely (from top to bottom) Silano, Rodez, Scamorza Bianca and grated Parmesan. For every product, the first two columns show a sample image and the corresponding occupancy grid-based product modelling, respectively. In the third column, the result of the OSA estimation algorithm (gray line) in emptying and/or refilling sequences is compared with the ground truth (black line). Specifically, in each test, the shelf box is gradually refilled starting from an OOS condition ($OSA_{\%} = 0$) or emptied after full replenishment ($OSA_{\%} = 100$). One every sixth frames is processed and the estimated OSA level is compared with the actual value. The average discrepancy $\bar{E}$, the standard deviation σ and the

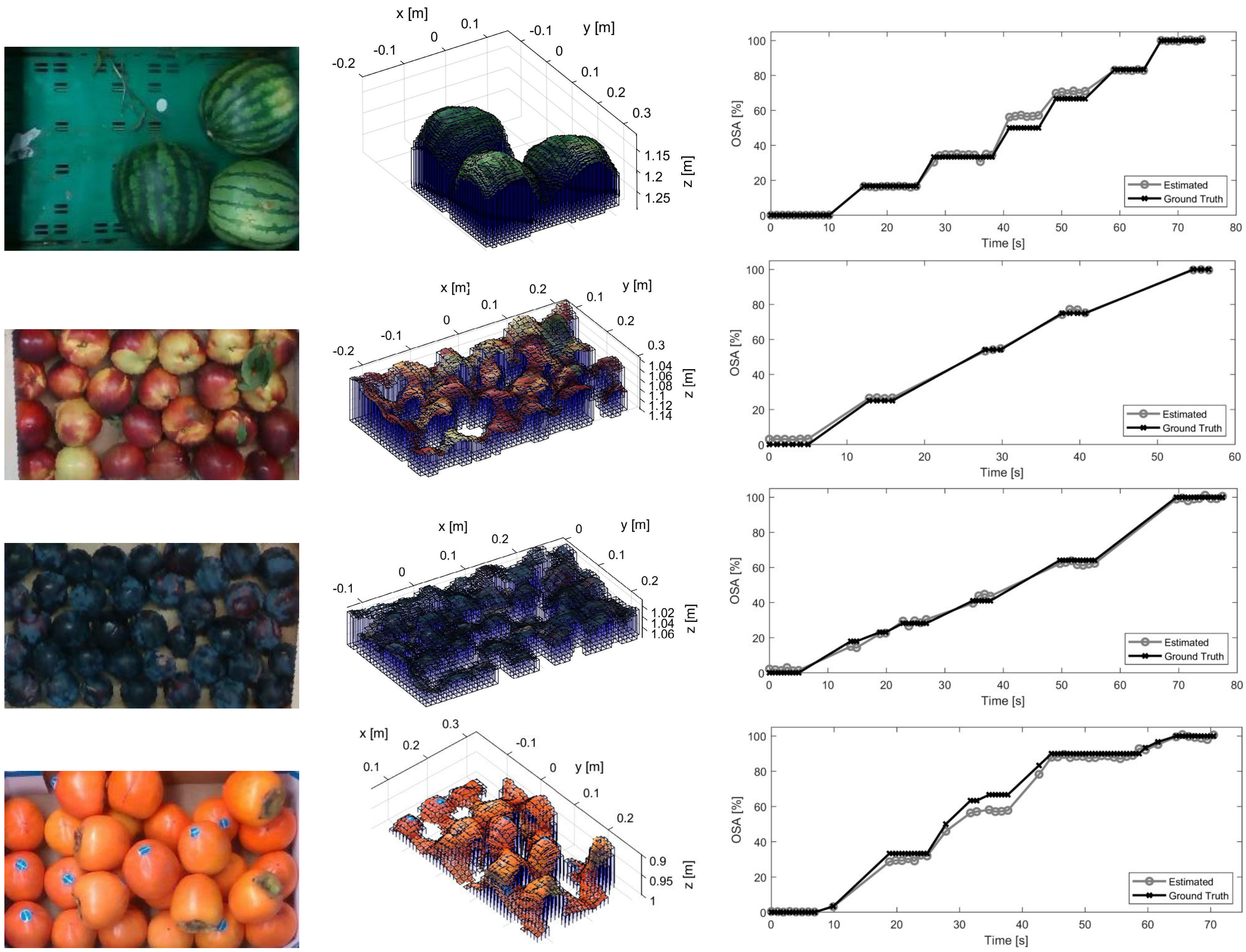


FIGURE 8. OSA estimation for different fruit varieties. From top to bottom: watermelons, peaches, plums and persimmons. From left to right: color images, occupancy grid-based product modelling and OSA estimates (gray) compared with ground-truth (black).

(a) (b) (c)

FIGURE 9. Example of OSA estimation for pre-packed artichokes on a vertical shelf (red rectangle), using a forward-looking camera configuration. In (a) the system reports a full product availability with an OSA estimate of 99.5%. However, once the foremost product row has been removed (b), it appears that only a few misplaced product units are actually available. The indication of the OSA level can be therefore misleading in such cases due to occlusions. Nevertheless, the system can be still effectively used for OOS detection, as in the case of figure (c), where all product items have been removed from the shelf.

r-squared coefficient are then computed. Numerical results for all tests are collected in Table 1. It can be seen that a maximum average discrepancy of about 5.0% is obtained for the Rodez cheese, which for a maximum number of products in the shelf box $N_{max} = 15$ corresponds to an average error of 0.75 (i.e., less than 1 product) over the estimated number of available items. An r-squared coefficient higher than 0.98 is reached for all tests showing a good agreement between estimated and ground truth OSA values.

Experiments carried out for different fruit types, namely (from top to bottom) watermelons, peaches, plums and persimmons, are shown in Fig. 8. Starting from an OOS situation,

TABLE 1. OSA estimation for different cheese types: error analysis. Average error $\bar{E}$, standard deviation σ and r-squared coefficient between estimated and actual OSA percentage.

Product Type (N_{max})	$\bar{E}$ [%]	σ [%]	r-squared
Silano (16)	3.991	4.228	0.981
Rodez (15)	5.007	4.591	0.981
Scamorza (12)	3.675	2.343	0.991
Grated Parmesan (12)	2.453	1.506	0.992

TABLE 2. OSA estimation for different fruit types: error analysis. Average error $\bar{E}$, standard deviation σ and r-squared coefficient between estimated and actual OSA percentage.

Product Type (N_{max})	$\bar{E}$ [%]	σ [%]	r-squared
Watermelon (6)	1.514	2.141	0.995
Peach (24)	1.594	1.164	0.999
Plum (39)	1.653	0.941	0.998
Persimmon (30)	2.488	2.592	0.994

the boxes are gradually replenished reaching the condition of full product availability. Numerical results are reported in Table 2. A maximum average discrepancy of about 2.5% is reached for Persimmons, which for a maximum number of available products $N_{max} = 30$ corresponds to an average error of 0.75 (i.e., less than 1 product) over the estimated number of available items. An r-squared coefficient higher than 0.99 is reached for all tests.

As previously noticed, the proposed system is independent of the orientation of the shelf in space and in particular it can be adopted for vertical shelves. A sample sequence is shown in Fig. 9 for pre-packed artichokes. Initially, the system estimates a product availability of 99.5% (Fig. 9(a)), since, from a frontal camera view, the shelf appears as fully replenished. However, after removal of the foremost product row, it appears that only a few misplaced product packages are actually available (Fig. 9(b)), showing that for vertical shelves, the indication of the OSA level may be misleading due to occlusions. Nevertheless, the system can be still successfully adopted for early OOS detection, as shown in the frame of Fig. 9(c).

V. DISCUSSION AND CONCLUSION

A novel framework to estimate online the on-shelf availability of products in a retail environment, based on 3D data provided by an Intel RealSense D435, is proposed. The system is intended to early detect out-of-stock events, as well as to provide continuously updated information on product availability for e-commerce apps and stock inventory purposes. Experimental results obtained in a real point of sale are presented to demonstrate the performance of the proposed approach. It is shown that the system is able to accurately estimate the on-shelf availability percentage of different types of fresh products with a maximum average discrepancy between estimated and actual OSA percentage of about 5.0% with respect to the actual one and an absolute error over the estimated number of products less than one unit for all the tested item categories. The coefficient of determination considering manually measured and estimated OSA levels appears also high, with values higher than 0.98 in all test cases. Overall, it is shown that the proposed framework is effective for online estimation of product on-shelf availability in a non-invasive and automatic way.

The system exploits 3D sensory data acquired by a consumer-grade RGB-D sensor (worth a few hundred Euros) as the only input. This proved to be a good trade-off between performance and cost-effectiveness. Obvious improvements would derive from the adoption of higher-end depth cameras available on the market, ensuring more accurate and less noisy depth estimation. However, vision-based approaches typically suffer from errors due to occlusions and under/over exposition. An optimal placement of visual sensors and lighting sources (such as shelf lighting) would be beneficial to maximize the camera coverage with good illumination conditions. The integration of complementary sensor types, like load or IR sensors, can be also considered to enhance the overall system reliability. Compared to most of previous approaches, one advantage of the proposed framework is that it does not require *a priori* knowledge of shelf or product characteristics, whereas the only assumption is that the shelf can be modelled as a planar surface whose parameters are learnt via an initial automatic calibration phase. In the current system implementation, the segmentation of different product categories within the camera field of view is based on manual segmentation of the RGB image returned by the sensor. However, the adoption of automated image segmentation techniques is under investigation to further reduce the need of human inputs. Further work will regard the use of RGB information in addition to 3D data for tasks such as automated product identification, detection of misplaced products or product occlusions and planogram compliance verification in general. Future research objectives will also deal with the integration of the proposed method on-board mobile platforms such as mobile robots or smart shopping carts to reduce the need for environment infrastructuring and increase the overall system flexibility.

ACKNOWLEDGMENT

The financial support of the grant Electronic Shopping & Home delivery of Edible goods with Low environmental Footprint (E-SHELF), POR Puglia FESR-FSE 2014–2020 (Grant Id. OSW3NO1) is gratefully acknowledged. The authors are thankful to Primo Prezzo (Carelli) for providing test facilities and valuable support in performing the experiments. The administrative and technical support by Vito Micunco (Software Design), Michele Attolico (CNR-STIIMA) and Giuseppe Bono (CNR-STIIMA) is also gratefully acknowledged.

REFERENCES

[1] T. W. Gruen. *A Comprehensive Guide to Retail Out-of-Stock Reduction in the Fast-Moving Consumer Goods Industry a Research Study Conducted*. Accessed: Jun. 7, 2019. [Online]. Available: https://www.semanticscholar.org/paper/A-Comprehensive-Guide-To-Retail-Out-of-Stock-In-the-Gruen/19280a78f21294dabcbe300c71791f0b7e9bf9bf

[2] R. Berger. *Optimal Shelf Availability: Increasing Shopper Satisfaction at the Moment of Truth*. Accessed Jun. 7, 2019. [Online]. Available: http://ecr-community.org/wp-content/uploads/2016/10/ecr-europe-osa-optimal-shelf-availability.pdf

[3] T. Gruen and D. Corsten, "Rising to the challenge of out-of-stocks," *ECR J.*, vol. 2, pp. 45–58, Dec. 2002.

[4] J. Aastrup and H. Kotzab, "Forty years of out-of-stock research—And shelves are still empty," *Int. Rev. Retail, Distrib. Consum. Res.*, vol. 20, pp. 147–164, Feb. 2010.

[5] M. Bertolini, G. Ferretti, G. Vignali, and A. Volpi, "Reducing out of stock, shrinkage and overstock through RFID in the fresh food supply chain: Evidence from an Italian retail pilot," *Int. J. RF Technol.*, vol. 4, no. 2, pp. 107–125, 2013.

[6] E. Frontoni, A. Mancini, and P. Zingaretti, "Embedded vision sensor network for planogram maintenance in retail environments," *Sensors*, vol. 15, no. 9, pp. 21114–21133, Aug. 2015.

[7] B. Hardgrave, S. Langford, M. Waller, and R. Miller, "Measuring the impact of RFID on out of stocks at Wal-Mart," *MIS Quart. Executive*, vol. 7, no. 4, pp. 181–192, Dec. 2008.

[8] E. Frontoni, A. Mancini, P. Zingaretti, and V. Placidi, "Information management for intelligent retail environment: The shelf detector system," *Information*, vol. 5, no. 2, pp. 255–271, May 2014.

[9] C. Metzger, J. Meyer, E. Fleisch, and G. Tröster, "Weight-sensitive foam to monitor product availability on retail shelves," in *Pervasive Computing*, A. LaMarca, M. Langheinrich, and K. N. Truong, Eds. Berlin, Germany: Springer, 2007, pp. 268–279.

[10] R. Moorthy, S. Behera, and S. Verma, "On-shelf availability in retailing," *Int. J. Comput. Appl.*, vol. 115, no. 23, pp. 47–51, Apr. 2015.

[11] R. Moorthy, S. Behera, S. Verma, S. Bhargave, and P. Ramanathan, "Applying image processing for detecting on-shelf availability and product positioning in retail stores," in *Proc. 3rd Int. Symp. Women Comput. Informat. (WCI)*, 2015, pp. 451–457.

[12] L. Rosado, J. Goncalves, J. Costa, D. Ribeiro, and F. Soares, "Supervised learning for out-of-stock detection in panoramas of retail shelves," in *Proc. IEEE Int. Conf. Imag. Syst. Techn. (IST)*, Oct. 2016, pp. 406–411.

[13] A. Saran, E. Hassan, and A. K. Maurya, "Robust visual analysis for planogram compliance problem," in *Proc. 14th IAPR Int. Conf. Mach. Vis. Appl. (MVA)*, May 2015, pp. 576–579.

[14] M. Paolanti, M. Sturari, A. Mancini, P. Zingaretti, and E. Frontoni, "Mobile robot for retail surveying and inventory using visual and textual analysis of monocular pictures based on deep learning," in *Proc. Eur. Conf. Mobile Robots (ECMR)*, Sep. 2017, pp. 1–6.

[15] R. Satapathy, S. Prahlad, and V. Kaulgud, "Smart shelfie—Internet of shelves: For higher on-shelf availability," in *Proc. IEEE Region Symp.*, May 2015, pp. 70–73.

[16] A. Milella, R. Marani, A. Petitti, and G. Reina, "In-field high throughput grapevine phenotyping with a consumer-grade depth camera," *Comput. Electron. Agricult.*, vol. 156, pp. 293–306, Jan. 2019.

[17] P. Henry, M. Krainin, E. Herbst, X. Ren, and D. Fox, "RGB-D mapping: Using Kinect-style depth cameras for dense 3D modeling of indoor environments," *Int. J. Robot. Res.*, vol. 31, no. 5, pp. 647–663, Apr. 2012.

[18] G. Cicirelli and T. D'Orazio, "Gesture recognition by using depth data: Comparison of different methodologies," in *Motion Tracking and Gesture Recognition*. Rijeka, Croatia: InTech, Jul. 2017, ch. 6.

[19] D. Liciotti, M. Contigiani, E. Frontoni, A. Mancini, P. Zingaretti, and V. Placidi, "Shopper analytics: A customer activity recognition system using a distributed RGB-D camera network," in *Proc. Int. Workshop Video Anal. Audience Meas. Retail Digit. Signage*, vol. 8811, Aug. 2014, pp. 146–157.

[20] *Smart Shelf. The Evolution of the Shelf*. Accessed: Sep. 18, 2019. [Online]. Available: https://www.intel.it/content/www/it/it/retail/digital-signage/intelligent-shelving.html

[21] P. Torr and A. Zisserman, "MLESAC: A new robust estimator with application to estimating image geometry," *Comput. Vis. Image Understand.*, vol. 78, no. 1, pp. 138–156, Apr. 2000.

[22] A. Grunnet-Jepsen, J. N. Sweetser, and J. Woodfill. *Best-Known-Methods for Tuning Intel RealSense D400 Depth Cameras for Best Performance*. Accessed: Dec. 13, 2019. [Online]. Available: https://dev.intelrealsense.com/docs/tuning-depth-cameras-for-best-performance

ANNALISA MILELLA was born on March 1979. She received the M.S. (summa cum laude) and Ph.D. degrees in mechanical engineering from the Politecnico of Bari, Italy, in 2002 and 2006, respectively. In 2005, she was a Visiting Ph.D. Student with the EPFL Autonomous Systems Laboratory. From 2006 to 2009, she worked as a Postdoctoral Researcher with the Institute of Intelligent Systems for Automation (ISSIA), National Research Council of Italy (CNR). From December 2009 to June 2018, she has been a Researcher with ISSIA. Since June 2018, she has been a Researcher with the Institute of Intelligent Industrial Systems and Technologies for Advanced Manufacturing (STIIMA), CNR. She is the Principal Investigator of the CNR Research Unit for the Project H2020 DT-ICT-08-2019 Agricultural Interoperability and Analysis System (ATLAS) and for the project Electronic Shopping & Home delivery of Edible goods with Low environmental Footprint (E-SHELF), POR Puglia FESR-FSE 2014–2020. From January 2016 to April 2018, she has been the Principal Investigator of the CNR Research Unit for the Project FP7 ERA-NET ICT-AGRI 2 Simultaneous Safety and Surveying for Collaborative Agricultural Vehicles (S3-CAV). She has authored more than 70 publications on international journals, conference proceedings, and book chapters.

ANTONIO PETITTI was born on August 1985. He received the B.S. and M.S. (Laurea Specialistica) degrees *(summa cum laude)* in automation engineering from the Politecnico di Bari, Italy, in 2008 and 2010, respectively, and the Ph.D. degree in 2015. From June 2011 to November 2018, he was a Research Fellow with the Institute of Intelligent Systems for Automation (ISSIA), Italian National Research Council (CNR), Italy. From December 2018 to September 2019, he served as a Fixed-Term Researcher with the Institute of Intelligent Industrial Systems and Technologies for Advanced Manufacturing (STIIMA), National Research Council. He was involved in the design of innovative estimation algorithms for robotic networks. Since October 2019, he has been a Permanent Researcher with CNR-STIIMA, where he cooperates with the Intelligent Sensing and Perception Group for distributed estimation and cooperative control of robotic networks, dedicated to industrial applications, precision agriculture, and smart retail systems.

ROBERTO MARANI was born on March 1985. He received the B.E., M.S., and Ph.D. degrees in 2006, 2008, and 2012. From January 2009 to December 2011, he worked with the Department of Electric and Information Engineering, Polytechnic University of Bari, where he carried out his research activity on optoelectronics, microelectronics, photonics, and plasmonics. From 2012 to 2018, he was a Research Fellow with the Institute of Intelligent Systems for Automation (Istituto di Studi sui Sistemi Intelligenti per l'Automazione - ISSIA), National Research Council (Consiglio Nazionale delle Ricerche - CNR), Bari, where he was involved in the design of hardware and software prototypes for autonomous systems. Since 2018, he has been a Permanent Staff Researcher of the Institute of Intelligent Industrial Technologies and Systems for Advanced Manufacturing (Istituto di Sistemi e Tecnologie Industriali Intelligenti per il Manifatturiero Avanzato - STIIMA), National Research Council, where he cooperates with the Intelligent Sensing and Perception Group for computer vision, machine and deep learning applications, dedicated to industrial applications, robotics, geology, archaeology, and reverse engineering. He is currently an Electronic Engineer with CNR-STIIMA. He has authored more than 80 scientific articles on international peer-reviewed journals and 40 proceedings of international conferences.

GRAZIA CICIRELLI received the Laurea degree *(summa cum laude)* in computer science from the University of Bari, Italy, in 1994. Until 2001, she held grants from the Italian National Research Council (CNR) for research activities in robotics and image processing. She is currently a Technologist Researcher with the Institute of Intelligent Industrial Technologies and Systems for Advanced Manufacturing (STIIMA), CNR, Bari, where she cooperates with the Intelligent Sensing and Perception Group. Her principal interests include computer vision and machine learning for different context applications such as pattern recognition, autonomous mobile robotics, industrial quality control, and intelligent systems for video-surveillance. She has authored numerous research articles published in international peer-reviewed journals and proceedings of international conferences.

TIZIANA D'ORAZIO received the degree *(summa cum laude)* in computer science from the University of Bari, Italy. Since 1997, she has been working at the Institute of Signal and Image Processing, Italian National Research Council (CNR). She is currently a Senior Researcher with the Institute of Intelligent Industrial Technologies and Systems for Advanced Manufacturing (STIIMA), CNR. She has authored more than 150 articles published in international journals and international conference proceedings, and coauthored three international patents on visual systems for event detection in sportive contexts. Her research interests include pattern recognition, artificial intelligence, image and signal processing for robotic application, and intelligent perception systems.