Directional People Counter Based on Head Tracking

Jorge García, Alfredo Gardel, Ignacio Bravo, José Luis Lázaro, Miguel Martínez, and David Rodríguez

***Abstract*—This paper presents an application for counting people through a single fixed camera. This system performs the count distinction between input and output of people moving through the supervised area. The counter requires two steps: detection and tracking. The detection is based on finding people's heads through preprocessed image correlation with several circular patterns. Tracking is made through the application of a Kalman filter to determine the trajectory of the candidates. Finally, the system updates the counters based on the direction of the trajectories. Different tests using a set of real video sequences taken from different indoor areas give results ranging between 87% and 98% accuracies depending on the volume of flow of people crossing the counting zone. Problematic situations, such as occlusions, people grouped in different ways, scene luminance changes, etc., were used to validate the performance of the system.**

***Index Terms*—Dynamic background subtraction, head detection, Kalman filter, people tracking.**

I. Introduction

NOWADAYS, there is a large demand for people-counting systems that offer greater efficiency and lower prices. These people-counting systems are widely used for tasks such as video surveillance, security, statistical analysis of people accessing an area, and other added value products.

Numerous studies have addressed the problem of people counting [1], [2]. In [3], a classification is presented in accordance with the type of sensor that the system uses, ranging from contact counters, which are not very effective because they reduce the flow of people, thus blocking the way, to photocells, passive infrared (PIR), microwave, etc. Most of them have the disadvantage of not being able to distinguish groups of people, their aim being only to detect the emptiness/occupancy of an area. Superior systems capable of counting people have a higher cost on the market, so they are solutions that price themselves out of widespread installation. An example of a high-level system might be a surveillance system based on a laser scanner [4] that consists of two lasers for resolving occlusions. Artificial vision systems are halfway houses which balance effectiveness and cost and attempt to overcome the drawbacks mentioned earlier by using off-the-shelf equipment.

There are stereo systems which use two cameras for people counting as proposed in [5] and [6]. In [6], people counting is performed by extracting an appearance vector based on a color region of interest (ROI) and a probabilistic model, using the stereoscopic disparity map to resolve possible uncertainties. Reference [7] presents a system for people counting in crowded environments; instead of executing a real count, it performs an estimate of the number of people from the occupancy volume. This system consists of a minimum of two cameras to prevent occlusions occurring in the scene. The foreground objects from each camera are obtained. Later, these are used to build up a flat projection of the supervised area, and from this projection, an estimate is made of the number of people.

Stereoscopic systems require a more complicated installation in order to precisely calibrate the stereo system and greater maintenance due to the possible undesirable movement of the camera housing. It is also worth pointing out that the greater the number of camera sensors, the higher the final price of the system. Thus, most people counter systems are based on monocamera systems which are more desirable due to lower installation and maintenance. In [8], the camera is placed overhead at a certain height. Two ROIs are defined at the top and bottom of the image. Next, the column histogram of the optical flow is computed in those areas. The number of people crossing the area is obtained from the histogram values considering a minimum threshold. The count is obtained from the information of blobs crossing the ROIs.

In [9], the position of the camera is also zenithal and performs optical flow. From this information considering two horizontal zones, optical flow values are accumulated vertically, and then, the number of persons crossing the zones is estimated. Neither system detects the shape of a person, such shapes possibly being confused with other objects such as cars, luggage, etc. There is great uncertainty when determining the number of people by area of occupation and using a ratio of pixels/person.

Also, as [9] shows, shadows or people crossing the area frequently will produce false positive detections. Finally, it is impossible to deal with the case of a person stopping and then continuing again. To deal with shadows in the image, one solution could be to use a stereo pair of cameras to take into account the circular shape of the head, as in our case. Disturbance shadows could be minimized if infrared LED rings coupled to the camera were used.

Other methods of people detection are based on background subtraction. With this information, in [10], tracking is based on area of occupancy and a color vector for each ROI. As the author states, this system does not perform properly for people walking together.

Manuscript received December 21, 2010; revised July 27, 2011, January 12, 2012, and May 5, 2012; accepted May 31, 2012. Date of publication July 6, 2012; date of current version May 2, 2013. This work was supported by the Spanish Research Program (Programa Nacional de Diseño y Producción Industrial, Ministerio de Ciencia y Tecnología), through the project "ESPIRA" (ref. DPI2009-10143), and by the University of Alcalá (ref.UAH2011/EXP-001), through the project "Sistema de Arrays de Cámaras Inteligentes."

The authors are with the Department of Electronics, Escuela Politécnica Superior, University of Alcalá, 28871 Alcalá de Henares, Spain (e-mail: jorge.garcia@depeca.uah.es; alfredo@depeca.uah.es; ibravo@depeca.uah.es; lazaro@depeca.uah.es; miguel.martinez@depeca.uah.es; david.rodriguez@depeca.uah.es).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIE.2012.2206330

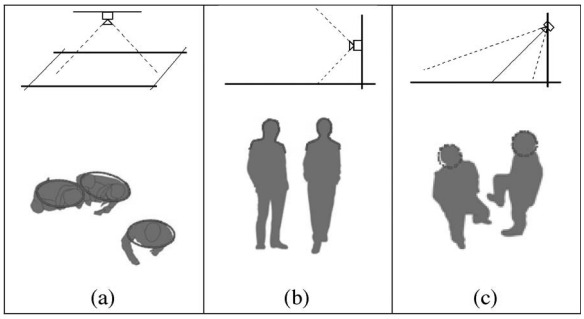

Fig. 1. Possible camera views. (a) Overhead view. (b) Front view. (c) Inclined view.

Other authors [11] propose a tilted camera system as in our system. Both systems require the detection of people heads in order to track each person detected in the monitored area. Reference [11] presents a system based on a support vector machine classifier to perform head detection. To track detected people, they also use a Kalman filter, but there is no feedback from the prediction of the tracking for the head detection. As the authors state, this system has a problem with occlusions that cannot be resolved.

The people-counting system we present here makes use of a single fixed camera positioned at a certain vertical angle. It aims to conform a system that processes errors in counting when a person stands in the area covered by the camera, that is to say, the false positives produced by objects that are associated with people like cars, bags, etc.

The main contributions of our innovative proposal are the use of an annular filter bank to quickly detect heads in the scene and the adaptive selection of head candidates to solve temporal errors or occlusions by using the estimates provided by the Kalman filter. Thus, the proposal solves the problem of people that are stopping and going and false positives from pets, handbags, shopping carts, etc.

The next section, Section II, discusses the significant characteristics of a people-counting system based on a single fixed camera. In Section III, the proposed computer vision algorithm is described. Section IV gives the results obtained when using the system in real video sequences. Finally, Section V sets out the conclusions and the main contributions of this paper.

II. Analysis of Scenario. Requirements and Constraints

People-counting systems operate within the ambit of people tracking in urban environments. These systems are located in transit areas such as entrances/exits of buildings, corridors, etc. The lighting in such locations is mostly artificial, with mainly constant features.

Another question is whether to use color cameras. Low-cost systems are often preferred for people counting; thus, color image processing is less suitable than gray-scale processing for addressing the embedded products.

An important feature of all counting systems is the location and orientation of the camera; thus, these characteristics are analyzed. An image taken from an overhead orientation is shown in Fig. 1(a). In this orientation, there is a drastic loss of features to distinguish a person from another object in the captured image. For example, the difficulty of separating the circular head shape from the body should be noted. Circular contour detection is related to the difference in luminance between the head and the body.

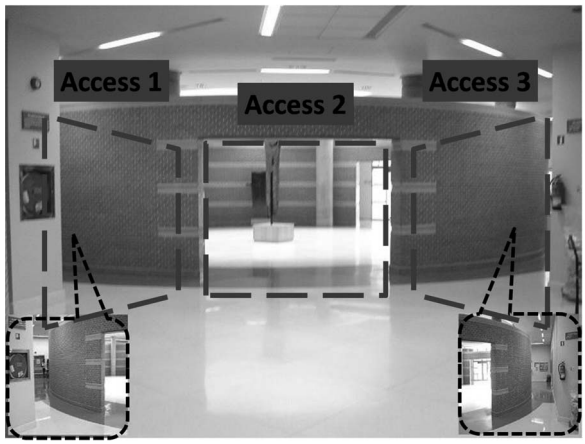

Fig. 2. Example of different accesses located in a large area.

Fig. 1(b) shows that this perspective could detect the shape of the head from background, but this location is often impossible to set up in real conditions. This position is used when there is the need for interaction between a robot and a person, as it is necessary if the person's movements are to be detected and tracked, as suggested in [12]. Our solution is to increase the height of the camera and direct it at an angle as is shown in Fig. 1(c), a position that increases the range covered, avoiding some occlusions. Head contours from head–body and head–background are both captured in the image.

The flatness of the floor does not introduce errors in the proposed algorithm. The perspective of people heads does not change enough, while the inclination of the camera with respect to the floor remains inside the considered range. If the floor contains steps or stairs, this situation is equivalent to a quick movement of the people. Both issues should be resolved by the tracking algorithm.

Another requirement is that the system should distinguish people from other types of "blobs" captured in the image. In tilted camera systems, the detection of a person's entire body is difficult because many occlusions usually occur [13]. The detection of head is more stable than other parts of the body such as arms or legs. In this way, the detection of an individual may be achieved by identifying the circular shape of his/her head.

Focal length, sensor image size, and camera height configure the supervised area. To cover larger zones, the scalability of the system should be possible, as shown in Fig. 2, which presents a situation in which three cameras count in–out people crossing collateral areas.

A minimum processing frame ratio is necessary to achieve enough information about the trajectory of each person and perform an accurate count. The expected maximum speed of people crossing the area should be taken into account. The average speed of a person is about 1–1.5 m/s. Considering a supervised area of 3–4 m with a standard 30 frames/s (fps),

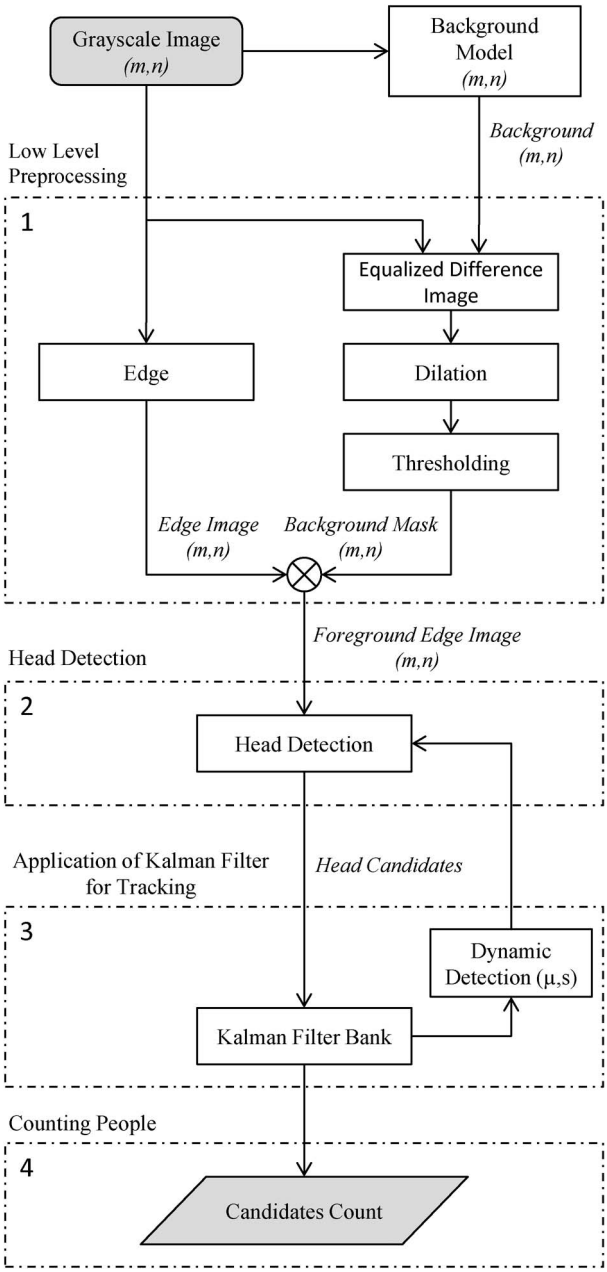

Fig. 3. Main steps in the proposed people-counting system.

the monocamera system captures 60–120 images of a person, which should be enough to perform a robust people tracking.

III. Algorithm Description

Fig. 3 shows the main steps of the proposed system/algorithm: low-level preprocessing (stage 1), head detection (stage 2), application of Kalman filter for tracking (stage 3), and counting people (stage 4).

An edge foreground image is generated in the preprocessing stage and used later to search for people by an annular filter bank for head detection. Next, tracking and counting are performed from the traces generated by the Kalman filter. The head detection makes use of information from the Kalman filter to achieve more robust detection. In what follows, each algorithm step is described in more detail.

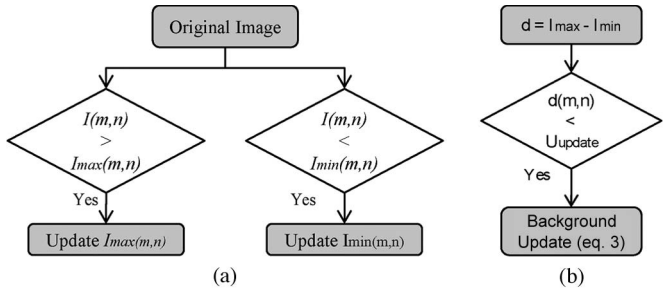

Fig. 4. Background update. (a) Comparison phase. (b) Update phase.

A. Background Model

For the proper operation of the system, a background image is dynamically updated to slowly capture small changes in scene illumination and introduce new static objects. This background model is affected by instant lighting changes or system vibration. These false foreground regions are considered in the head detection algorithm, which increases the processing time, but do not introduce counting errors.

The background modeling has been divided into two parts: a comparison phase to obtain min/max values for each image pixel and the update phase, as shown in Fig. 4. In the comparison phase [Fig. 4(a)], the original image is compared pixel by pixel with one image containing the maximum values $I_{\max}$ and another with minimum values $I_{\min}$.

If an original image pixel $I(m,n)$ is greater than the corresponding pixel from the maximum image $I_{\max}(m,n)$, the pixel value of the maximum image is modified with the image value. The comparison with the minimum-value image is similar.

This process takes place over a certain number of frames N_f. In our system, with a video frame rate of 30 fps, the value of N_f is ten, so the background is updated at a frequency of 3 Hz. For each N_f frame, the images $I_{\max}$ and $I_{\min}$ are initialized with the first frame of each run, which is copied into both images.

Once the max/min values of N_f frames are obtained, the update phase [Fig. 4(b)] modifies the background image accordingly. In the update phase, the difference between maximum and minimum images is obtained (pixel variation). Each pixel variation is compared with a threshold U_{update} fixed experimentally. Only the pixel coordinates with a low value for pixel variation, i.e., pixels belonging to the background, will be updated. Pixels with high variation are not updated as they represent pixels where there has been movement.

The update of a background pixel is performed using (1) and (2). In (1), the value of the pixel to update is given by $v_n(m,n)$, the mean value for the maximum and minimum values for that pixel coordinates. The pixel is updated progressively according to (2) so the background image does not present abrupt changes between consecutive updates

$$v_m(m,n) = \frac{I_{\max}(m,n) - I_{\min}(m,n)}{2} \quad (1)$$

$$\text{Back}_{\text{update}}(m,n) = \alpha \text{Back}(m,n) + (1-\alpha) v_m(m,n) \quad (2)$$

where parameter α determines the influence of the previous background value $\text{Back}(m,n)$ and a new value $v_m(m,n)$.

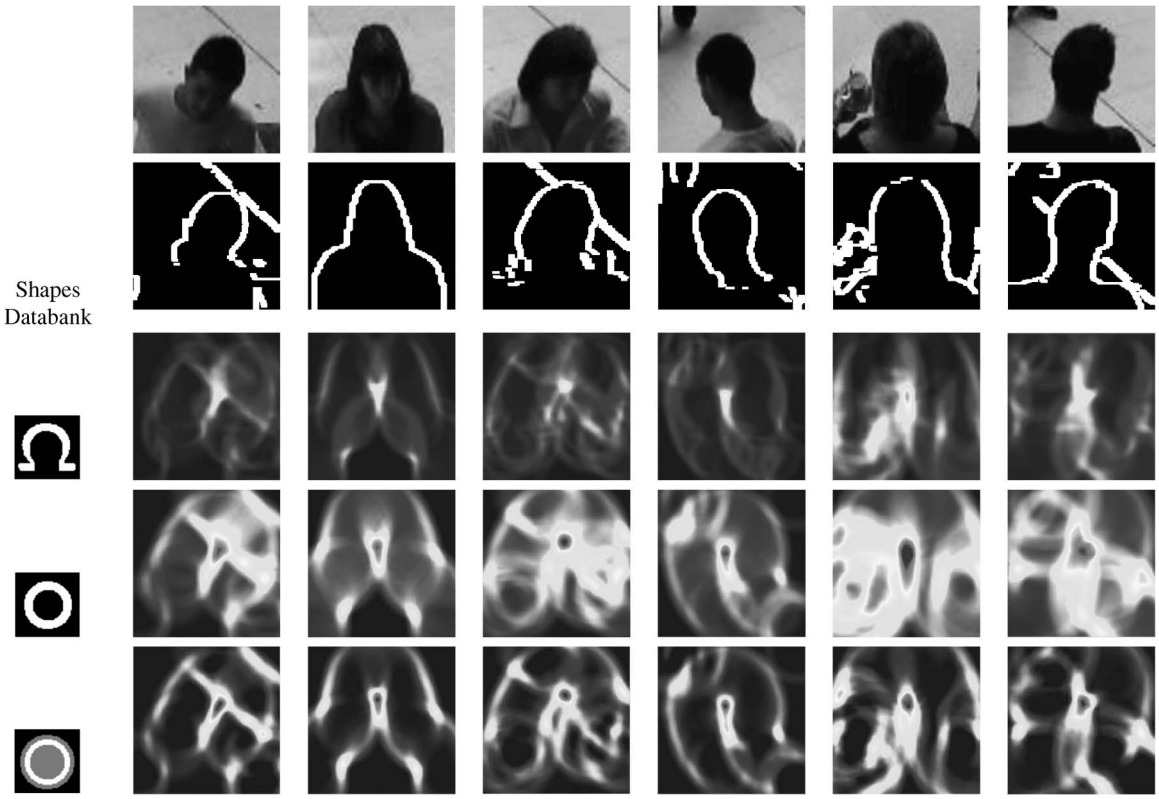

Fig. 5. Examples of different heads to be detected by synthetic shapes. First, we present the original images accompanied by their images of edges (*Sobel* detector). In the left column, the three synthetic shapes are shown: Omega shape, circular shape, and the proposed shape. Shown below are the results of the different correlations, where the blue color represents the lowest correlation, while red represents the highest values.

If new objects are introduced in the scene background, they will be smoothly updated to the image background every N_f frames.

B. Gray-Scale Image Preprocessing

A very common way to detect foreground objects in an image is by means of the background subtraction method [14]. In our case, the subtraction of background image is used to create an enlarged area that includes any possible edge from any object crossing the area. To obtain this mask, different low-level operations are carried out on the captured image.

First, the image difference I_{diff} is calculated from the original and background images. Next, the equalized difference image I_n is obtained by maximizing the contrast of the difference between [0, 255] using the normalization shown in

$$I_n(m,n) = 255x\frac{I_{\text{diff}}(m,n) - I_{\text{diff}_{\min}}}{(I_{\text{diff}_{\max}} - I_{\text{diff}_{\min}})}. \tag{3}$$

The I_n image is then dilated and binarized for use as a foreground mask (FM). The threshold value, 20, is empirical. This value is conservative in order to not eliminate any edge of interest.

Edge detectors used in the preprocessing stage commonly obtain information from an image [15] using a Canny edge detector to detect line segments in the images of a road. There are different techniques for obtaining the edge image (EI) [16]. In [17], some methods for edge detection are compared, before concluding that the Canny detector has more to recommend than other detectors such as *Sobel*. It should be noted that Canny extracts almost all the edges of the captured image, whereas *Sobel* highlights more the contours of objects [18]. For our system, it is more useful to highlight the outline of the person with a *Sobel* edge detector than to obtain any edge detected on the image since the latter detection is performed by means of the head contour. Then, the *EI* is obtained from the sum of the horizontal edges and vertical edges using a pair of size 5×5 *Sobel* masks.

The *EI* is masked with the *FM* to obtain the foreground EI *(FEI)*.

C. Head Detection

There are different methods in order to obtain the location of the head of people in an image. In [19], with the camera placed at face level, an Ω model is used to improve the efficiency of face tracking based on other features such as skin color, etc. Due to the restriction on the camera's location, the Ω model does not yield good detection values. The perspective of the image obtained of a person crossing the area undergoes significant changes, from a perspective view to a vertical view. Thus, our proposed system uses a more general elliptical model for head detection.

The proposed algorithm for head detection is to search for circular shapes through a 2-D correlation using a bank of annular patterns. This detection method is not dependent on person size in the image, and it does not introduce significant errors in the calculation of the number of people in the scene. Fig. 5 shows a comparison between the models described earlier and

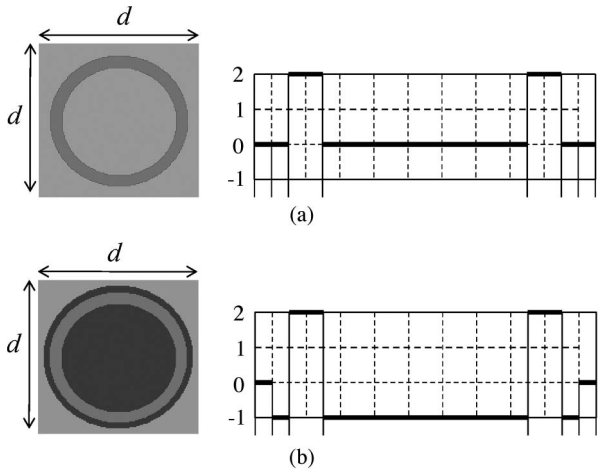

Fig. 6. Detection masks and profile values. (a) Basic shape circular. (b) Proposed shape.

concludes that the model chosen gives better results. A color camera does not introduce any advantages because the detection algorithm analyzes image contours.

Moreover, if the image size increases, a better analysis of contours can be done because the EI is better defined. In the head detection step, this effect produces an improvement because correlation values are higher, once results are normalized. In addition, this increase in size leads to increased computation time.

The detection algorithm is based on a bank of annular filters as shown in Fig. 6(a). Different proposals for annular models are presented. The basic model consists of a white ring mask. This model is not valid because areas with many edge pixels would give a high detection value not necessarily with a circular shape. The profile of the proposed model in our system introduces a penalty for edge pixels found inside the circular shape [Fig. 6(b)], being capable of distinguishing the head shape correctly. Different tests have been carried out to show that the correlation with the proposed annular filter gives better results.

Our system should be independent of the size of people captured in the image. Moreover, according to the distance from the person to the camera, the head size will be different. To solve this problem, a bank of N annular patterns with different sizes is defined. For our camera and area covered, empirical tests have shown that a bank composed of $N = 4$ masks provides enough resolution, but this value depends on the camera and size of the area supervised.

The effects of perspective mean that there is an area of the image where a determined size of annular pattern is more suitable. Thus, the *FEI* image is divided into different zones in order to assign a given annular pattern related to the size of the heads in that image area. As the zones overlap by 50%, two different sizes of patterns are applied each time. This improvement makes the processing more efficient and therefore faster. Then, two correlations with annular patterns are executed at each image pixel.

The correlation values are normalized after taking into account the values and size of the annular patterns applied. Then, 2-D correlation maxima are retrieved in order to ascertain the location of head candidates taking into account their ring affinity. Peak values that are lower than a fixed threshold U_c do not form part of the final set of coordinates. Nearby peaks from different annular correlations which might belong to the same head candidate are unified, discarding the lower value or, if they have the same values, taking the larger size correlation. Ring affinity values will be included in the state vector in order to obtain a more robust tracking. Assuming frame rates around 20–30 fps, minimum similarity differences are found between zones of interest in the person trajectory.

D. People Tracking

The people counting is based on the people trajectories. These are achieved owing to people crossing the area captured by the camera. By tracking detected objects in image sequences, different pathways are constructed. These time series are well suited for the application of a Kalman filter or any of its variants. Different techniques are used according to the nature of the process to be estimated. In the case of people tracking, the capture rate of the camera (in frames per second) should be sufficient to capture individual movements a number of consecutive times so that the estimation process may be considered linear in time. In that case, it is not necessary to use specific nonlinear processes, such as extended Kalman filter, unscented Kalman filter, or nonlinear/non-Gaussian particle filter [20] used for nonlinear processes.

The aim of these techniques is to obtain a good model to follow an object at each instant of time through an analysis of state variables [21]. In [22], Kalman filters used to estimate the orientation of the tool and the position of the center of mass of the tool are shown. In [23], they propose the use of the Kalman filter for face tracking in an image sequence. In [24], the use of Kalman filter is combined with the information from a 2-D stochastic model in order to identify the shape of a person within an image.

Other possible techniques, such as optical flow analysis, may be used for people tracking. These analyses involve high computation times and numerous detection errors [25]. Contributions have been made to solve these problems. An adaptive optical flow for people trajectories is presented in [25], considering image regions which present differences with respect to the estimated image background. The main restriction for these techniques is that people should be in continuous movement, while the detection of heads is not dependent on the people movement.

As in the aforementioned described cases, we propose the use of the estimation provided by the Kalman filter to generate the tracking of a person while also reinforcing the detection stage in an area of interest centered on the estimate. This contribution is achieved by reintroducing the prediction values provided by the Kalman filter to improve the detection of new head candidates in the next sample time. In addition to the common state variables to perform a trajectory tracking—the position and velocity of the person in the image—three further additional parameters are considered in the state vector: mean, standard deviation, and affinity of the model. These parameters are extracted from the head candidate and should be tracked

as another feature along the pathway. The calculation of these parameters is explained in the section *Data Association*. The state vector is shown in

$$\begin{bmatrix} X_k \\ Y_k \\ X'_k \\ Y'_k \\ \mu_k \\ s_k \\ V_k \end{bmatrix} = \begin{bmatrix} \text{coordinate } x \\ \text{coordinate } y \\ \text{velocity } x \\ \text{velocity } y \\ \text{mean of circular region} \\ \text{variance of circular region} \\ \text{ring affinity} \end{bmatrix}. \quad (4)$$

In large areas, using a tilted camera system, the motion model of a person can be compared with a model of constant acceleration [11]. In our case, the tracking area is small, about 3 or 4 m long, so that the motion movement of a person can be compared with a constant velocity model. Therefore, it is possible to track a person only on position and velocity in the image, not being necessary to know the real position and velocity of an individual. This approach avoids the use of a calibrated camera system with respect to the scene to get the real position of the individual.

1) Discrete Kalman Filter: The Kalman filter [26] is a recursive procedure consisting of two main stages: prediction and correction. The first stage aims to estimate the motion, while the second is responsible for correcting the error in the motion. The following paragraph describes these two stages.

1) Update in time (prediction).

Equation (5) determines the evolution of the state vector $\hat{X}_k^*$ from the motion model defined by A matrix (6) and the previous state vector of the object. Matrix A defines a constant velocity model, where parameter T indicates the time between two consecutive measurements. Second, the projection of error covariance (7) is calculated P_k^*. Matrix $W \sim N(0, Q)$ represents the noise in the process, where Q represents the covariance of the random disturbance of the process that attempts to estimate the state

$$\hat{X}_k^* = A\hat{X}_{k-1}^* + W \quad (5)$$

$$A = \begin{bmatrix} 1 & 0 & T & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & T & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad (6)$$

$$P_k^* = AP_{k-1}A^T + Q. \quad (7)$$

2) Update of observation (correction).

In (8), the value of the Kalman constant K_t is calculated from the *a priori* error covariance P_k^*, matrix H (11) linking the state to the measurement, and R which represents the covariance of the random disturbance of the measurement. The estimate is updated with new measurements obtained from the process using (9), and the new estimated state vector $\hat{X}_k$ is achieved. The error covariance is updated using (10). Finally, in (12), the output $\hat{Y}_k$ is obtained, where $V \sim N(0, R)$ represents the noise in the measurement

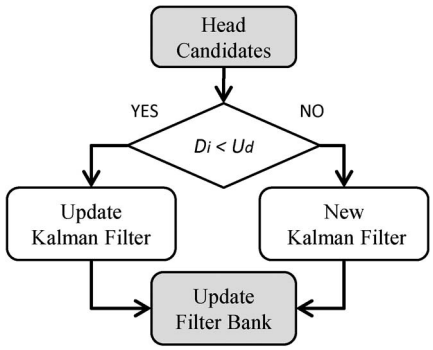

Fig. 7. Data association algorithm.

$$K_k = P_k^* H^T \left(HP_k^* H^T + R\right)^{-1} \quad (8)$$

$$\hat{X}_k = \hat{X}_k^* + K_k \left(Z_k + H\hat{X}_k^*\right) \quad (9)$$

$$P_k = (I - K_k H)P_k^* \quad (10)$$

$$H = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad (11)$$

$$\hat{Y}_k = H\hat{X}_k^* + V. \quad (12)$$

In order to obtain an efficient and robust data filtering (convergence filter), a correct definition of the parameters Q and R is needed. However, in real conditions, there is little or no knowledge of these parameters; therefore, they are obtained experimentally *a priori*.

2) Data Association: The Kalman filter does not behave in multimodal manner, i.e., each filter is capable of representing only one estimate. Thus, for each new object detected, it is necessary to implement one new filter. Each head candidate (measurement) is evaluated to determine whether it belongs to a particular filter (see Fig. 7). Taking into account these restrictions, it is necessary to design a management system for multiple hypotheses and multiple objects. The nearest neighbor approach is adopted to achieve the track association. This approach is commonly used for tracking objects whose state has little interaction or is based on a very specific likelihood model, as it is our case. More advanced techniques have been proposed in order to improve the efficiency. However, the choice of this approach is also justified by its low computational load and high performance in embedded systems.

The association of a head candidate to a tracking filter is performed using the minimum Euclidean distance D_i between the estimation filter (est_i) and the measurement set $(m_1, m_2, \ldots, m_j)$, as expressed in (13) and (14), where $\triangle$ represents the difference of the magnitude between the prediction filter and the measurement. The value of the Euclidean distance

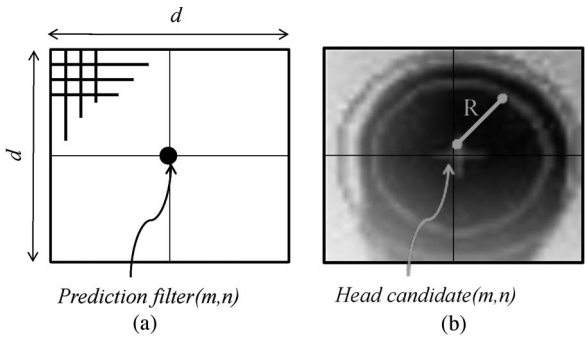

Fig. 8. (a) ROI (ROI_search) for search centered on the prediction. (b) Zoom of the region in the original image in gray scale for the calculation of parameters: Mean and variance, centered on the measurement associated.

D_i is compared with a threshold U_d (fixed empirically); if the distance value D_i is less than the threshold U_d, the measurement is associated to that filter. Otherwise, the measurement generates a new filter

$$d_{\mathrm{est}_i,m_j} = \sqrt{(\triangle x)^2+(\triangle y)^2+(\triangle \mu)^2+(\triangle s)^2+(\triangle V)^2} \quad (13)$$

$$D_i = \min\left\{d_{\mathrm{est}_i,m_0}, d_{\mathrm{est}_i,m_1}, \ldots, d_{\mathrm{est}_i,m_j}\right\}. \quad (14)$$

Temporal errors or partial occlusions at the head detection stage could lead to a state where there are tracking filters without any correspondence measurement. For those cases, a new ROI (ROI_search) centered at coordinates (x, y) of prediction is considered for a new refined processing [see Fig. 8(a)].

Then, from the ROI_search, the highest value of correlation ($ring\ affinity$) is obtained and compared with the threshold U_c. The search process starts from the predicted Kalman coordinates, so if some points have the same maximum value, the nearest point is considered. If that correlation value is greater than the threshold U_c, that coordinate is considered to be a peak of correlation, and its location/velocity is used in the Kalman filter under analysis. As noted earlier, this process enhances the detection stage, generating measurements that could not be detected in the previous process.

Those filters that have not been assigned any measurement are not directly eliminated, but their predictions are fed up to L consecutive times. If the value L (live time) is exceeded, then that instance of the filter is removed. The maximum value of L is related to the update time Nf of the background image. In the proposed algorithm, the value of L varies according to the history of the path, considering the total number of tracking points. The initial value L for a new instance of the filter is low (3), in order to eliminate false positive errors. The value of L is increased with new valid detections or reduced if there are no valid detections.

As discussed earlier, each measurement for each Kalman filter is formed by five parameters. The first two parameters are the correlation peak coordinates (x, y). The value $ring\ affinity$ is also introduced as another measurement to track as it was calculated at the head detection stage. Fig. 8(b) shows the circular region with radius R and centered on the coordinates (x, y) used in the calculation of several statistics:

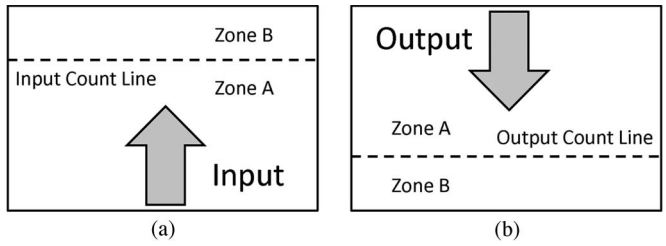

Fig. 9. Counting cases. (a) Input. (b) Output.

mean (μ) and standard deviation (s), as expressed in (15) and (16), respectively

$$\mu = \frac{\sum \mathrm{pixel}_{(x,y)}}{N}_{|\mathrm{circular\ region}} \quad (15)$$

$$s^2 = \frac{1}{N}\sum \left|\mathrm{pixel}_{(x,y)} - \mu\right|^2_{|\mathrm{circular\ region}} \quad (16)$$

where $\mathrm{pixel}_{(x,y)}$ is the pixel value of coordinates (x, y) and N is the total number of pixels of the circular region. The parameter R, which defines the area of interest for calculating the mean and standard deviation, should be appropriate to the size of the head under analysis. To do this, R is determined as the radius of the inside penalty area [Fig. 7(b)], given a detection profile. This value is updated in each sample time according to the detection profile corresponding to the latter measure added in its tracking.

3) Counting People: Different interpretations for counting people could be applied. Here, we present the traditional approach, counting by means of two virtual lines, one line for input and another line for output. When an object crosses the area, the trajectory followed is evaluated, and the in–out counting is updated. These lines represent relative positions to decide when a person has entered or exited. In any case, the pose of a person is not determined with respect to the scene.

Fig. 9 shows the different options, case (a) representing the input count and case (b) representing the output count. As may be seen, in both cases, two different zones are defined, zones A and B. To update the in–out counting, the path analyzed must start in A and end in B, a minimum lifetime being considered for each trajectory analyzed.

IV. Results

The camera used in the tests is a standard *Firewire* camera configured with a resolution of 320 × 240 pixels and 30 fps. Different scenes were recorded such as the following: doors, cafeteria accesses, entry to public centers, etc. In the processed image sequences, these videos contain many specific situations such as people with backpacks, handbags, etc.

After many tests, the camera was positioned as shown in Fig. 10. The camera was located 3 m from the ground and at an angle of 30 from the vertical view. Minor changes in the camera orientation are accepted (around ±5 from the optimal orientation) without increasing the error rate. The orientation of the camera was perpendicular to the flow of people in–out, so the video captures the main part of the path of the people entering the area. Fig. 11 shows the camera screwed to an aluminum support.

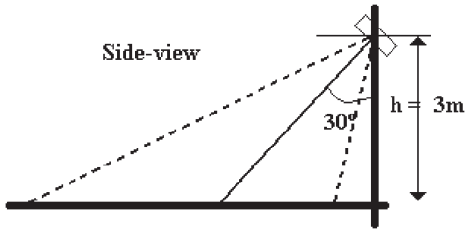

Fig. 10. Location and orientation of the camera in the proposed system.

Fig. 11. Real system image located on the access of a room from a corridor.

TABLE I
STATISTICS ON THE NUMBER OF PEOPLE CROSSING

People / Area	Total people (800)	% cases from total	% Detection
1	320	40%	98%
2	280	35%	96%
3	120	15%	93%
4	48	6%	91%
5 or +	32	4%	87%

Table I shows the counting results of the proposed system depending on the number of people crossing the counting zones. The first column identifies the number of people crossing the counting area simultaneously. The second and third columns show the number of people in each category displayed on the video test and the percentage of the total, respectively. Finally, the fourth column shows the percentage of detection and counting for each category. These values have been obtained without discriminating the counting errors.

As can be seen in Fig. 12, the detection error increases with the number of people in the counting area, causing a decrease in the effectiveness of the system.

It is worth noting that the lower the probability of compact groups of people crossing the area, the higher the number of people in the group. Many times, the detection errors are due to total occlusions that occur in the testing videos. These errors cannot be resolved due to the configuration of the system (height and orientation of the camera). In some cases, when

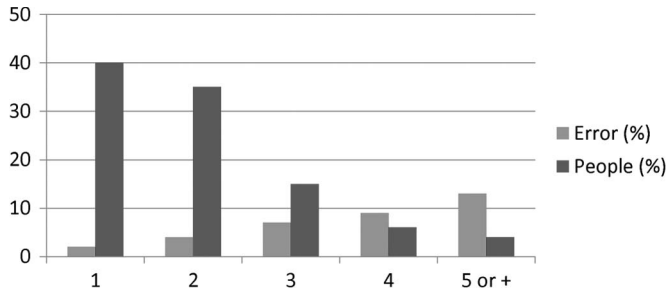

Fig. 12. Graphical error and people depending on the number of people crossing the counting zone, expressed as a percentage.

the occlusion occurs over a short period of time, the problem is solved by the use of the Kalman filter which estimates the trajectory of the individual. Moreover, false negatives occur when the contrast between the person and the background is weak, thus producing less than desired detection levels. Other errors (false positives) occur if the correlation with a border area has a head similar annular appearance and generates a wrong detection. For different objects, their edges change shape over the trajectory frames, so their circular shape disappears, and instead of progressing, the pathway is eliminated. Other cases such as when a person stops within the counting area, the tracking does not stop because the detection step is based on pattern recognition. Different systems based on optical flow of the people have problems in these situations.

In test videos, some shades were added to the foreground object; this event did not cause further problems because, in these regions of interest, there was no head detection. With background updating working properly, various static elements were introduced into the scene and were added perfectly to the background image.

All experiments were performed with a Core2Duo 2.6-GHz notebook, achieving a processing rate of approximately 10 fps. The algorithm has been codified in *C++* language using open source libraries *OpenCV 2.1*, and it is executed without multiple-thread support. In the near future, to increase the frame rate, the computer vision algorithm will be ported to a multiple-thread implementation with GPU parallel computation. Another line of implementation should take into account setting up an embedded system, such as hardware/software (HW/SW) codesign in a single field programmable gate array (FPGA). Hardware blocks would be responsible of convolutions with the different annular patterns, considerably increasing the performance. Other sequential parts of the algorithm would be executed by the embedded microprocessor inside the FPGA.

V. CONCLUSION

In this paper, a directional people-counting system through a single fixed camera has been presented. Tracking people is based on a head-detection procedure using a bank of annular patterns and a Kalman filter to smoothly follow the path and make for a more robust system. One key point is the feedback between the tracking and detection stages, allowing a more robust algorithm to be achieved, resolving temporal errors and partial occlusions that could occur in real-image sequences such as the test videos used.

This system determines the people count for different numbers of people crossing simultaneously the counting area in different directions.

The performance of the algorithm is not dependent on the background, as it encounters no problems with new static elements, and is capable of updating small changes in the background scene. Floor markings are not necessary. The system is easily installable and does not require highly skilled labor.

The effectiveness of our system decreases slightly with increasing numbers of people in the scene. This is mainly due to complete occlusions that occur in the test videos. Also, some false negatives are due to low pattern-head coincidence.

Summarizing, the test videos evaluated have obtained a weighted average of 97.67% effectiveness without discriminating the counting errors mentioned earlier.

REFERENCES

[1] A. Chan and N. Vasconcelos, "Counting people with low-level features and Bayesian regression," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 2160–2177, Apr. 2012.

[2] Y.-L. Hou and G. Pang, "People counting and human detection in a challenging situation," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 41, no. 1, pp. 24–33, Jan. 2011.

[3] S. Velipasalar, Y.-L. Tian, and A. Hampapur, "Automatic counting of interacting people by using a single uncalibrated camera," in *Proc. IEEE Int. Multimedia Expo. Conf.*, 2006, pp. 1265–1268.

[4] J. H. Lee, Y.-S. Kim, B. K. Kim, K. Ohba, H. Kawata, A. Ohya, and S. Yuta, "Security door system using human tracking method with laser range finders," in *Proc. ICMA*, 2007, pp. 2060–2065.

[5] N. Cottini, M. De Nicola, M. Gottardi, and R. Manduchi, "A low-power stereo vision system based on a custom CMOS imager with positional data coding," in *Proc. 7th Conf. PRIME*, Jul. 2011, pp. 161–164.

[6] G. Englebienne, T. van Oosterhout, and B. Krose, "Tracking in sparse multi-camera setups using stereo vision," in *Proc. 3rd ACM/IEEE ICDSC*, 2009, pp. 1–6.

[7] C. Fookes, S. Denman, R. Lakemond, D. Ryan, S. Sridharan, and M. Piccardi, "Semi-supervised intelligent surveillance system for secure environments," in *Proc. ISIE*, Jul. 2010, pp. 2815–2820.

[8] L. Rizzon, N. Massari, M. Gottardi, and L. Gasparini, "A low-power people counting system based on a," in *Proc. IEEE ISCAS*, 2009, p. 786.

[9] J. Barandiaran, B. Murguia, and F. Boto, "Real-time people counting using multiple lines," in *Proc. 9th Int. WIAMIS*, 2008, pp. 159–162.

[10] T.-H. Chen, T.-Y. Chen, and Z.-X. Chen, "An intelligent people-flow counting method for passing through a gate," in *Proc. IEEE Conf. Robot., Autom. Mechatron.*, 2006, pp. 1–6.

[11] H. Xu, P. Lv, and L. Meng, "A people counting system based on head–shoulder detection and tracking in surveillance video," in *Proc. ICCDA*, 2010, vol. 1, pp. V1-394–V1-398.

[12] P. Vadakkepat, P. Lim, L. C. De Silva, L. Jing, and L. L. Ling, "Multimodal approach to human-face detection and tracking," *IEEE Trans. Ind. Electron.*, vol. 55, no. 3, pp. 1385–1393, Mar. 2008.

[13] Y. Ishii, H. Hongo, K. Yamamoto, and Y. Niwa, "Face and head detection for a real-time surveillance system," in *Proc. 17th ICPR*, 2004, vol. 3, pp. 298–301.

[14] S. Yu, X. Chen, W. Sun, and D. Xie, "A robust method for detecting and counting people," in *Proc. ICALIP*, 2008, pp. 1545–1549.

[15] T. Bucher, C. Curio, J. Edelbrunner, C. Igel, D. Kastrup, I. Leefken, G. Lorenz, A. Steinhage, and W. von Seelen, "Image processing and behavior planning for intelligent vehicles," *IEEE Trans. Ind. Electron.*, vol. 50, no. 1, pp. 62–75, Feb. 2003.

[16] M. Sharifi, M. Fathy, and M. T. Mahmoudi, "A classified and comparative study of edge detection algorithms," in *Proc. Int Inf. Technol., Coding Comput. Conf.*, 2002, pp. 117–120.

[17] M. Hassan, N. E. A. Khalid, A. Ibrahim, and N. M. Noor, "Evaluation of Sobel, Canny, Shen & Castan using sample line histogram method," in *Proc. ITSim*, 2008, vol. 3, pp. 1–7.

[18] Q. Liao, J. Hong, and M. Jiang, "A comparison of edge detection algorithm using for driver fatigue detection system," in *Proc. 2nd Int ICIMA*, 2010, vol. 1, pp. 80–83.

[19] R. Patil, P. E. Rybski, T. Kanade, and M. M. Veloso, "People detection and tracking in high resolution panoramic video mosaic," in *Proc. IEEE/RSJ IROS*, 2004, vol. 2, pp. 1323–1328.

[20] M. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking," *IEEE Trans. Signal Process.*, vol. 50, no. 2, pp. 174–188, Feb. 2002.

[21] Z. Wang and D. Gu, "Cooperative target tracking control of multiple robots," *IEEE Trans. Ind. Electron.*, vol. 59, no. 8, pp. 3232–3240, Aug. 2012.

[22] S.-H. P. Won, F. Golnaraghi, and W. W. Melek, "A fastening tool tracking system using an IMU and a position sensor with Kalman filters and a fuzzy expert system," *IEEE Trans. Ind. Electron.*, vol. 56, no. 5, pp. 1782–1792, May 2009.

[23] Z. Shaik and V. Asari, "A robust method for multiple face tracking using Kalman filter," in *Proc. 36th IEEE AIPR*, 2007, pp. 125–130.

[24] G. Rigoll, S. Eickeler, and S. Muller, "Person tracking in real-world scenarios using statistical methods," in *Proc. 4th IEEE Int. Autom. Face Gesture Recog. Conf.*, 2000, pp. 342–347.

[25] S. Denman, V. Chandran, and S. Sridharan, "Adaptive optical flow for person tracking," in *Proc. Digital Image Comput., Tech. Appl.*, 2005, pp. 1–8.

[26] G. Welch and G. Bishop, "An Introduction to the Kalman Filter," Univ. North Carolina, Chapel Hill, NC, 1995, Tech. Rep.

Jorge García received the B.S. degree in telecommunications engineering and the M.Sc. degree in electronics system engineering from the University of Alcalá, Alcalá de Henares, Spain, in 2009 and 2011, respectively, where he is currently working toward the Ph.D. degree in electronics.

Since 2009, he has been with the Department of Electronics, University of Alcalá. His current research interests include computer vision and system based on FPGAs.

Alfredo Gardel received the B.S. degree in telecommunication engineering from the Polytechnic University of Madrid, Madrid, Spain, in 1999, and the Ph.D. degree in telecommunication from the University of Alcalá, Alcalá de Henares, Spain, in 2004.

Since 1997, he has been a Lecturer with the Department of Electronics, University of Alcalá. His main areas of research comprise infrared and computer vision, monocular metrology, robotics sensorial systems, and the design of advanced digital systems.

Ignacio Bravo received the B.S. degree in telecommunications engineering, the M.Sc. degree in electronics engineering, and the Ph.D. degree in electronics from the University of Alcalá, Alcalá de Henares, Spain, in 1997, 2000, 2007, respectively.

Since 2002, he has been a Lecturer with the Department of Electronics, University of Alcalá, where he is currently an Associate Professor. His areas of research are reconfigurable hardware, vision architectures based in FPGAs, and electronic design.

José Luis Lázaro received the B.S. degree in electronic engineering and the M.S. degree in telecommunication engineering from the Polytechnic University of Madrid, Madrid, Spain, in 1985 and 1992, respectively, and the Ph.D. degree in telecommunication from the University of Alcalá, Alcalá de Henares, Spain, in 1998.

Since 1986, he has been a Lecturer with the Department of Electronics, University of Alcalá, where he is currently a Professor. His areas of research are the following: robotics sensorial systems by laser, optical fibers, infrared and artificial vision, motion planning, monocular metrology, and electronics systems with advanced microprocessors.

Miguel Martínez received the B.S. degree in telecommunication engineering from the University of Alcalá, Alcalá de Henares, Spain, in 2010, where he is currently working toward the Ph.D. degree in electronic engineering.

His current research interests include machine learning, battery modeling, battery management systems, and intervehicle communications.

David Rodríguez received the B.Sc. degree in telecommunications technical engineering from the Miguel Hernández University of Alicante, Alicante, Spain, in 2011.

He is currently with the Department of Electronics, University of Alcalá, á de Henares, Spain. His research interests are related to systems based on FPGAs and optical positioning systems.