

Visual sensor fusion for active security in robotic industrial environments

Article

Published Version

Creative Commons: Attribution 3.0 (CC-BY)

Open Access

Robla, S., Llata, J. R., Torre-Ferrero, C., Sarabia, E. G., Becerra, V. and Perez-Oria, J. (2014) Visual sensor fusion for active security in robotic industrial environments. *Eurasip Journal on Advances in Signal Processing*, 2014. 88. ISSN 1110-8657 doi: <https://doi.org/10.1186/1687-6180-2014-88>
Available at <https://centaur.reading.ac.uk/37138/>

It is advisable to refer to the publisher's version if you intend to cite from the work. See [Guidance on citing](#).

Published version at: <http://asp.eurasipjournals.com/content/2014/1/88>

To link to this article DOI: <http://dx.doi.org/10.1186/1687-6180-2014-88>

Publisher: SpringerOpen

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the [End User Agreement](#).

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

Reading's research outputs online



RESEARCH

Open Access

Visual sensor fusion for active security in robotic industrial environments

Sandra Robla^{1*}, Jose R Llata¹, Carlos Torre-Ferrero¹, Esther G Sarabia¹, Victor Becerra² and Juan Perez-Oria¹

Abstract

This work presents a method of information fusion involving data captured by both a standard charge-coupled device (CCD) camera and a time-of-flight (ToF) camera to be used in the detection of the proximity between a manipulator robot and a human. Both cameras are assumed to be located above the work area of an industrial robot. The fusion of colour images and time-of-flight information makes it possible to know the 3D localization of objects with respect to a world coordinate system. At the same time, this allows to know their colour information. Considering that ToF information given by the range camera contains inaccuracies including distance error, border error, and pixel saturation, some corrections over the ToF information are proposed and developed to improve the results. The proposed fusion method uses the calibration parameters of both cameras to reproject 3D ToF points, expressed in a common coordinate system for both cameras and a robot arm, in 2D colour images. In addition to this, using the 3D information, the motion detection in a robot industrial environment is achieved, and the fusion of information is applied to the foreground objects previously detected. This combination of information results in a matrix that links colour and 3D information, giving the possibility of characterising the object by its colour in addition to its 3D localisation. Further development of these methods will make it possible to identify objects and their position in the real world and to use this information to prevent possible collisions between the robot and such objects.

Keywords: Active security; Industrial robot; ToF and colour cameras; Information fusion

1 Introduction

Since the 1960s, industrial robots have been used in the manufacturing industry and they have substituted humans in various repetitive, dangerous, hostile tasks. A consequence associated with the incorporation of robots in industry is the emergence of new risks of accidents for workers. The normatives which incorporate, among many other aspects, these robot-related risks, include the international standard ISO 10218, the American ANSI/RIA R15.06, the European EN 775, and national normatives such as the Spanish UNE-EN 755. To prevent accidents, the selection of a security system must be based on the analysis of these risks. Traditionally, these security systems separate the robot workspace from the human one. One example of this requirement was reflected in the Spanish normative UNE-EN 755:1996 [1]. It is established

that sensor systems have to be incorporated to prevent the entrance of humans in a hazardous area in case the operating state of the robotic system implies dangers to the human. According to traditional normatives, maintenance, repair, or programming personnel can only be inside the robot workspace if the industrial robot is not in automatic mode.

However, in recent years, due in part to the flexible design of products, the optimization of production methods, and the introduction of new technologies, the tasks performed by industrial robots are no longer restricted to the transfer of objects, or other repetitive tasks. Instead, there is an increasing number of tasks in which humans and robots combine their skills in collaborative work.

To enable collaboration between human and robot, safety measures that establish a rigid separation between human and robot workspaces have to be removed. Instead, the introduction of other types of security systems is required so that collisions can be avoided by detecting obstacles as well as their dynamic characteristics, and harm to the human can be mitigated in case of

*Correspondence: sandra.robla@unican.es

¹Electronic Technology and Automatic Systems Department, School of Industrial and Telecommunications Engineering, University of Cantabria, Avda. de Los Castros s/n, Santander 39005, Spain

Full list of author information is available at the end of the article

an unexpected impact. For this reason, research in this field is directed towards changing the way a human interacts with a robot so that the trend is that both human and robot can share the same workspace at the same time. This change in the working relationship is reflected in the updates carried out from the year 2006 in the international normatives ISO10218 [2] and guidelines for the implementation of these regulations, such as [3]. In these guidelines, new concepts are presented, such as collaborative robots, collaborative operations, and spaces of collaborative work.

Taking into account that security is a fundamental aspect in the design of robotic manufacturing systems, the development of systems and security strategies that allow safe collaborative work between human and robot is essential. The aim of this paper is to contribute at the initial stage of the design of a system for collision prevention between a human and a robot manipulator sharing a workspace at the same time. A method for processing of information acquired from two different types of vision sensors located above an industrial robot environment is proposed. The method, which is mainly focused on information captured from a time-of-flight camera, allows the fusion of both colour and 3D information, as an initial step towards the development of an active security system for application in an industrial robotics environment. This information fusion generates a colour and 3D information matrix which allows simultaneously estimating colour characteristics from an object and its three-dimensional position in a world coordinate frame. At a later step, the use of this combination of information

will allow to associate a security volume around each characterised object, in order to prevent possible collisions between industrial robot and human.

2 Related work on shared human robot workspaces

A brief summary of different types of security applied to industrial robotic environments is provided in order to give the context to the work presented in this paper. With the aim of giving context to the work presented in this paper, Figure 1 presents a possible classification of these types of security, as well as goals to achieve for each type of security, systems and devices used, and actions to apply on the robotic system.

Security systems in industrial robotic environments can be classified as passive and active. Passive security systems are hazard warning elements which do not alter the robot behaviour. These systems are audible or visible signals such as alarms or lights or systems that prevent the inadvertent access to a restricted area. Active security systems in industrial robotic environments can be defined as the methods used to prevent the intrusion of humans to the robot workspace when it is in automatic mode. The difference with the passive methods is that active methods can modify the robot behaviour. Historically, devices such as movement, proximity, force, acceleration, or light sensors are used to detect human access to the robot workspace and to stop the execution of the robot task. However, as it has been discussed previously, research in this field is moving towards allowing humans and robots to share workspaces.

TYPES	PRINCIPAL AIM	SECONDARY AIM	SYSTEMS	DEVICES	ACTIONS
PASSIVE SECURITY SYSTEMS	SEPARATED HUMAN AND ROBOT WORKSPACES	HUMAN ACTIONS RESTRICTED	WARNING SIGNALS	Optical, acoustic, ligh, signals	No actions
		ROBOT BEHAVIOUR MODIFICATION	ACCESS RESTRICTED	Fences, chains	Robot stop/reduction of velocity
ACTIVE SECURITY SYSTEMS	SHARED HUMAN AND ROBOT WORK / WORKSPACES	COLLISION AVOIDANCE	Combination passive and active security systems	Interlocking devices. Proximity, tactile sensors	Robot stop/ reduction of velocity/ motion planing
			Sensors	Capacitive, ultrasonic, laserscaner sensors	
			Artificial vision systems	One/Several cameras Standard/Fisheye	
			Range systems	ToF laser sensor One/ several ToF cameras	
Combination of vision and range systems	Standard CCD and TOF cameras One/several RGB-D devices				

Figure 1 Security systems industrial robot environments. This scheme aims to summarize a classification of the types of security applied to industrial robotic environments, as well as goals to achieve for each type of security, systems and devices used, and the action to apply on the robotic system.

2.1 Collision avoidance

A further way to enhance safety in shared human/robot work/workspaces is to implement collision avoidance systems. Robots have been provided with sensors capturing local information. Ultrasonic sensors [4], capacitive sensors [5,6], and laser scanner systems [7] have been tried to avoid collisions. However, the information provided by these sensors does not cover the whole scene, and so these systems can only provide a limited contribution to enhance safety in human-robot collaboration tasks [8]. Moreover, geometric representations of human and robotic manipulators have been used to obtain a spatial representation in human-robot collaboration tasks. Numerical algorithms are then used to compute the minimum distance between human and robot and to search for collision-free paths [9-12]. Methods have been proposed involving the combination of different types of devices to help avoid collisions. This idea has been applied into a cell production line for component exchange between human and robot in [13], where the safety module uses commands from light curtain sensors, joint angle sensors, and a control panel to prevent the collision with the human when exchanging an object. The discussion concentrates below in artificial vision systems, range systems, and their combination.

2.1.1 Artificial vision systems

Artificial vision systems have also been used to prevent human-robot collisions. This information can be used on its own or in the combination with information from other types of devices. In order to achieve safe human-robot collaboration, [14] describes a safety system made up of two modules. One module is based on a camera and computer vision techniques to obtain the human location. The other module, which is based on accelerometers and joint position information, is used to prevent an unexpected robot motion due to a failure of robot hardware or software. Research work such [15] investigates safety strategies for human-robot coexistence and cooperation. The use of a combination of visual information from two cameras and information from a force/torque sensor is proposed. In order to perform collision tests, other work has used visual information acquired by cameras [16,17] to generate a 3D environment. Also, visual information is used to separate humans and other dynamic unknown objects from the background [18] or to alter the behaviour of the robot [19]. In [20-22], visual information has been used to develop safety strategies based on fuzzy logic, probabilistic methods, or the calculation of warning index, respectively.

2.1.2 Range systems

The depth map of a scene can be obtained by using depth sensors such as laser range finders and stereo camera

systems. The results of using a laser time-of-flight (ToF) sensor are presented in [23] and [24] with the latter using several depth sensors in combination with presence sensors. Recently, a new type of camera has become available. These cameras, denominated as range-imaging cameras, 3D ToF cameras, or PMD cameras, capture information providing a 3D point cloud, among other information. They are starting to be used in active security systems for robotic industrial environments, among other applications. An example is a single framework for human-robot cooperation whose purpose is to achieve a scene reconstruction of a robotic environment by markerless kinematic estimation. For example, [8,25] use the information delivered by a 3D ToF camera mounted to the top of a robotic cell. This information is employed with the purpose of extracting robust features from the scene, which are the inputs to a module that estimates risks and controls the robot. In [26], the fusion of 3D information obtained from several range imaging cameras and the application of the visual hull technique are used to estimate the presence of obstacles within the area of interest. The configurations of a robot model and its future trajectory along with information on the detected obstacles are used to check for possible collisions.

2.1.3 Combination of vision and range systems

This technique is based on the combination of 3D information from range cameras and 2D information from standard charge-coupled device (CCD) cameras. Although this technique is being used in other applications, such as hand following [27,28] or mixed reality applications [29-31], not much work has been reported using this technique in the area of active security in robotic environments. In [32], an analysis of human safety in cooperation with a robot arm is performed. This analysis is based on information acquired by a 3D ToF camera and a 2D/3D Multicam. This 2D/3D Multicam consists of a monocular hybrid vision system which fuses range data from a PMD ToF sensor, with 2D images from a conventional CMOS grey scale sensor. The proposed method establishes that while the 3D ToF camera monitors the whole area, any motion in the shared zones is analysed using the 2D/3D information from the Multicam. In [33], a general approach is introduced for surveillance of robotic environments using depth images from standard colour cameras or depth cameras. The fusion of data from CCD colour cameras or from ToF cameras is performed to obtain the object hull and its distance with respect to the known geometry of an industrial robot. They also present a comparison between distance information from colour and ToF cameras and a comparison between a ToF camera and ToF information fusion. One of the conclusions of this work is that the fusion of information from several ToF cameras provides better resolution and less noise than

the information obtained from a single camera. Finally, [34] describes a hybrid system based on a ToF camera and a stereo camera pair which is proposed to be applied in human-robot collaboration task. Stereo information is used in unreliable ToF data points to generate a depth map which is fused with the depth map from the ToF camera. Colour feature is not taken into account. On the other hand, nearly a decade after that ToF cameras emerged into the industrial trade [35], a new type of 3D sensors (RGB-D sensors), which are fitted with a RGB camera and a 3D depth sensor, were launched for non-commercial use [36]. The RGB-D sensor has several advantages over ToF cameras such as higher resolution, lower price, and the availability of depth and colour information. Hence, its study and application have been objective of research work such as [37] that presents a review of Kinect-based computer vision algorithms and applications. Several topics are presented like preprocessing tasks including a review of Kinect recalibration techniques, object tracking and recognition, and human activity analysis. These authors propose in [38] an adaptive learning methodology to extract spatio-temporal features, simultaneously fusing the RGB and depth information. In addition to this, a review of several solutions to carry out information fusion of RGB-D data is presented. Also, a website for downloading a dataset made of RGB and depth information for hand gesture recognition is introduced. Related to active security system in industrial robotic environments, the use of the Kinect sensor is being incorporated as it is shown in [39] where a real-time collision avoidance approach based on this sensor is presented.

3 Method for the fusion of colour and 3D information

The presented method for fusion of acquired information from a ToF camera and a colour camera has a different standpoint from the ones proposed in the consulted papers. According to papers that are not related to active security in robotic industrial environments such as [27], the spatial transformation is performed establishing the ToF camera coordinate system as the reference coordinate system. Therefore, if an object position in a world coordinate system wanted to be known, another calibration should be done to establish the rotation matrix and translation vector that connected both coordinate systems. Nevertheless, in the present paper, this aspect has been considered. Therefore, it was needed to define a common coordinate system for an industrial robot, a colour camera, and a ToF camera, in order to know at the same time 3D object location at the robot arm workspace and its colour feature. According to papers focusing on mixed reality applications as paper [29], the used setup includes a CCD firewire camera, a ToF camera, and a fisheye camera. After performing the calibration and establishing

relative transformations between the different cameras, a background model, whose use eliminates the need for chroma keying and also supports planning and alignment of virtual content, was generated allowing to segment the actor from the scene. Paper [31] presents a survey of ToF basic measurement principles of ToF cameras including, among other issues, camera calibration, range image preprocessing, and sensor fusion. Several studies which study different combinations of high-resolution cameras and lower-resolution ToF cameras are mentioned.

In relation to the paper focused on active security, the most closely related to our work is [32]. Though a common world coordinate system for cameras and robot is also used, the method seem to present certain differences because a spatial transform function is identified in order to map the image coordinates of the 2D sensor to the corresponding coordinates of the PMD sensor. Moreover, saturated pixels errors do not seem to have been considered. Here, the presented work shows a different standpoint since the obtained parameters from the cameras calibration are used to transform 3D point cloud given in the ToF camera coordinate system to the world coordinate system, and finally, the obtained internal and external parameters are used to achieve the reprojection of corrected 3D points (distance error, saturated pixels, and jump edge effect) into colour images.

With the aim of allowing any researcher to implement the proposed method of fusion of information exactly like it that has been carried out at the present work, this paper gives a mathematical detailed description of the steps involved in the proposed method.

In what follows, it is assumed that a 3D ToF camera and a colour camera are fixed and placed over the workspace of a robot arm and that the fields of view of both cameras are overlapped. Also, it is assumed that external temperature conditions are constant, and that the integration time parameter of the 3D ToF camera is automatically updated at each data acquisition. Image and 3D data from the scene is captured and processed as described in the next sub-sections. Assume that the ToF camera has a resolution $n_x \times n_y$ and that the CCD camera has a resolution $\hat{n}_x \times \hat{n}_y$.

In what follows, vectors and matrices are denoted by Roman bold characters (e.g. \mathbf{x}). The j th element of a vector \mathbf{x} is denoted as x_j , element (i, k) of a matrix \mathbf{A} is denoted as $A_{i,k}$, a super-index in parenthesis (j) denotes a node within a range of distances, a sub-index within square brackets such as $[i]$ denotes an element of a set.

3.1 Reduction TOF range camera errors

The reduction of range camera errors is a fundamental step to achieve an acceptable fusion of colour and 3D information. The existence of these errors cause the fused information to have issues that range from minor,

such as border inaccuracy, to serious such as the loss of information in saturated pixels coordinates.

3.1.1 Distance error reduction

As it is well documented that ToF cameras suffer from a non-linear distance error, several experiments have been developed in order to model and correct the distance error (or circular error) [35,40-43]. With the purpose of decreasing the influence of this error in distance measurements, a procedure is described below to correct the ToF distance values based on a study of the the behaviour of the camera. This study requires a ToF camera to be positioned parallel to the floor, and a flat panel of light colour and low reflectance, to be mounted on a robot arm. The panel position is also parallel to the floor. The robot arm allows to displace the panel along a distance range and ToF data at different distances can be captured.

The distance error analysis from the acquired data can be performed in two ways: a global analysis of all the pixels without taking pixel position into account and an analysis which takes into account the position of each pixel. The first analysis is easier to perform as it only requires a relatively small panel; it is assumed that there is no error due to pixel localization and only a reduced region of the 3D ToF data is analysed. The second analysis can be carried out to check the suitability of the assumption of negligible error due to pixel localization of the first analysis. The second analysis requires a larger panel, as the distance image captured by the camera has to be based only on the panel for different distances. Both methods are described in the steps below.

1. *Image capture.* Since distance measurements are influenced by the camera internal temperature, a minimum time period is necessary to obtain stable measurements [43]. After the camera warms up, ToF information is captured at each of the P different nodes in which the distance range D was divided. Each captured data is defined by an amplitude matrix \mathbf{A} of dimensions $n_x \times n_y$, and 3D information made up of three coordinates matrices \mathbf{X} , \mathbf{Y} , and \mathbf{Z} , each one of dimensions $n_x \times n_y$. In order to generate a model of distance error, a set $\mathcal{Z}_{\mathcal{T}}^{(j)} = \{\mathbf{Z}_{T[1]}^{(j)}, \mathbf{Z}_{T[2]}^{(j)}, \dots, \mathbf{Z}_{T[N]}^{(j)}\}$ of distance information in the z axis is formed by capturing N images at each node j , with $j = 1, \dots, P$. Similarly, sets of distance information for training are defined for the x and y axes, which are denoted as $\mathcal{X}_{\mathcal{T}}^{(j)}$ and $\mathcal{Y}_{\mathcal{T}}^{(j)}$, respectively. In order to validate the model so obtained, a set $\mathcal{Z}_{\mathcal{V}}^{(j)} = \{\mathbf{Z}_{V[1]}^{(j)}, \mathbf{Z}_{V[2]}^{(j)}, \dots, \mathbf{Z}_{V[M]}^{(j)}\}$ of distance information is also formed by capturing M additional images at each node j , with $j = 1, \dots, P$. Similarly, sets of distance information for validation

are defined for the x and y axes, which are denoted as $\mathcal{X}_{\mathcal{V}}^{(j)}$ and $\mathcal{Y}_{\mathcal{V}}^{(j)}$, respectively. In this article, the sets of information $\mathcal{Z}_{\mathcal{T}}$ and $\mathcal{Z}_{\mathcal{V}}$ are also called *ToF distance images* and are defined as $\mathcal{Z}_{\mathcal{T}} = \{\mathcal{Z}_{\mathcal{T}}^{(1)}, \dots, \mathcal{Z}_{\mathcal{T}}^{(P)}\}$, and $\mathcal{Z}_{\mathcal{V}} = \{\mathcal{Z}_{\mathcal{V}}^{(1)}, \dots, \mathcal{Z}_{\mathcal{V}}^{(P)}\}$.

2. *Angle correction.* Correction angles are applied to the ToF information sets for each axis x , y , and z , with the aim of compensating for any 2D angular deviation between the the (x, y) plane of the range camera and the plane defined by the floor. This 2D angular deviation is denoted by the angles θ_x and θ_y . This correction allows obtaining parameter values as if both camera and panel were perfectly parallel. Given an x axis distance image \mathbf{X}_T of dimensions $n_x \times n_y$, define its sub-matrix $\hat{\mathbf{x}}$ of dimensions $n_1 \times n_2$, where $n_1 < \text{int}(n_x/2)$ and $n_2 < \text{int}(n_y/2)$, as a matrix formed such that its top left element $\hat{x}_{1,1}$ corresponds to element $X_{T_{i_c j_c}}$. Index i_c is chosen as $\text{int}(n_x/2)$, and index j_c is chosen as $\text{int}(n_y/2)$. Similarly, sub-matrices $\hat{\mathbf{y}}$ and $\hat{\mathbf{z}}$ are defined for axes y and z , respectively. Define $\bar{\mathbf{x}}$, $\bar{\mathbf{y}}$, and $\bar{\mathbf{z}}$ as the column-wise vectorised forms of sub-matrices $\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$, $\hat{\mathbf{z}}$, each with dimension $n \times 1$, where $n = n_1 n_2$, with n as the number of pixels from the selected area. This central region is taken from each ToF distance image to estimate and correct the 2D angle inclination between the panel and the ToF camera. Hence, for each image region, 3D points are modified using the rotation matrices \mathbf{R}_x and \mathbf{R}_y :

$$\mathbf{R}_x = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \theta_x & -\sin \theta_x \\ 0 & \sin \theta_x & \cos \theta_x \end{bmatrix} \quad (1)$$

$$\mathbf{R}_y = \begin{bmatrix} \cos \theta_y & 0 & \sin \theta_y \\ 0 & 1 & 0 \\ -\sin \theta_y & 0 & \cos \theta_y \end{bmatrix}$$

such that

$$\mathbf{G} = \mathbf{R}_y[\bar{\mathbf{x}} \bar{\mathbf{y}} \bar{\mathbf{z}}]^T \quad (2)$$

where \mathbf{G} has dimensions $3 \times n$. The transformed image region for the z coordinate is obtained from the rows of \mathbf{G} :

$$\bar{z}'_k = G_{3,k}, \quad k = 1, \dots, n \quad (3)$$

and in this way, a vector $\bar{\mathbf{z}}'$ of dimensions $n \times 1$ is defined.

A second rotation transformation is applied around the y axis such that

$$\mathbf{H} = \mathbf{R}_x[\bar{\mathbf{x}} \bar{\mathbf{y}} \bar{\mathbf{z}}']^T \quad (4)$$

The transformed image region for the y coordinate is obtained from the rows of \mathbf{H} :

$$\bar{z}''_k = H_{3,k}, k = 1, \dots, n \quad (5)$$

where $\bar{\mathbf{z}}''$ is of dimension $n \times 1$. Since the above rotation causes a displacement of the 3D points along the y axis, the $\bar{\mathbf{y}}$ vector is used to represent ToF information after angle correction. Then, in this way, the 3D ToF vectors after angle correction are $\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\mathbf{z}}''$, each one of dimensions $n \times 1$.

3. If the pixel position is not considered, then:

- (a) *Discrepancy curve calculation stage.* In order to test the angle correction effect over the distance error, the same procedure is applied using data before and after angle correction. However, the method is described using data after angle correction. The selected area is used to calculate several parameters including the mean distance value, discrepancy distance value, and mean squared error (MSE). Define a set of distances after angle correction $\bar{\mathbf{z}}''^{(j)} = \{\bar{z}''_{[1]}^{(j)}, \bar{z}''_{[2]}^{(j)}, \dots, \bar{z}''_{[N]}^{(j)}\}$ at each node j , with $j = 1 \dots P$. The mean distance ToF over the selected area in all ToF distance images, $\bar{\mathbf{Z}}_j$, at each node j , is calculated by means of:

$$\bar{\mathbf{Z}}_j = \frac{1}{nN} \sum_{i=1}^N \sum_{k=1}^n \bar{z}''_{[i]k}^{(j)} \quad (6)$$

where the resulting $\bar{\mathbf{Z}}$ is a vector with dimensions $P \times 1$.

Defining L_j as a distance value obtained by a laser distance meter at each node j (henceforth this value is treated as ground truth), and a vector $\mathbf{L} = [L_1, \dots, L_P]^T$, with dimensions $P \times 1$. Then, the discrepancy distance vector, $\delta_{\mathbf{d}}$, is calculated as the difference between the mean distance from the ToF camera after angle correction, $\bar{\mathbf{Z}}$, and the ground truth vector \mathbf{L} :

$$\delta_{\mathbf{d}} = \bar{\mathbf{Z}} - \mathbf{L} \quad (7)$$

In order to obtain correction values to be applied in new ToF distances images, a cubic spline is used for fitting this discrepancy information for each distance. The cubic spline is modelled as a function s that passes through all the points $(\bar{\mathbf{Z}}, \delta_{\mathbf{d}})$ and at each interval $[\bar{\mathbf{Z}}_j, \bar{\mathbf{Z}}_{j+1}]$ and is expressed as a polynomial.

$$s(\bar{\mathbf{Z}}) = a_0 + a_1(\bar{\mathbf{Z}} - \bar{\mathbf{Z}}_j) + a_2(\bar{\mathbf{Z}} - \bar{\mathbf{Z}}_j)^2 + a_3(\bar{\mathbf{Z}} - \bar{\mathbf{Z}}_j)^3 \quad (8)$$

where $j = 1, \dots, P - 1$. For each sub-interval, the coefficients a_0, a_1, a_2, a_3 are calculated so

that the curve passes through the points $(\bar{\mathbf{Z}}_j, \delta_{d_j})$ and $(\bar{\mathbf{Z}}_{j+1}, \delta_{d_{j+1}})$ [44]. The resulting spline, henceforth called the *discrepancy curve*, allows to estimate the discrepancy correction value, given a ToF distance.

- (b) *Discrepancy correction.* In order to reduce the errors in the distance estimates obtained from the ToF information, the set of ToF distance images for validation \mathbf{Z}_V is used to validate the discrepancy curve. To this end, a vector of validation ToF distance images after angle correction $\bar{\mathbf{z}}''_V$ (dimension $n \times 1$) is defined and evaluated on the discrepancy curve to obtain the vector of correction values \mathbf{C} (dimension $n \times 1$). Then, the corrected distance value for a distance image after its angle correction is calculated as follows:

$$\hat{\mathbf{z}} = \bar{\mathbf{z}}''_V - \mathbf{C} \quad (9)$$

Define $\hat{\mathbf{z}}^{(j)} = \{\hat{z}_{[1]}^{(j)}, \hat{z}_{[2]}^{(j)}, \dots, \hat{z}_{[M]}^{(j)}\}$ as a set of distances after discrepancy correction for each node j , with $j = 1 \dots P$, the mean value after discrepancy correction for the M ToF distance images obtained at each j node is calculated as follows:

$$\hat{\mathbf{z}}_j = \frac{1}{nM} \sum_{i=1}^M \sum_{k=1}^n \hat{z}_{[i]k}^{(j)} \quad (10)$$

with $j = 1 \dots P$ and where the resulting $\hat{\mathbf{Z}}$ is a vector with dimensions $P \times 1$.

In order to observe the effect that these corrections have over the 3D ToF points, the MSE can be calculated before and after the discrepancy correction. Defining for each node j a vector with the corresponding laser distance meter values $\mathbf{L}^{(j)} = [L^{(j)}, \dots, L^{(j)}]^T$ with dimension $n \times 1$ (treated here as ground truth), then the mean squared error at each pixel k and for node j can be calculated as

$$\text{MSE}_k^{(j)} = \sqrt{\frac{1}{N'} \sum_{i=1}^{N'} \|\mathbf{z}_i^{(j)} - \mathbf{L}^{(j)}\|^2} \quad (11)$$

where $\|\cdot\|$ is the euclidean norm, N' is the number of ToF distance images used, \mathbf{Z} is a vector of ToF distance values that can be substituted by the angle corrected vector $\bar{\mathbf{z}}''$ of each distance image, or by the discrepancy corrected vector $\hat{\mathbf{z}}$ of each distance image, each one with dimension $n \times 1$, and with $j = 1, \dots, P$. The set of $\text{MSE}_k^{(j)}$ values for $k = 1, \dots, n$ gives an indication of the planar distribution of the distance error for a given

node j . Then, for a given node j , it is possible to average the mean square errors to obtain an indication of the error depending on the node position

$$\overline{\text{MSE}}^{(j)} = \frac{1}{n} \sum_{k=1}^n \text{MSE}_k^{(j)} \quad (12)$$

4. If the position of each pixel is taken into account, then:

- (a) *Discrepancy curves calculation stage.* Using the N angle corrected ToF distance images represented by $\bar{\mathbf{z}}''$, a discrepancy curve is calculated for each pixel at each distance node. At this stage, using N images at each node j , the mean value of each pixel k , where $k = 1, \dots, n$, is calculated as follows:

$$\bar{V}_{kj} = \frac{1}{N} \sum_{i=1}^N \bar{z}_{[i]k}''^{(j)} \quad (13)$$

where the resulting $\bar{\mathbf{V}}$, whose elements are the values \bar{V}_{kj} , is a matrix with dimensions $n \times P$. Define a new matrix \mathbf{L}'' of dimension $n \times P$ which is obtained by replicating n times the laser distances vector \mathbf{L}^T as follows:

$$\mathbf{L}'' = \begin{bmatrix} L_{[1]}^{(1)} & \dots & L_{[1]}^{(P)} \\ \dots & \dots & \dots \\ L_{[n]}^{(1)} & \dots & L_{[n]}^{(P)} \end{bmatrix} \quad (14)$$

Then, the discrepancy distance vector δ_v for all the j nodes is calculated for each pixel $k = 1, \dots, n$ as the difference between the mean distance from the ToF camera after angle correction, $\bar{\mathbf{V}}$, and the ground truth distance vector \mathbf{L}'' obtained using a laser distance meter:

$$\delta_v = \bar{\mathbf{V}} - \mathbf{L}'' \quad (15)$$

with δ_v of dimension $n \times P$.

In order to obtain n correction values to be applied to any new ToF distances images, a cubic spline is calculated to fit this discrepancy information along the distance range for each pixel. The cubic spline is modelled at each pixel k using Equation 8 and the data points $(\bar{\mathbf{V}}, \delta_v)$.

- (b) *Correction using a discrepancy curve at each pixel.* In order to reduce the errors in the ToF distances images, the set of ToF distances images \mathcal{Z}_v is used to validate each discrepancy curve at each pixel. To this end, each pixel k of the validation vector after angle corrections $\bar{\mathbf{z}}_v''$ (dimension $n \times 1$) is

evaluated on its discrepancy curve to obtain the vector of correction values \mathbf{C}_v (dimension $n \times 1$). Then, the corrected distance vector $\hat{\mathbf{v}}$ (dimension $n \times 1$) is obtained using the expression

$$\hat{\mathbf{v}} = \bar{\mathbf{z}}_v'' - \mathbf{C}_v \quad (16)$$

Define $\hat{\mathcal{V}}^{(j)} = \{\hat{v}_{[1]}^{(j)}, \hat{v}_{[2]}^{(j)}, \dots, \hat{v}_{[M]}^{(j)}\}$ as the set of distances after discrepancy correction, where the mean value at each pixel k for each j node is calculated as follows:

$$\hat{V}_{kj} = \frac{1}{M} \sum_{i=1}^M \hat{v}_{[i]k}^{(j)} \quad (17)$$

with $k = 1, \dots, n, j = 1, \dots, P$, and where the resulting $\hat{\mathbf{V}}$ is a matrix with elements \hat{V}_{kj} and dimensions $n \times P$ of mean ToF distances values at each pixel for each node.

The mean squared error is obtained by means of Equation 11, where \mathbf{Z} is replaced by the corrected values $\hat{\mathbf{v}}$.

A comparison of the MSE values for discrepancy corrected and non-corrected measurements gives a measure of improvement in accuracy due to the discrepancy correction. If no such improvement is detected, then it is recommended to revise the experimental conditions as this may indicate the existence of problems with the experiment.

3.1.2 Correcting the values of saturated pixels

Information from range cameras can be affected by pixel saturation, which is caused by an excessive reflectance of light over objects. Though its effect can be reduced by an automatic updated of the integration time parameter of the ToF camera [31], in some circumstances like the presence of metal or reflecting paints, this tool is not enough.

The saturation of range camera information affects the amplitude and distance values returned by the range camera. These values are very different from the remaining pixel values of the scene. The proposed strategy to detect saturated pixels is based on this fact, and an analysis of amplitude signal is made. The method has two stages.

1. *Looking for saturated pixels.* According to [45], pixel saturation occurs when the amplitude values are greater than a given threshold value ζ , which depends on the camera being employed. Hence, the amplitude image is searched for values greater or equal than this value in order to generate a saturation binary mask \mathbf{M} with ones at the positions of the saturated pixels and zeros elsewhere. To be able to perform the correction on pixels located at the edges

of the image, the amplitude and 3D information matrices are augmented by replicating rows and columns located at the edges of the matrix. Define p as the number of rows and columns of \mathbf{A} to be replicated. Define the p upper rows of \mathbf{A} as $B_{i,j} = A_{i,j}$, such that \mathbf{B} is of dimension $p \times n_y$, where $i = 1, \dots, p$ and $j = 1, \dots, n_y$, and the p lower rows as $B'_{i,j} = A_{n_x-i+1,j}$, such that \mathbf{B}' is of dimension $p \times n_y$, where $i = 1, \dots, p$, and $j = 1, \dots, n_y$. Define the intermediate matrix $\hat{\mathbf{A}}$ as follows:

$$\hat{\mathbf{A}} = \begin{bmatrix} \mathbf{B} \\ \mathbf{A} \\ \mathbf{B}' \end{bmatrix} \quad (18)$$

where $\hat{\mathbf{A}}$ is matrix of dimension $2p + n_x \times n_y$. Then, define the left p columns of \mathbf{A} as $B''_{i,j} = A_{i,j}$, such that \mathbf{B}'' is of dimension $2p + n_x \times p$, where $i = 1, \dots, 2p + n_x$ and $j = 1, \dots, p$ and the p right columns as $B'''_{i,j} = A_{i,n_y-j+1}$ such that \mathbf{B}''' is of dimension $2p + n_x \times p$, where $i = 1, \dots, 2p + n_x$ and $j = 1, \dots, p$. Then, the augmented amplitude matrix $\tilde{\mathbf{A}}$ of dimensions $2p + n_x \times 2p + n_y$ is given by:

$$\tilde{\mathbf{A}} = [\mathbf{B}'', \hat{\mathbf{A}}, \mathbf{B}'''] \quad (19)$$

To represent saturated pixels in $\tilde{\mathbf{A}}$, the binary mask $\tilde{\mathbf{M}}$ matrix of dimensions $2p + n_x \times 2p + n_y$, is defined by

$$\tilde{M}_{i,j} = \begin{cases} 1 & \text{if } \tilde{A}_{i,j} \geq \zeta \\ 0 & \text{otherwise} \end{cases} \quad (20)$$

where $i = 1, \dots, n_x + 2p, j = 1, \dots, n_y + 2p$. The set \mathcal{Q} of index pairs indicating the positions of saturated pixels is defined as follows

$$\mathcal{Q} = \{(r, c) \in \mathcal{I} \times \mathcal{J} \mid \tilde{M}_{r,c} = 1\} \quad (21)$$

where $\mathcal{I} = [1, \dots, n_x + 2p], \mathcal{J} = [1, \dots, n_y + 2p]$.

2. *Correction of saturated pixels.* In order to replace an incorrect value with the average of its neighbours, the saturation binary mask is used to find the coordinates of saturated values in the amplitude and 3D matrices and to calculate the mean value of surrounding pixels. Saturated values are not taken into account in this calculation. Define a window-mask $\tilde{M}_{i,j} = \tilde{M}_{r-p+i-1, c-p+j-1}$, with $i = 1, \dots, 2p + 1$ and $j = 1, \dots, 2p + 1$, of dimensions $2p + 1 \times 2p + 1$, whose center is each saturated pixel with position $(r, c) \in \mathcal{Q}$. In order to calculate a new pixel value to replace a saturated pixel value, define a window of amplitude values $\hat{A}_{i,j} = \tilde{A}_{r-p+i-1, c-p+j-1}$, with $i = 1, \dots, 2p + 1$ and $j = 1, \dots, 2p + 1$, of dimensions $2p + 1 \times 2p + 1$, whose center corresponds to each saturated pixel

with position $(p + 1, p + 1)$. The new value $\tilde{A}_{r,c}$ for each saturated pixel $(r, c) \in \mathcal{Q}$ is calculated as

$$\tilde{A}_{r,c} = \frac{1}{(2p + 1)(2p + 1) - 1} \left(\sum_{k=1}^{2p+1} \sum_{l=1}^{2p+1} \hat{A}_{k,l} - \hat{A}_{p+1,p+1} \right) \quad (22)$$

With the aim of selecting and replacing values in the amplitude/3D information matrices, Figure 2 shows an example of the movement of a search window obtained from the binary saturation mask. Define (\mathbf{X}, \mathbf{Y}) as the initial ToF data, $\hat{\mathbf{z}}$ as the distance TOF data after discrepancy correction, and using the index set \mathcal{Q} of amplitude saturated values, a similar procedure to correct the corresponding values of these matrices is applied, obtaining matrices $(\tilde{\mathbf{X}}, \tilde{\mathbf{Y}}, \tilde{\mathbf{Z}})$, as these values are affected by the amplitude saturation. Once saturated pixels are corrected, all matrices are resized to their initial dimensions by removing the rows and columns previously added, which results in matrices $\mathbf{X}', \mathbf{Y}', \mathbf{Z}'$, and \mathbf{A}' .

3.1.3 Jump edge reduction

Another error that may affect the 3D data from a range camera is known as *jump edge*. This error produces spurious pixels which are 3D inaccurate measures of the real scene. In order to reduce this effect, the use of a median filter followed by a jump edge filter based on a local neighbourhood is proposed in [46]. Other solutions which implement non-local means filter or edge-directed re-sampling techniques are enumerated in paper [31]. In the present work, the use of 2D techniques applied to 3D points is proposed to prevent border inaccuracy in fused information. Traditionally, the technique of morphological gradient is used in grey scale images to emphasize transitions of grey levels [47,48]. In this work, only distance values from 3D data are used, generating a distance image. With the objective of finding pixels suffering from this effect, the morphological gradient is calculated, using the following expression [48]:

$$\mathbf{g} = (\mathbf{f} \oplus \mathbf{S}) - (\mathbf{f} \otimes \mathbf{S}) \quad (23)$$

where \mathbf{g} is of dimension $n_x \times n_y$, \mathbf{f} is a ToF distance matrix of same dimension as \mathbf{g} , \mathbf{S} is a 3×3 generalised dilation or erosion mask, and \oplus and \otimes are dilation and erosion operations, respectively.

A threshold value to discriminate non-desirable pixels from the remaining ones is then searched. With this aim, the distance image \mathbf{g} is transformed into a new distance image \mathbf{G} with values ranging from 0 to 255, by means of the following transformation:

$$\mathbf{G} = 255 (\mathbf{g} / \max(\mathbf{g})) \quad (24)$$

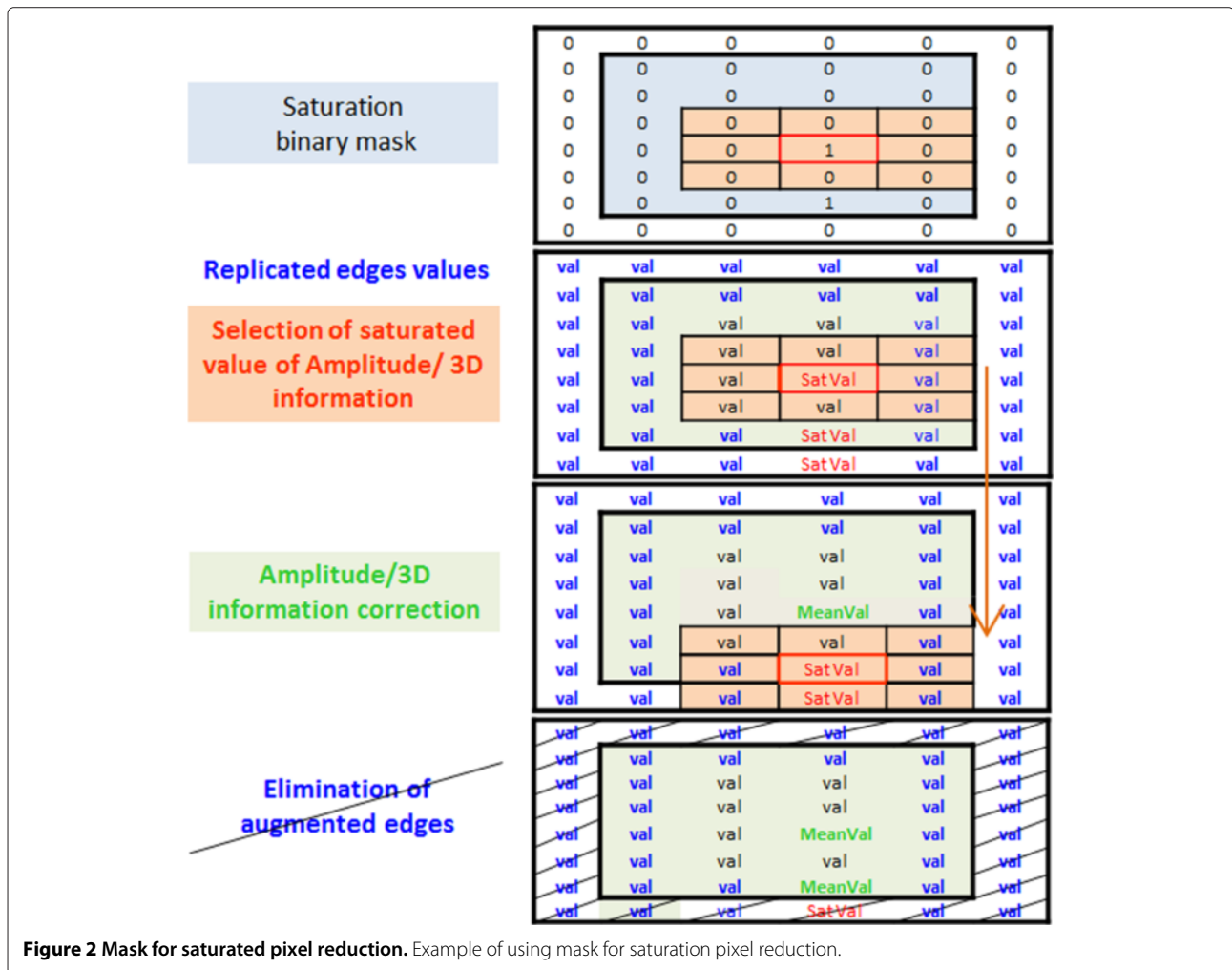


Figure 2 Mask for saturated pixel reduction. Example of using mask for saturation pixel reduction.

After that, the histogram of \mathbf{G} is calculated and then smoothed by means of a Butterworth filter. Finally, a threshold value η is defined by searching along the smoothed histogram for the first minimum to the right of the first maximum. A new distance matrix \mathbf{f} is generated by forcing to zero spurious pixels which are found and keeping the same distance values for the remaining pixels:

$$f'_{i,j} = \begin{cases} 0 & \text{if } G_{i,j} \geq \eta \\ f_{i,j} & \text{otherwise} \end{cases} \quad (25)$$

When performing the fusion of ToF and colour information, jump edge reduction is carried out after scaling up the ToF information, as discussed below.

3.2 Colour and 3D information fusion

Information fusion from a standard CCD camera and a ToF camera allows the simultaneous use of 3D and colour information. This can be achieved by means of the reprojection of 3D ToF points into a colour image. In

an active security system, moving objects, such as robots and humans, have to be detected to prevent possible collisions between them. To obtain information about these objects and develop the algorithms that make it possible to avoid collisions, the foreground detection is carried out in such way that the fused information is obtained only through those pixels classified previously as foreground pixels. The foreground object detection in a scene is carried out using 2D techniques over 3D ToF points, and subsequently, colour and 3D information from foreground objects is fused.

3.2.1 3D information analysis for detecting foreground objects

Background subtraction methods for detecting moving objects have been proposed, analysed, and employed to locate object motion in a 2D image sequence [49-51]. In this work, for the purpose of motion detection in 3D point cloud, and considering that ToF camera is static, and illumination changes do not affect the acquired 3D points, the

background subtraction technique has been considered suitable to be adapted and applied to three-dimensional information. Therefore, after performing distance and saturated pixel correction, a background subtraction method based on the reference image model is adapted to be used in a 3D point cloud. The goal is to discriminate the static part of the 3D scene from the moving objects, so an offline background reference image $\mathbf{B}_{\mathcal{T}}$ is calculated as the average image during a time period $\mathcal{T} = 1, \dots, t$. Define a set of t ToF distance images after discrepancy and pixel saturation correction captured in a time period \mathcal{T} , such that $\mathcal{Z}' = \{\mathbf{Z}'_1, \mathbf{Z}'_2, \dots, \mathbf{Z}'_t\}$, then, the background reference image is calculated as

$$\mathbf{B}_{\mathcal{T}} = \frac{1}{tn} \sum_{i=1}^t \sum_{k=1}^n \mathbf{Z}'_{[i],k} \quad (26)$$

where n is the number of pixels in each ToF distance image. With the aim of detecting pixels that show motion, the difference image \mathbf{Z}'_d between the reference and a current image \mathbf{Z}'_c is calculated as:

$$\mathbf{Z}'_d = |\mathbf{B}_{\mathcal{T}} - \mathbf{Z}'_c| \quad (27)$$

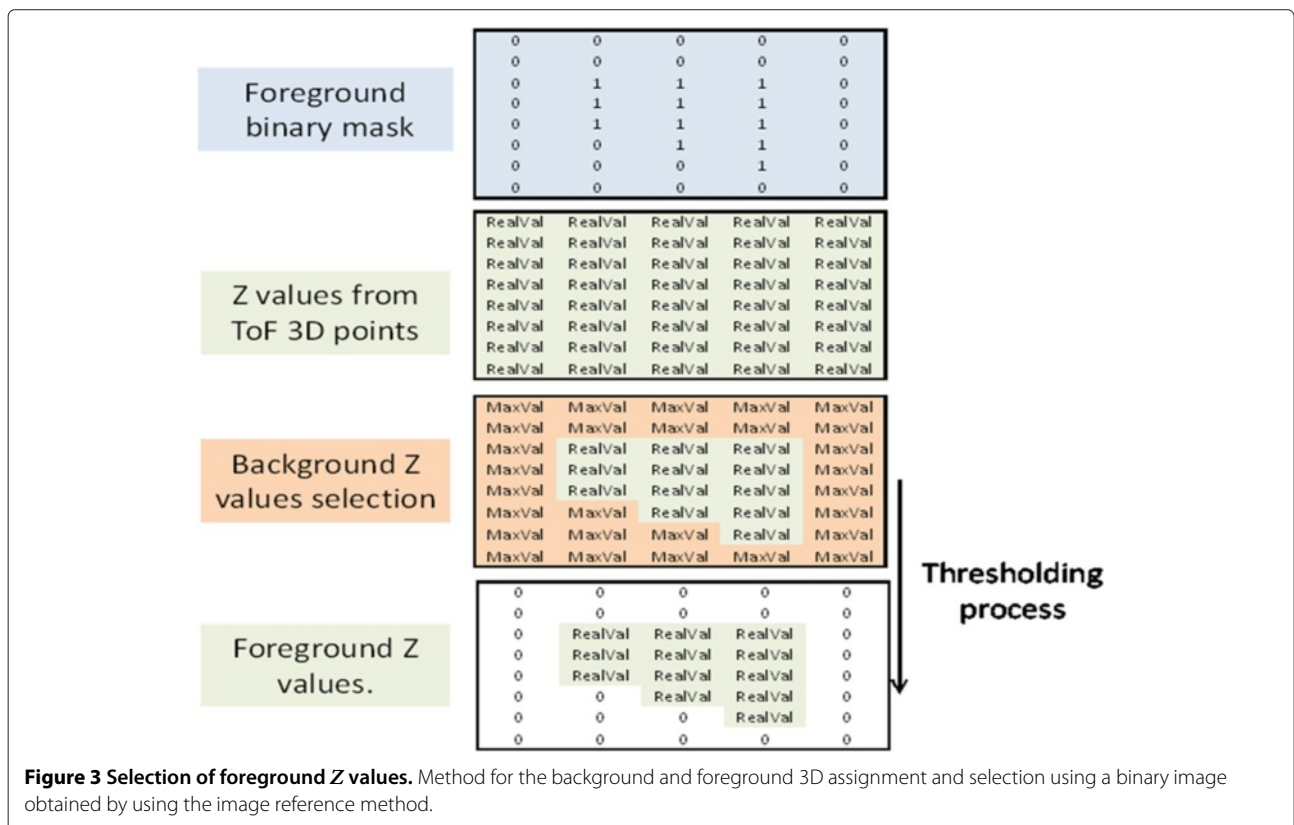
where $|\cdot|$ indicates an element-wise absolute value operation.

Foreground detection is performed in those pixels whose distance value, \mathbf{Z}'_d , exceeds a threshold value, T_h ,

which results in a binary image \mathbf{Z}'_b . In order to automatically determine T_h , the distance matrix \mathbf{Z}'_d is processed as if it was 2D information by means of Equation 24, where \mathbf{g} is replaced by \mathbf{Z}'_d , resulting in a grey scale image \mathbf{G}' . Then, the calculation of the smoothed histogram of \mathbf{G}' and the search for threshold value are carried out in a similar way as presented in the 'Jump edge reduction' section. The binarisation process to detect pixels that show motion is given by

$$\mathbf{Z}'_b = \begin{cases} 1 & \text{if } \mathbf{G}' \geq T_h \\ 0 & \text{otherwise} \end{cases} \quad (28)$$

In the resulting binary image, isolated pixels are removed using morphological operations (dilation, hole filling, and erosion). This enhanced binary image is used as a mask over the 3D points of \mathbf{Z}' to set the maximum value to the coordinate of 3D points whose coordinates in the binary image are considered as background (0 value) and to leave as real Z values those 3D points whose coordinates in the binary image are considered as foreground (1 value), then a new ToF distance matrix \mathbf{Z}'' is obtained. Figure 3 illustrates this method for the background and foreground 3D value assignment and selection.



3.2.2 Reprojection of 3D ToF information into a colour image

With the aim of giving additional colour information to the 3D foreground points previously detected, the reprojection of these points into a colour image was carried out. Using colour and amplitude images, both cameras are calibrated with respect to the world coordinate frame. Since both cameras can be represented by the pinhole camera model [42,52], a tool such as the *Camera Calibration Toolbox for Matlab* [53] can be used to extract internal and external parameters for both cameras. External parameters are used to transform 3D ToF information given in the camera coordinate system into the world coordinate system. On the other hand, internal and external parameters are used to reproject 3D information into colour images. Hence, based on calibration camera theory [48,54,55] and after the range camera error reduction, the reprojection process is applied over the corrected and transformed 3D points following the transformations described below.

The transformation of ToF information after discrepancy and saturation corrections and foreground detection from world frame coordinates $\mathbf{P}_w = [X', Y', Z']^T$ to camera frame coordinates $\mathbf{P}_c = [X_c, Y_c, Z_c]^T$ is given by

$$\mathbf{P}_c = \mathbf{R}\mathbf{P}_w + \mathbf{T} \quad (29)$$

where extrinsic parameters are expressed by the 3×3 rotation matrix \mathbf{R} and by the 1×3 translation vector \mathbf{T} .

Frequently, standard CCD colour cameras have a higher resolution than range cameras, so the reprojection of 3D points does not have a one-to-one equivalence. Hence, the ToF information is scaled up by bilinear interpolation. In addition to this, as only information of foreground 3D points will be extracted, the automatic thresholding process is applied to the 3D points \mathbf{P}_c in order to remove those points classified as background, which results in a new 3D point cloud $\mathbf{P}'_c = [X'_c, Y'_c, Z'_c]^T$.

Image coordinates are affected by tangential and radial distortions; therefore, the models of this systematic distortions are added to the pinhole model following the method proposed in [55]. The transformation between a three-dimensional coordinate frame and the image coordinate frame without distortion (x_u, y_u) is given by

$$\begin{aligned} x_u &= fX'_c/Z'_c \\ y_u &= fY'_c/Z'_c \end{aligned} \quad (30)$$

where the intrinsic parameter f is the focal length in millimetres.

The relation between image coordinates with (x_d, y_d) and without distortion (x_u, y_u) , considering the radial $D^{(r)}$, and tangential $D^{(t)}$ distortions are defined by pinhole model as

$$\begin{aligned} x_u + D_x^{(r)} + D_x^{(t)} &= x_d \\ y_u + D_y^{(r)} + D_y^{(t)} &= y_d \end{aligned} \quad (31)$$

The transformation between distorted image coordinates to pixel coordinates is given by

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} k_u & s & u_0 \\ 0 & k_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_d \\ y_d \\ 1 \end{bmatrix} \quad (32)$$

where the intrinsic parameters k_u and k_v are the number of pixels per millimetre (horizontally and vertically, respectively), s is the skew factor whose value is usually zero in most cameras, and (u_0, v_0) are the coordinates of the centre of projection.

After obtaining the pixel coordinates (u, v) of the 3D foreground points, these values are adjusted into pixel values by rounding them to the nearest integer to the values obtained. Furthermore, as the captured area by both cameras (ToF and colour camera) is not exactly the same, pixels in non-common areas are eliminated. A diagram which illustrates the proposed method is shown in Figure 4.

4 Experiments

4.1 Experimental setting

In this article, a method for the fusion of colour and 3D information that is suitable for active security systems in industrial robotic environments is presented. To verify the proposed methods, a colour camera, AXIS 205, and a range camera, SR4000, have been located over the workspace of the robot arm FANUC ARC MATE 100iBe. The AXIS 205 Network Camera used has a resolution of 640×480 pixels and a pixel size of 5.08×3.81 mm. The SR4000 range camera has a resolution of 176×140 pixels and a pixel size of 40×40 μm . This camera has a modulation frequency of 29/20/31 Mhz and a detection range from 0.1 to 5 m.

4.2 Camera calibration

This initial stage is intended to obtain extrinsic and intrinsic parameters from the standard CCD camera and ToF camera by means of a calibration process using a common reference frame.

4.3 Reduction of distance error

To correct for any misalignment between the range camera and the experimental panel employed, the angular deviations in x and y coordinates have been estimated and their effects have been corrected. Figure 5 shows the effect of the angle and displacement corrections in a 3D point cloud. Discrepancy curves which do not take into account pixel position have been calculated before and after angular correction. These curves are shown in Figure 6. It can be seen that the distance error is a function of the measured distance, and discrepancies values show a small improvement after the angle correction, as it was expected.

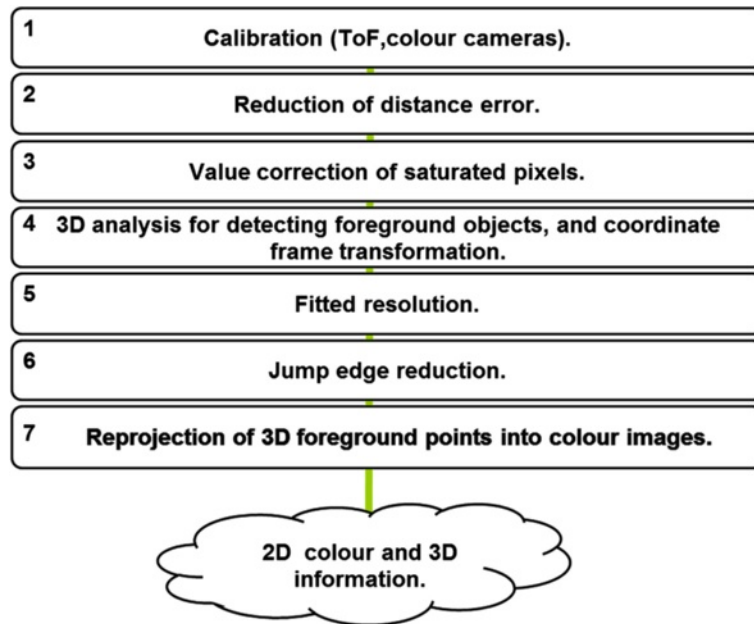


Figure 4 Proposed method for fusion 3D ToF and colour information. Stages proposed to achieve the 3D ToF and colour information fusion.

To take into account the effect of pixel position in the distance error, a discrepancy curve has been generated for each pixel. These curves are tested by using the set Z_y as input, resulting in a correction value to be applied at each pixel. Figure 6 shows in green colour the discrepancy

values after the correction with a cubic spline for each pixel.

The results indicate that the improvement achieved using a discrepancy curve at each pixel is almost imperceptible, which can be explained by the selected area being

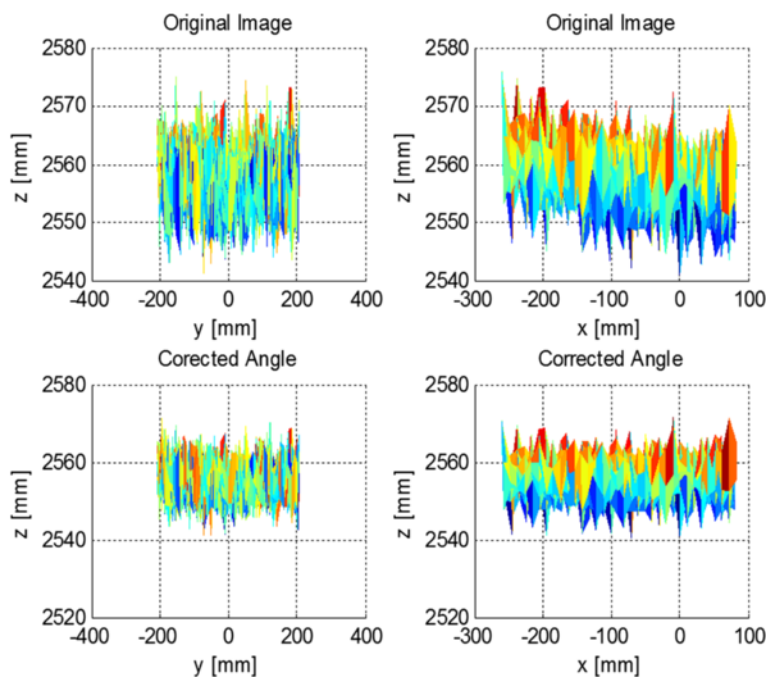
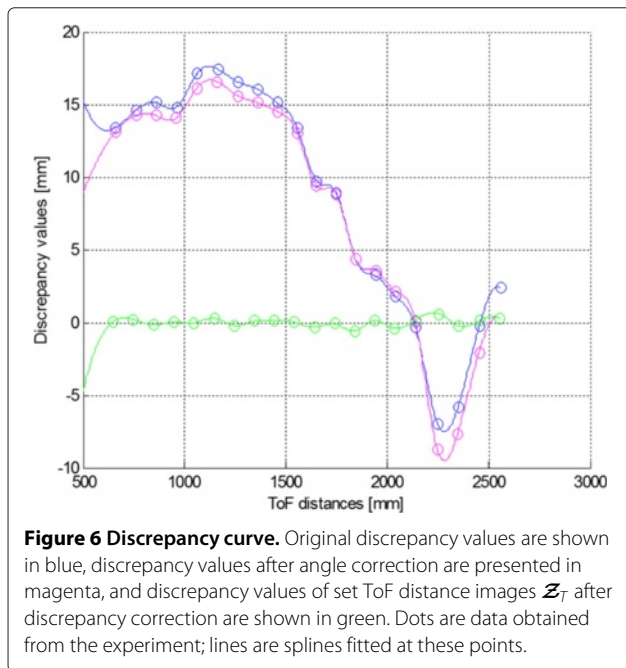


Figure 5 3D point cloud before and after angle correction. Initial 3D point cloud before and after angle correction. 3D point cloud are presented in (x, z) coordinates and in (y, z) coordinates.



too small and centred in the middle of the image, so the influence of the pixel position is very low. To check the influence of pixel position on the distance error, a larger area has been selected from a reduced set of images taken from Z_V . Figure 7a shows the MSE before these corrections, while the MSE after discrepancy correction using the same discrepancy curve at each pixel is shown in Figure 7b. It can be seen that there is a reduction in the MSE over the selected area. However, the distance error is not just a function of the distance value but also it depends on the location of the pixel, as can be observed. The results obtained using a different discrepancy curve at each pixel (Figure 7c) suggest that this kind of correction leads to better results. Hence, a discrepancy correction which takes into account pixel position and distance value is suggested for future work.

Since as a result of using the larger area, there is incomplete information along the whole of the distance range, a discrepancy curve which does not take into account pixel position is used in the experiment. Then, the discrepancy curve calculated after angle correction is tested by using a ToF distance image from a real scene where data from a human and a robot arm are captured and used as input, resulting in a correction value to be applied at each pixel. Figure 8a shows in red colour the discrepancy values selected to use in discrepancy correction together with the generated cubic spline showed in cyan colour. In order to verify the effect of applying the distance correction, the initial 3D point cloud and the results after applying the distance correction are shown in Figure 8b.

4.4 Value correction of saturated pixels

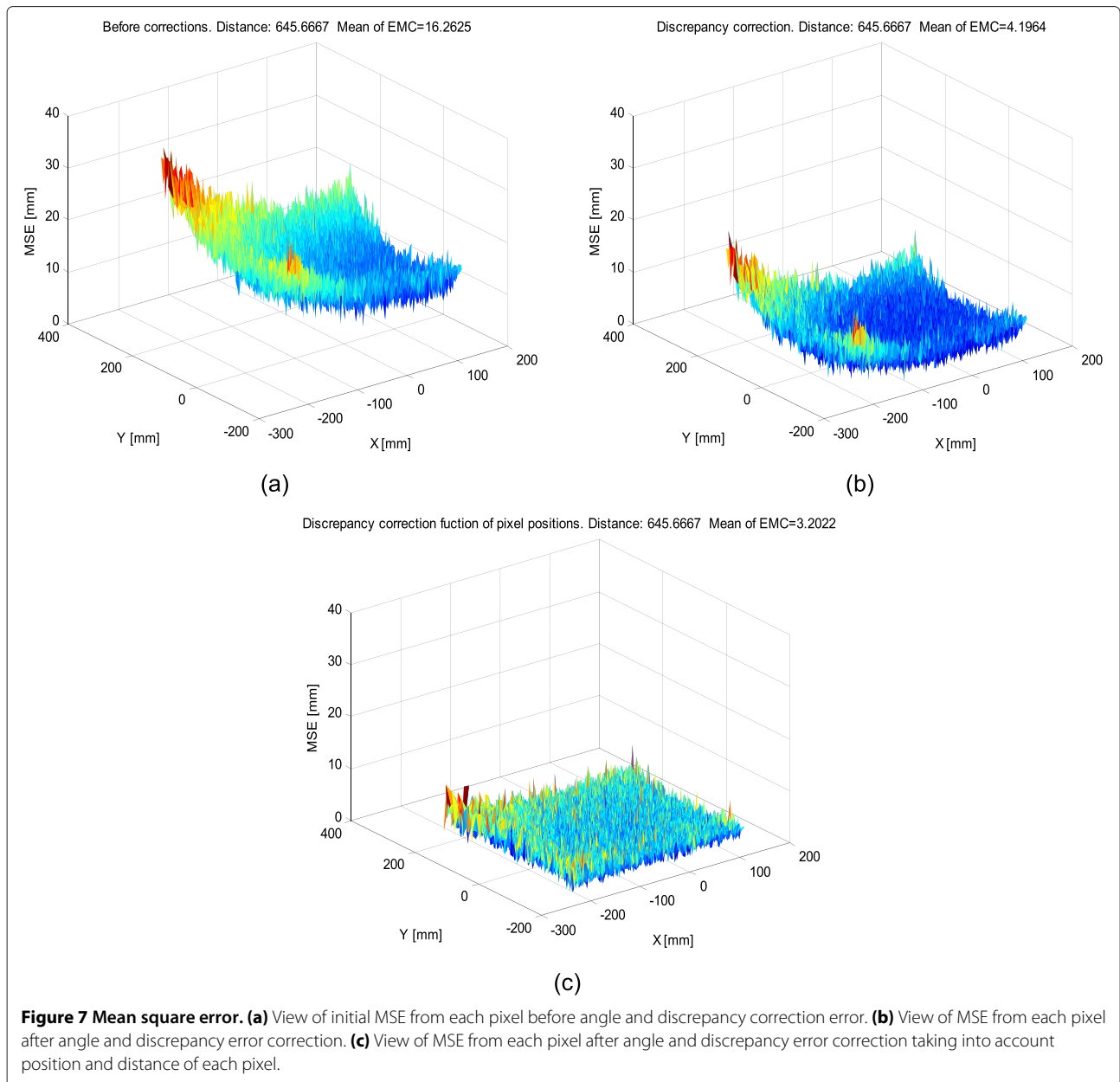
A real scene in which a human and a robot arm appear is used to illustrate the proposed methods of error corrections and the fusion of 3D and colour information. The value correction of saturated pixels in ToF information captured from this real scene has been carried out. This example of the effect of saturation is illustrated in Figure 9a which shows an amplitude image in which several saturated pixels are located on an area of the robot arm. These high values do not allow the correct visualization of the scene. Figure 9b shows the effect of saturated information over 3D data where saturation produces pixels with zero coordinate values. According to [45], pixel saturation occurs when the amplitude values are greater than 20,000, so this value has been used as threshold in Equation 20. After applying the proposed method to saturated pixels using this threshold value, Figure 10a shows the improvement achieved with this correction, allowing the view of the total scene. Figure 10b shows 3D points in which pixels with zero coordinate values have been corrected.

4.5 3D analysis for detecting foreground objects and coordinate frame transformation

To illustrate the detection of foreground objects using 3D ToF information, the background subtraction method based on the reference image model has been used. Figure 11a shows 3D information with a generated reference matrix with Z values. Figure 11b shows three-dimensional information resulting from subtracting the reference distance matrix from the real scene distance matrix, in which positive values indicate possible motion points. In order to take into account only foreground 3D points, an automatic thresholding process and the proposed method for the background and foreground 3D values assignment and selection have been applied using Equation 28. After that, the modified 3D points are shown in red in Figure 12, whereas the initial 3D points are shown in cyan colour. It can be observed that in the modified 3D points, all background points have equal Z values. However, as the points of interest are the foreground points, the background points are not taken into account; therefore, the final scene 3D representation is not affected by those equal Z values. After coordinate frame transformation using Equation 30, and another automatic thresholding process to remove points classified as background, the result achieved in this example is shown in Figure 13, where the foreground object detected is represented in the world coordinate system.

4.6 Resolution increase

As the standard CCD camera employed provides a colour image which has higher resolution (480×640) than the



3D ToF information (176×144) provided by the range camera, the reprojection of 3D points does not have a one-to-one equivalence. Then, ToF matrices dimensions have been scaled up using a bilinear interpolation and reprojected to the colour image using Equations 30 to 32.

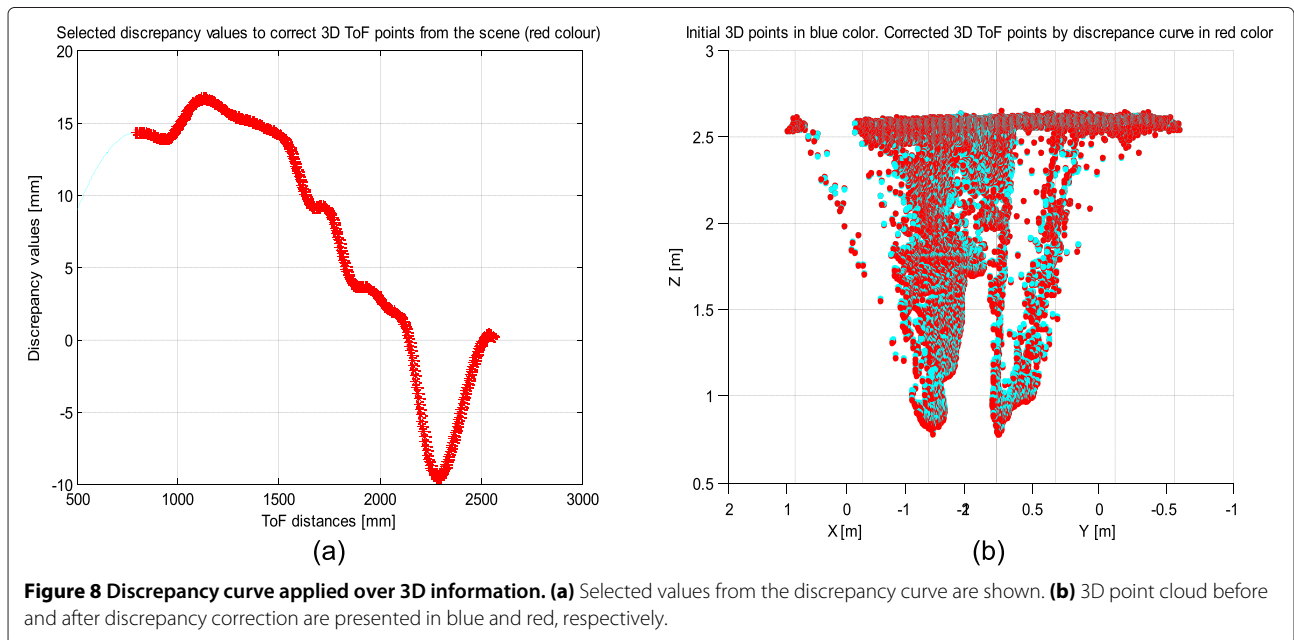
4.7 Jump edge reduction

With the aim to compare some usual edge filters and the morphological filter used in the detecting edge jump effect, ToF information from the scene, after interpolation of 3D points, has been processed. Figure 14a shows the results achieved using a Sobel filter in distance values

from 3D information, and Figure 14b shows the results obtained using the morphological filter in distance values, obtained by using Equation 23 and establishing a dilation and erosion mask S as follows

$$S = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 0 \end{bmatrix} \quad (33)$$

It can be observed that the edges found by applying the Sobel filter are not continuous and also are narrower than the edges found by morphological filter, so using this,



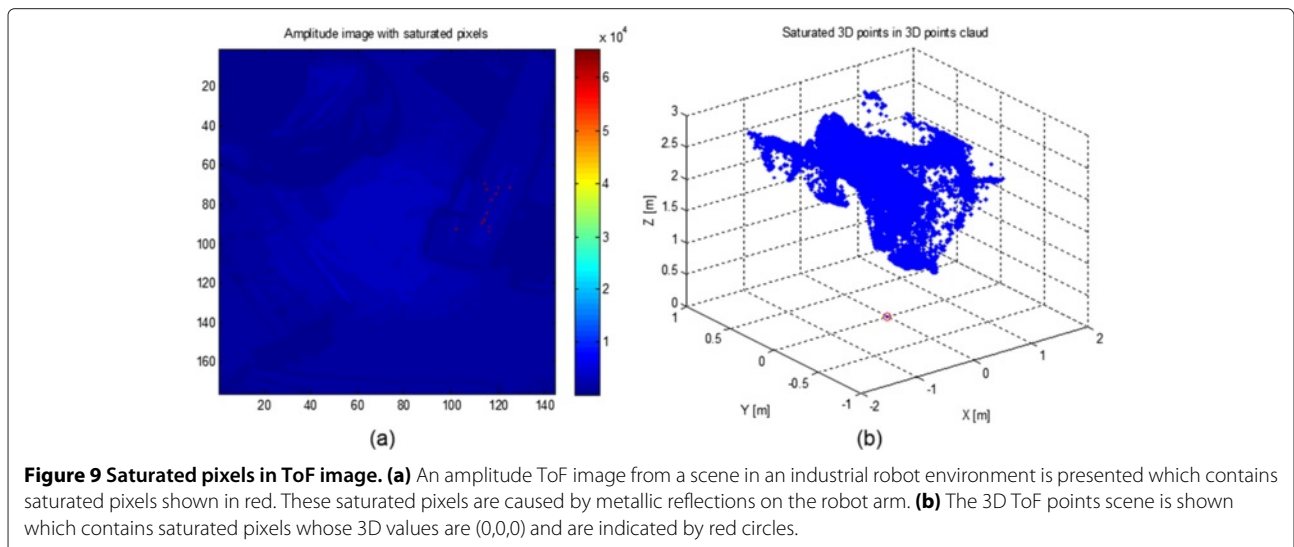
most of the spurious pixels can be detected and removed from the 3D ToF points.

In order to smooth the histogram of the gray scale distance image, a first-order lowpass Butterworth filter with normalized cutoff frequency value of 0.5 is used. As an example of the application of the proposed method for the jump edge reduction, Figure 15a shows the spurious pixels produced in the object contours by the jump edge effect and Figure 15b shows the 3D points after the reduction of spurious pixels by the proposed method. Although not all spurious pixels have been eliminated, the results show a significant improvement in the reduction

of this effect as most of them have been detected and eliminated.

4.8 Reprojection of 3D foreground points into colour images

In order to obtain a matrix that contains 3D and 2D information, using the calibration parameters of the cameras, the reprojection of 3D foreground points into a colour image has been carried out. Then, the reprojected points are adjusted into pixel values and those that are in the non-common area of both cameras are removed. A selection mask is generated by using the resulting pixels,



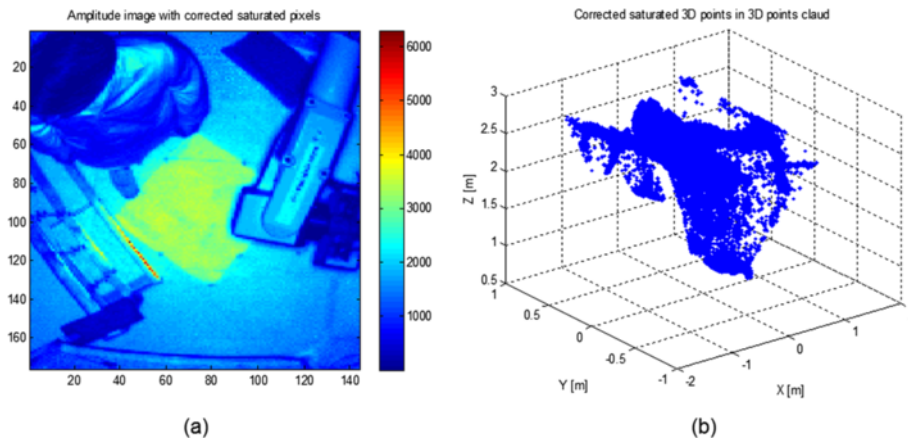


Figure 10 Saturated pixels in 3D ToF points. (a) The same amplitude ToF image is presented after pixel saturation reduction. **(b)** 3D ToF points after pixel saturation reduction.

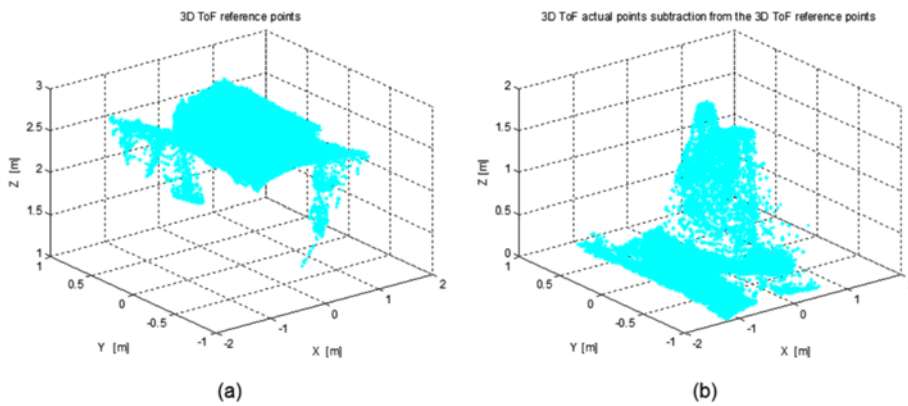


Figure 11 Foreground detection of 3D points expressed in ToF camera frame. (a) Background reference matrix $\mathbf{B}'_{\mathcal{T}}$ to be used in foreground detection of 3D ToF points. **(b)** Value absolute of the difference image Z'_d between the reference, $\mathbf{B}'_{\mathcal{T}}$, and a current image Z'_c .

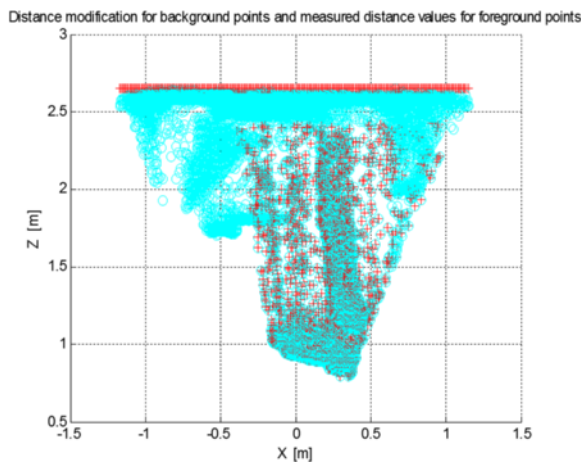


Figure 12 Foreground and background Z values from ToF points in ToF camera frame. Background distance modification of distance values. View of axis x, z .

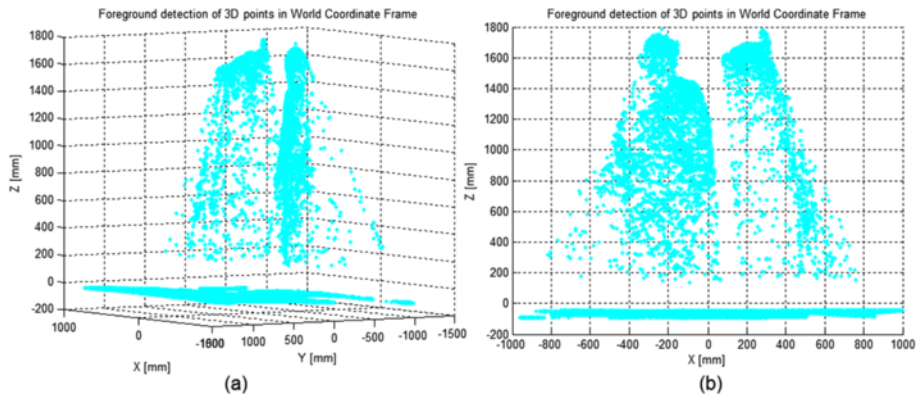


Figure 13 Foreground and background 3D points detection in world reference frame. **(a)** View of axis x, y, z . **(b)** View of axis x, z .

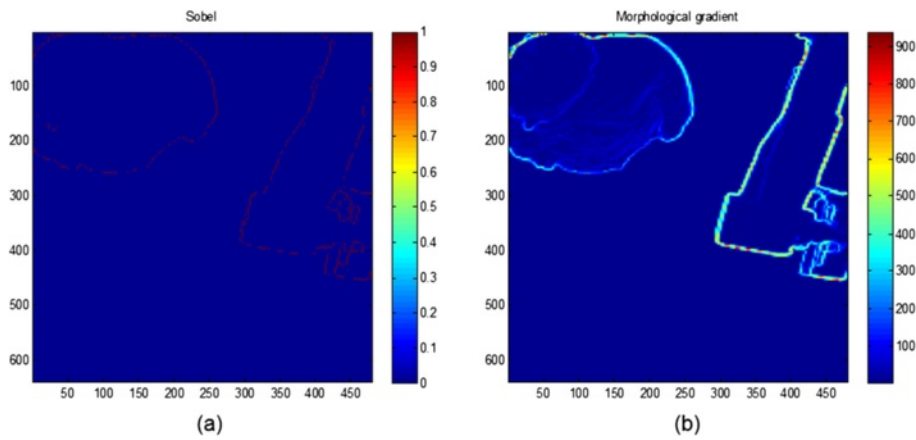


Figure 14 Strategy for jump edge error reduction. **(a)** View of Sobel filter used on the z information. The edges found are not suitable as they are not continuous and have only a thickness of one pixel. **(b)** View of morphological gradient used on z information. The edges found are continuous and have thickness of several pixels, which allows jump edge reduction.

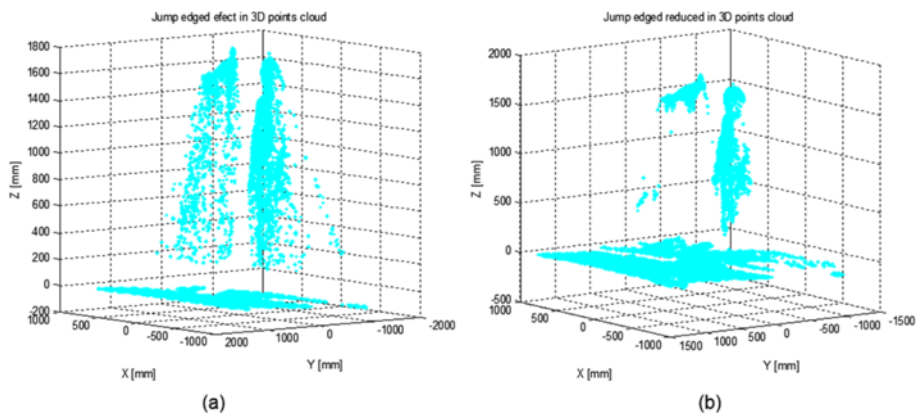


Figure 15 View of of jump edge error in 3D ToF points. **(a)** View of 3D ToF points from the scene presented in Figure 9 in which jump edge effect appears. **(b)** View of the same 3D scene in which jump edge has been reduced using the morphological gradient operation.

and this mask is used to select the coincident coordinates of the colour pixels. This method makes it possible to achieve a colour segmentation based on 3D information and to have 2D and 3D information in a single matrix. Figure 16 shows foreground segmentation in the colour image based on foreground detection of 3D points in the world coordinate system.

5 Discussion

The aim of this work is to achieve the fusion of colour and 3D ToF information in order to apply it in active security tasks for industrial robotic environments, so given the coordinates of a 3D point, this fusion allows knowing colour information and 3D position in a common world coordinate system to both cameras and the robot arm, at the same time.

After obtaining intrinsic and extrinsic parameters by a calibration process, the proposed method of distance error reduction improves the distance measurement values, and the achieved effect in the scene can be observed after the application of the information fusion method. The correction curves obtained are consistent with curves reported by other authors such as [41] and also consistent with the use of cubic splines in order to approach and correct the distance error [42]. This consistence occurs despite some differences in experimental setup, such as a reduced range of measurement, different ToF camera models, target material, and camera configuration parameters.

As a second stage, saturation error correction must be performed given that in industrial environments, certain materials such as metal or reflecting paints are often present and can produce saturated pixels in the range camera information. The results obtained show that this method works well as it allows the correct visualization of the amplitude image, and more importantly, it corrects

values of saturated pixels of 3D points. If these points have incorrect values, the reprojection stage would fail in these positions, as the 3D values would be reprojected as 0, and so its 2D information would be lost.

In order to detect foreground objects, the reference image technique applied to 3D data, after error corrections, has been used and presented as a simple and fast method which yields acceptable results. The 3D points of foreground objects are correctly identified and only a few false positives are detected which can be removed easily using 2D image morphological operations. Traditionally, this technique is used in colour and grey scale images, but illumination variations result in false foreground detection. The advantage of using ToF information is that it has a more stable behaviour in these illumination conditions. In addition, this technique has a short computational time, which is an important factor in order to be develop a suitable strategy for active security of robotic industrial environments. Then, using extrinsic parameters, the transformation of foreground 3D points from the camera to the world reference frame is carried out and scaled up by bilinear interpolation. The proposed method of jump edge reduction, applied to the resulting distance points, minimises false positives and false negatives around an object edge which arise in the pixel reprojection process as a consequence of the presence of spurious pixels that do not have correct 3D values. The achieved results can be considered acceptable since most spurious pixels are removed without changing the object shape, and therefore, a softer 3D point reprojection over objects edges in colour images is achieved.

Finally, the reprojection of the resulting 3D points to the colour image is performed. Nevertheless, as can be seen in Figure 16, this reprojection is not perfect, since in spite of having applied distance error reduction, the position of the pixels in the image has not been taken into account,

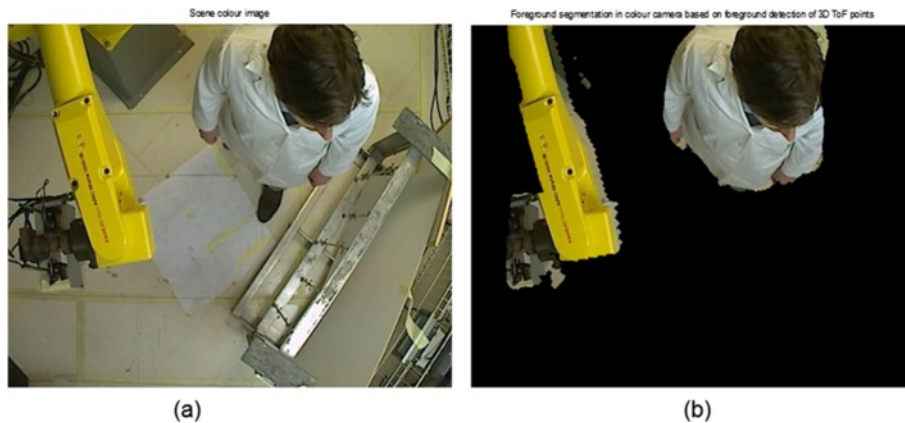


Figure 16 Foreground segmentation in colour images based on foreground detection of 3D points. **a)** Original image **b)** Foreground segmentation in the colour image.

and a single correction value is applied which is a function of the measurement distance but not of the pixel position in the image.

6 Conclusions

This paper aims to contribute to the research area of active security systems in industrial robotic environments using ToF cameras.

Despite the fact that active security in robotic industrial environments is a wellstudied topic, few previously published methods have dealt with this subject using the combination ToF cameras and colour cameras. The paper describes the development of methods for the fusion of colour and 3D ToF information as an initial step in the design of a system for collision prevention between human and manipulator robot sharing a workspace at the same time. Furthermore, this work provides a detailed mathematical description of the steps involved in the proposed method, so that any researcher can implement it.

The presented method has a different standpoint from the methods previously proposed in the literature, since a common coordinate system is defined for a robot arm, colour camera and ToF camera. The obtained calibration parameters are used to transform the 3D points from the ToF camera coordinate system into the defined common coordinate system, which are reprojected in 2D colour images. This procedure has the advantage that it gives a single matrix made of colour and three-dimensional information; therefore, 3D coordinates of objects inside the robot arm's workspace are known at the same time as their colour information. In addition to this, the proposed method for jump edge error detection, which is based on morphological gradient, allows the detection and reduction of jump edge error at points which are affected by this error. Also, in order to obtain a suitable fusion of information, a method for detection and reduction of saturated pixels, which is based on neighbour pixels information, has been proposed.

As future work, in order to improve the accuracy of fused information, a modification of the applied distance correction method is suggested. A preliminary study carried out with a small range of distances shows the influence of the pixel position in the distance measurements. Hence, a suggestion for future work is to modify the error correction so that it takes into account the position of the 3D point (measured distance and pixel location).

A possible application to prevent collisions between an industrial robot and a human would be to use colour information to characterise the detected foreground objects and to associate a security volume around each object.

Competing interests

The authors declare that they have no competing interests.

Acknowledgements

This work has been supported by the Ministry of Economy and Competitiveness of the Spanish Government (project DPI2012-36959).

Author details

¹Electronic Technology and Automatic Systems Department, School of Industrial and Telecommunications Engineering, University of Cantabria, Avda. de Los Castros s/n, Santander 39005, Spain. ²University of Reading, School of Systems Engineering, Whiteknights, Reading, Berkshire RG6 6AY, UK.

Received: 16 December 2013 Accepted: 23 May 2014

Published: 12 June 2014

References

1. UNE-EN 755, Robots manipuladores industriales. Seguridad. Ed. by AENOR. (Asociacion Española de Normalizacion y Certificacion, Madrid, 1996)
2. ISO 10218-1, Robots for industrial environments. Safety requirements. Part 1: Robots. Ed. by ISO. (International Organization for Standardization, Switzerland, 2006)
3. RIA TR R15.206-2008, *Guidelines for implementing ANS/RIA/ISO 10218-1-2007. For industrial robots and robot systems. Safety requirements*. Ed. by RIA. (Robotic Industries Association, USA, 2008)
4. JR Llata, EG Sarabia, J Arce, JP Oria, Fuzzy controller for obstacle avoidance in robotic manipulators using ultrasonic sensors, in *Advanced Motion Control, 1998. AMC '98-Coimbra, 1998 5th International Workshop On* (IEEE, 1998), pp. 647-652
5. JT Feddema, JL Novak, Whole arm obstacle avoidance for teleoperated robots, in *Robotics and Automation, 1994. Proceedings., 1994 IEEE International Conference On*, (1994), pp. 3303-33094
6. JL Novak, JT Feddema, A capacitance-based proximity sensor for whole arm obstacle avoidance, in *Robotics and Automation, 1992. Proceedings., 1992 IEEE International Conference On*, (1992), pp. 1307-13142
7. Y Yu, K Gupta, Sensor-based roadmaps for motion planning for articulated robots in unknown environments: some experiments with an eye-in-hand system, in *Intelligent Robots and Systems, 1999. IROS '99. Proceedings 1999, IEEE/RSJ International Conference On*, vol. 3 (Kyonjju, Korea, 1999), pp. 1707-17143
8. S Puls, J Graf, H Wörn, *Cognitive Robotics in Industrial Environments*. Human Machine Interaction - Getting Closer (InTech, 2012), pp. 213-234. <http://www.intechopen.com/books/human-machine-interaction-getting-closer/cognitive-robotics-in-industrial-environments>
9. B Martinez-Salvador, AP del Pobil, M Perez-Francisco, A hierarchy of detail for fast collision detection, in *Intelligent Robots and Systems, 2000. (IROS 2000). Proceedings 2000 IEEE/RSJ International Conference On*, vol. 1 (Takamatsu, Japan, 2000), pp. 745-7501
10. L Balan, GM Bone, Real-time 3d collision avoidance method for safe human and robot coexistence, in *Intelligent Robots and Systems, 2006 IEEE/RSJ International Conference On* (Beijing, China, 2006), pp. 276-282
11. JA Corrales, F Torres, FA Candelas, IEEE ICRA 2010 Workshop on Multimodal Human-Robot Interfaces (Anchorage, Alaska, 2010)
12. JA Corrales, FA Candelas, F Torres, Safe human-robot interaction based on dynamic sphere-swept line bounding volumes. *Robot. Comput. Integr. Manuf.* **27**(1), 177-185 (2011)
13. Y Nakabo, H Saito, T Ogure, SH Jeong, Y Yamada, Development of a safety module for robots sharing workspace with humans, in *Intelligent Robots and Systems. IROS 2009. IEEE/RSJ International Conference On* (St. Louis, MO, USA, 2009), pp. 5345-5349
14. A-J BaerVELdt, Cooperation between man and robot: interface and safety, in *Robot and Human Communication, 1992. Proceedings. IEEE International Workshop On* (Tokyo, 1992), pp. 183-187
15. S Kuhn, T Gecks, D Henrich, Velocity control for safe robot guidance based on fused vision and force/torque data, in *2006 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems* (Heidelberg, Germany, 2006), pp. 485-492
16. DM Ebert, DD Henrich, Safe human-robot-cooperation: image-based collision detection for industrial robots, in *Intelligent Robots and Systems, 2002. IEEE/RSJ International Conference On*, vol. 2 (Lausanne, Switzerland, 2002), pp. 1826-18312
17. T Gecks, D Henrich, Simero: camera supervised workspace for service robots, in *2nd Workshop on Advances in Service Robotics, Fraunhofer IPA* (Germany, 2004)

18. T Gecks, D Henrich, Multi-camera collision detection allowing for object occlusions, in *37th International Symposium on Robotics (ISR 2006)/4th German Conference on Robotics (Robotik 2006) VDI/VDE-Gesellschaft Mess- und Automatisierungstechnik* (München, Germany, 2006)
19. D Henrich, S Kuhn, Modeling intuitive behavior for safe human/robot coexistence cooperation, in *Robotics and Automation, 2006. ICRA 2006. Proceedings 2006 IEEE International Conference On* (Orlando, FL, 2006), pp. 3929–3934
20. B Fevery, B Wyns, L Boullart, JR Llata, C Torre-Ferrero, Industrial robot manipulator guarding using artificial vision, in *Robot Vision*, ed. by Ales U (In-Tech, Vukovar, 2010), pp. 429–454
21. L Bascetta, G Ferretti, P Rocco, H Ardo, H Bruyninckx, E Demeester, ED Lello, Towards safe human-robot interaction in robotic cells: an approach based on visual tracking and intention estimation, in *Intelligent Robots and Systems (IROS), 2011 IEEE/RSJ International Conference on* (San Francisco, CA, 2011), pp. 2971–2978
22. D Kulic, EA Croft, Real-time safety for human robot interaction. *Robot. Autonom. Syst.* **54**(1), 1–12 (2006)
23. L Bascetta, G Magnani, P Rocco, R Migliorini, M Pelagatti, Anti-collision systems for robotic applications based on laser time-of-flight sensors, in *In Proceedings of the IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM)*. Montreal, Canada, 2010, pp. 278–284
24. F Flacco, A De Luca, Multiple depth/presence sensors: integration and optimal placement for human/robot coexistence, in *Robotics and Automation (ICRA), 2010 IEEE International Conference On* (Anchorage, Alaska, 2010), pp. 3916–3923
25. J Graf, P Czapiewski, H Woern, Evaluating risk estimation methods and path planning for safe human-robot cooperation, in *Proceedings of the joint conference of ISR 2010 (41st International Symposium on Robotics) and ROBOTIK 2010 (6th German Conference on Robotics)* (VDE Verlag, Munich, 2010), pp. 1–7
26. M Fischer, D Henrich, 3d collision detection for industrial robots and unknown obstacles using multiple depth images, in *Depth Images, German Workshop on Robotics* (Technical University of Braunschweig, Braunschweig, Germany, 2009)
27. M Van den Bergh, L Van Gool, Combining rgb and tof cameras for real-time 3d hand gesture interaction, in *Applications of Computer Vision (WACV), 2011 IEEE Workshop On* (Kona, Hawaii, 2011), pp. 66–72
28. S Park, S Yu, J Kim, S Kim, S Lee, 3d hand tracking using kalman filter in depth space. *EURASIP J. Adv. Sig. Proc.* 36–2012 (2012)
29. B Bartczak, I Schiller, C Beder, R Koch, Integration of a time-of-flight camera into a mixed reality system for handling dynamic scenes, moving viewpoints and occlusions in real-time, in *In Proceedings of the International Symposium on 3D Data Processing, Visualization and Transmission Workshop, Georgia Institute of Technology, Atlanta, GA, USA, (2008)*
30. A Kolb, E Barth, R Koch, Tof-sensors: new dimensions for realism and interactivity, in *Computer Vision and Pattern Recognition Workshops, 2008. CVPRW'08. IEEE Computer Society Conference On* (Anchorage, Alaska, USA, 2008), pp. 1–6
31. A Kolb, E Barth, R Koch, R Larsen, Time-of-flight cameras in computer graphics. *Comput. Graph. Forum.* **29**(1), 141–159 (2010)
32. S Ghobadi, O Loepprich, O Lottner, K Hartmann, O Loffeld, W Weihs, Analysis of the personnel safety in a man-machine-cooperation using 2d/3d images, in *Proceedings of the EURON/IARP International Workshop on Robotics for Risky Interventions and Surveillance of the Environment*, ed. by Pender J, Baudoin EMR, and Cervera Y, vol. 4 (Benicassim – Spain, 2008), p. 59
33. M Fischer, D Henrich, Surveillance of robots using multiple colour or depth cameras with distributed processing, in *Distributed Smart Cameras, 2009. ICDS 2009. Third ACM/IEEE International Conference On* (Como, Italy, 2009), pp. 1–8
34. C Walter, C Vogel, N Elkmann, A stationary sensor system to support manipulators for safe human-robot interaction, in *Robotics (ISR), 41st International Symposium on and 2010 6th German Conference on Robotics (ROBOTIK)* (Curran Associates, Inc. sMunich, 2010), pp. 1–6
35. R Lange, *3d time-of-flight distance measurement with custom solid-state image sensors in cmos/ccd-technology*, PhD thesis. (University of Siegen, Siegen, Germany, 2000)
36. S Clayton, Kinect for Windows SDK to Arrive Spring 2011. http://blogs.technet.com/b/microsoft_blog/archive/2011/02/21/kinect-for-windows-sdk-to-arrive-spring-2011.aspx. Accessed 21 April 2014
37. J Han, L Shao, D Xu, J Shotton, Enhanced computer vision with microsoft kinect sensor: a review. *IEEE Trans. Cybern.* **43**(5), 1318–1334 (2013)
38. L Liu, L Shao, Learning discriminative representations from rgb-d video data, in *Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence (AAAI Press/International Joint Conferences on Artificial Intelligence Beijing, 2013)*, pp. 1493–1500
39. F Flacco, T Kroger, A De, O Luca Khatib, A depth space approach to human-robot collision avoidance, in *Robotics and Automation (ICRA), 2012 IEEE International Conference On* (St Paul, MN, USA, 2012), pp. 338–345
40. S Fuchs, Calibration and multipath mitigation for increased accuracy of time-of-flight camera measurements in robotic applications. PhD thesis, (Technische Universität Berlin, 2012)
41. T Kahlmann, Range imaging metrology: Investigation, calibration and development. PhD thesis, Institute of Geodesy and Photogrammetry. (ETH Zurich, 2007)
42. M Lindner, A Kolb, Lateral and depth calibration of pmd-distance sensors. *Ger. Res.* **4292**(4292/2006), 524–533 (2006)
43. F Chiabrando, R Chiabrando, D Piatti, F Rinaudo, Sensors for 3d imaging: metric evaluation and calibration of a ccd/cmos time-of-flight camera. *Sensors.* **9**(12), 10080–10096 (2009)
44. MJD Powell, *Approximation Theory and Methods*. (Cambridge University Press, New York, 1981)
45. (MESA Imaging AG, Zurich, Switzerland, 2008)
46. S May, 3D Time-of-flight ranging for robotic perception in dynamic environments (Doctoral Dissertation). Dusseldorf VDI-Verl. Univ. Osnabrück, Germany (2009)
47. S Robla, JR Llata, C Torre, EG Sarabia, An approach for tracking oil slicks by using active contours on satellite images, in *OCEANS 2009 - EUROPE*, (2009), pp. 1–8
48. ER Davies, *Machine Vision, Third Edition: Theory, Algorithms, Practicalities (Signal Processing and Its Applications)*, 3rd edn. (Elsevier, Morgan Kaufmann, University of London, UK, 2005)
49. N Friedman, S Russell, Image segmentation in video sequences: a probabilistic approach, in *Proceedings of the Thirteenth Conference on Uncertainty in Artificial Intelligence (UAI'97 Morgan Kaufmann, San Francisco, Providence, 1997)*, pp. 175–181
50. M Piccardi, Background subtraction techniques: a review, in *Systems, Man and Cybernetics, 2004 IEEE International Conference On*, vol. 4 (The Hague, The Netherlands, 2004), pp. 3099–31044
51. M Cristani, M Farenzena, D Bloisi, V Murino, Background subtraction for automated multisensor surveillance: a comprehensive review. *EURASIP J. Adv. Sig. Proc.* **2010**, 1–24 (2010)
52. S Fuchs, S May, Calibration and registration for precise surface reconstruction with tof cameras, in *International Journal of Intelligent Systems Technologies and Applications*, vol. 5 (Inderscience Publishers, 2008), pp. 274–284
53. J-Y Bouguet, Camera Calibration Toolbox for Matlab (2000). http://www.vision.caltech.edu/bouguetj/calib_doc/. Accessed 03 March 2010
54. R Hartley, A Zisserman, *Multiple View Geometry in Computer Vision*, 2nd edn. (Cambridge University Press, New York, 2003)
55. J Heikkilä, O Silven, A four-step camera calibration procedure with implicit image correction, in *Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference On*, (1997), pp. 1106–1112

doi:10.1186/1687-6180-2014-88

Cite this article as: Robla et al.: Visual sensor fusion for active security in robotic industrial environments. *EURASIP Journal on Advances in Signal Processing* 2014 **2014:88**.