Surveillance Patrol Robot for People Tracking in Indoor Environments

Neerparaj Rai, Shakti Dhar, Rupam Kakati

***Abstract*-There is a great challenge that a mobile robot reliably and continuously tracks a specific person in indoor environments. In this paper, a novel method is presented, which can effectively recognize and reliably track a target person based on mobile robot vision. Such a robot is equipped with a camera which senses a moving object and starts tracking the object. The on-board camera develops a computer vision system for detection of the object/target to control and guide the movement of mobile robot. In order to effectively track the specific person, upper body vision clothes region is proposed for extracting the pattern features. The system applies Centre-Of-Mass based computation, filtering and color segmentation algorithm to locate the target and the position of the robot. Artificial Neural Network (ANN) is introduced for controlling the robot to follow the person with voice aided instructions from the robot. Experimental results validate the robustness and the reliability of this approach.**

Keywords: Mobile Robot, People tracking, Color segmentation, Filtering.

I. INTRODUCTION

The origins of robot manipulators date back to the 1940s, when Walters invented "Machina" the first robotic manipulator. Then, Devol invented the first industrial robot, "Unimate" in the late 1950s. Since then, industrial robot arms have proliferated and in today's automated industries, robot arms are well employed in various tasks on the factory floor. The current technological revolution in robotics and automation has trans- formed the concept by accomplishing industrial tasks in a safer, optimized and much more efficient manner.

A service robot, which is designed to serve people in special domains or help them in their daily life, must be able to detect and continuously track its users. The detection and reliable tracking of people in real time is a difficult problem in dynamic indoor environments, where lighting conditions are uncontrolled, multi- ple people wearing the similar color clothes are moving about, and temporary occlusions can occur at any time. The problem would become more intricate, if the robot need reliably track a recognized person in these complex environments. In order to enable a robot to automatically track the specific person, it is necessary to develop a technique that allows it robustly to recognize and track the person.

Recently, much work on detecting and tracking people with mobile robots has focused on visual methods [1–3]. There are also recent approaches that make use of the laser range finders to detect and track people [4,5]. A few approaches have attempted to use a fusion of multi-cues, such as combining visual and range information from laser and sonar sensors [6] and using additional microphones for sound source localization [7]. Most of these systems are not aimed at tracking a specific person. In this paper, a novel-tracking method is presented, which can continuously track a specific person based on monocular vision mounted on the mobile robot.

There are many tracking methods. A common approach is to employ predictive filtering and use the statistics of object's color and location in the distance computation while updating the object model by constant weights. When the measurement noises are assumed to be Gaussian, the optimal solution is provided by the Kalman filter [9]. When the process is non-linear, the EKF and the UKF were developed. The most general class of filters is represented by particle filters, which are based on Monte Carlo integration methods. Blake and Isard [10] introduced the particle filter to vision tracking. Although it provides tractable solutions to non-linear and non-Gaussian systems, it relies on important sampling and, as result, require the design of proposal distributions that can approximate the posterior distribution reasonably well. In general, it is hard to design such proposal. To solve this problem, Merve et al. [11] had developed the unscented particle filter (UPF), which utilized the UKF to generate proposal distributions.

In the context of visual-based surveillance applications, there are many conditions for which deciding a prior placement of vision sensor puts limits on system performance. We refer, for example, to those cases in which an alarm situation can occur with the same probability in any area of the monitored environment or to those situations that require the tracking of mobile objects in wide areas (e.g., a vehicle moving on a road or a people moving in a building). In these situations, especially in the context of indoor environments, the employment of mobile robots equipped with specific visual sensors for surveillance purposes can become an important issue. The integration of a mobile robot in a visual-based surveillance system can allow the coverage of all types of

environments, can extend the perceiving capabilities of the system (e.g., acquire images of higher quality by reducing the distance from the camera to the target).

II. PLATFORM DESCRIPTION

Figure 1 shows the experimental setup of Person Tracking Robot (PTR). The overall control system one Webcam and Laptop mounted on a robot with gear motors. The surveillance robot also includes one microcontroller and its necessary external circuitry for its control operation. Each motor control is done by sending a PWM (pulse width modulation) signal, a series of repeating pulses of variable width. The proposed system is a 2D-Image based vision system which is equipped with an intelligent image analysis and object detection algorithm that was developed in MATLAB®. The corresponding inputs to PC are the position error I_{ex} in horizontal axis with respect to the object image centre i.e., (I_x, I_y) and the rectangular area of the object and also the distance of the object from PTR. The details can be found in Section 3 and the output is Δθ for corresponding gear motors. The algorithms, including image-processing algorithm for Webcam, and the routines for receiving the image information and sending the reference command, are implemented in the Laptop.

Figure 1. Image of the complete setup used in this paper

The proposed framework combines control and image-processing to perform desired operations. In order to find the specific person and reliably track him, the robot needs to first find the target. PTR follows a circular path for finding the target. In a typical application, once a user's command is encountered, the workspace is scanned and corresponding images are captured. These images are processed to identify target object and their coordinates. The acquired coordinates are then passed to the Laptop (Matlab Program) to compute the necessary movement required to reach the target position.

III. PROPOSED METHODOLOGY

The proposed system has been build for performing two main activities: vision and control. The vision process aims to identify all objects moving in the scene and to verify which one must be considered the most important to monitor. To achieve this objective, a tracking module is involved. The motion detection module aims to solve the problem of the detection of moving objects in the scene. This problem has been addressed by applying an image differencing technique after the alignment of two consecutive frames I(x, t) and I(x, t + 1). The threshold output of this process is a binary image I(x, y) representing the pixels belonging to moving objects. Once the objects have been identified, a tracking module is applied to maintain track of their movements, by maintaining the objects inside the field of view of the camera.

3.1. Image acquisition

This intermediate block was focused on describing and applying standard signal processing techniques in images [20]. These techniques cover image enhancement algorithms such as noise reduction, filtering, contrast adaptation, etc., and also image analysis procedures applied after segmentation and morphological filtering, such as size, position, orientation, distance and average color estimate. The Matlab environment has enormous library of functions (or toolbox) dedicated to "*Image Acquisition*". This toolbox enables simple, fast and powerful access to image grabbers or cameras connected to the computer either with the USB or FireWire busses.

However, this toolbox is an optional purchase and may not be available in a standard computer laboratory facility equipped with Matlab. The image acquisition toolbox can be used to access image acquisition devices and store a sequence of images for offline analysis. The destination of the images acquired can be set as logging to disk or as logging to memory. Fig.2 shows the image acquired by PTR along with the target detection by rectangular box.

Figure 2. Image captured by Webcamera

The logging to disk mode has a 'preview'function that forces the connection of the camera and shows a preview of the last image acquired by the active camera. The most interesting feature is that this live image can be retrieved from the Matlab environment as an image matrix with a simple getsnapshot function. Fig. 3 shows an illustrative script with the basic toolbox functions required for image acquisition. The first step is the creation of the input video object link (for continuous image acquisition) with one camera connected to the computer with the function *videoinput* that requires as parameters the adaptor name, the number of the camera and the image format requested (which must be one of the formats offered by the camera). The format specified is the correct codification to obtain RGB color images with 24-bit depth and resolution of 640×480 pixels.

```
% input video object creation
% videoinput(ADAPTOR,DEVICE,FORMAT);
object = videoinput('winvideo',1,'RGB24_320x240');
% activate preview (opens an auxiliary window)
preview(object);
% starting the image acquisition loop
while condition
% get an RGB image matrix
image = getsnapshot(object);
% image processing and robot control
end
% delete the video object when finishing
delete(object);
```

Figure 3. Sample Matlab script to acquire live images with the "Image Acquisition" toolbox.

3.2. Conversion of Color Space

Most digital images use the *RGB* color space. However, individual *R*, *G*, and *B* components may have unstable variations under changing illumination conditions. On the other hand, it is easier to use the *Y UV* color space (where *Y* represents the luminance component, and *U*, *V* are chrominance components) for the segmentation of desired features. Vadakkepat *et al.* [22] verified that the *UV* color space for the face tracking problem is more effective and robust than that of the RGB. Hence, the color base of *RGB* is first transformed into that of *Y UV* with *U* and *V* ∈[0, 255] as

$$\begin{bmatrix} Y \\ U \\ V \end{bmatrix} = \begin{pmatrix} 0.299 & 0.587 & 0.114 \\ -0.169 & -0.331 & -0.500 \\ 0.5 & -0.419 & -0.018 \end{pmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} + \begin{pmatrix} 0 \\ 128 \\ 128 \end{pmatrix} \quad (1)$$

3.3. Image Segmentation

Since the value of Y is strongly dependent on luminance, it will not be considered in this paper. The values of U and V are set to different ranges as U ∈[Umin, Umax] and V ∈[Vmin, Vmax] for the identification of different objects, e.g., object (red color of the upper body) in Fig 3. Hence, the ranges [Umin, Umax] and [Vmin, Vmax] for the image features i.e., red color, are assigned as [0, 110] and [0, 120], respectively. The process of image segmentation is more robust in distinguishing the object on the ground with non uniform illumination, and strong reflection.

3.4. Binary

The type of image processing followed is real-time processing, which has fast image processing time and is immune to the varying of the size irrespective of how far the object is from the camera. Color filtering technique is used to extract the current position of the gripper and the static position of the object. The object of interest colors to be considered is red for the current position of the gripper and light green color for the object while other colors are discarded. The binary of the image is to choose a suitable threshold value *Tb* for *U* and *V*. The corresponding relation is

$$P(Ix, Iy) = \begin{cases} 1 \text{ if } f(Ix, Iy) \geq Tb \\ 0 \text{ if } f(Ix, Iy) < Tb \end{cases} \quad (2)$$

where f(*Ix, Iy*) denotes the values of *U* and *V* on the image plane (*Ix, Iy*), *P*(*Ix, Iy*) = 1 stands for the white pixels, and *P*(*Ix, Iy*) = 0 stands for the black pixels. The purpose of the binary is to reduce the storage amount as well as the computation load. The value of *Tb* is less sensitive to lighting conditions because the *Y* component is not considered for the binary operation. In Fig.5 the red color object are segmented from the background for further processing.

3.5. Filtering

A median filtering is used to remove noises produced by additional factors such as lighting intensities and the presence of unwanted particles. Median filtering is similar to using an averaging filter, in that each output pixel is set to an average of the pixel values in the neighbourhood of the corresponding input pixel. However, with median filtering, the value of an output pixel is determined by the *median* of the neighbourhood pixels, rather than the mean. The median is much less sensitive than the mean to extreme values (called *outliers*). Median filtering is therefore better able to remove these outliers without reducing the sharpness of the image. Then a Gaussian filter is used to further smoothen the image but will preserve edges better than the more basic mean filter. The resulting object is shown in Fig. 6 forWebcam**.** By weighting a pixels contribution to the final pixel value this filter can better preserve edges than the mean filter which specifies equal weights to all pixels within the filter window. For a 1-D Gaussian filter the single filter values are defined as

$$G(x) = \sqrt{\frac{1}{\sqrt{2\pi}}} \cdot e^{\frac{x^2}{2^2}} \quad (3)$$

Figure 4. Binary segmented image for Webcam

Figure 5. Noise reduced and filtered image for Webcam

3.6. Image Representation and Description

The centre and coordinate of the image features are considered to represent the pose of the PTR on the image plane coordinate. For the digital image, their 2-D centred moments are defined as follows [21]–[23]:

$$(I_{1ix}, I_{1iy}) = \sum_{(Ix,Iy)\in\Omega i} \sum(Ix, Iy) /N \quad (4)$$

$$(I_{1ox}, I_{1oy}) = \sum_{(Ix,Iy)\in\Omega o} \sum(Ix, Iy)/N \quad (5)$$

The centres of the whole image frame and image object i.e. (I_{1ix}, I_{1iy}) and (I_{1ox}, I_{1oy}) in fig 7 can be calculated by (4) and (5) respectively for the desired image features Ωi and Ωo. The error between centres of the whole image frame and image object in horizontal axis is determined as:

$$(I_{1ex}) = (\left| I_{1ix} - I_{1ox} \right|) \quad (6)$$

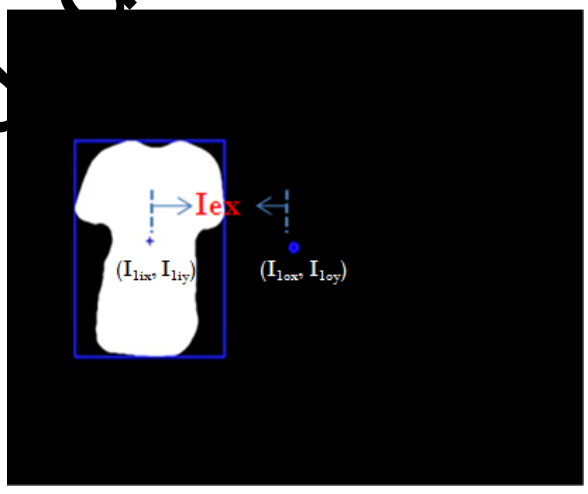

Figure 6. Image representing gripper and object centre coordinates for Webcam.

IV. NEURAL NETWORKS IN AREA CALCULATION

Multi-layer perceptrons are one of many different types of existing neural networks. They comprise a number of neurons connected together to form a network. The “strengths” or “weights” of the links between the neurons is where the functionality of the network resides. Its basic structure is shown in Fig. 7.

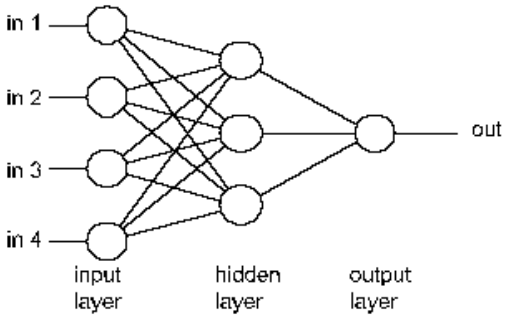

Figure 7. Structure of a multi-layer perceptron.

The idea behind neural networks stems from studies of the structure and function of the human brain. Neural networks are useful to model the behaviours of real-world phenomena. Being able to model the behaviours of certain phenomena, a neural network is able to subsequently classify the different aspects of those behaviours, recognise what is going on at the moment, diagnose whether this is correct or faulty, predict what it will do next, and if necessary respond to what it will do next.

Feed-forward networks [17] often have one or more hidden layers of sigmoid neurons followed by an output layer of linear neurons. Multiple layers of neurons with nonlinear transfer functions allow the network to learn nonlinear and linear relationships between input and output vectors. Table 1 shows the training and testing data collected in this project. The inputs of the network are the areas from the selected frames at different distances.

TABLE I:
TRAINING AND TESTING DATA

Sl.No	Area (in pixel square)	Distance (in metres)
1.	52198	2
2.	25162	3
3.	12365	4
4.	9682	5
5.	6728	6
6.	5053	7
7.	4123	8
8	2881	9
9.	1305	10
10.	789	11

One of the common problems when using Multilayer Perceptrons is how to choose the number of neurons in the hidden layer. There are many suggestions on how to choose the number of hidden neurons in Multilayer Perceptrons. For example, the minimum number of neurons, h, can be:

$$h \geq \frac{p-1}{n+2} \tag{5}$$

where p is the number of training examples and n is the number of inputs of the network [19]. The ANN model used to approximate the distance between the PTR and object comprises of one input and one output neuron along with four hidden neurons. The fitting function obtained after the training of the model in shown in Fig. 8.

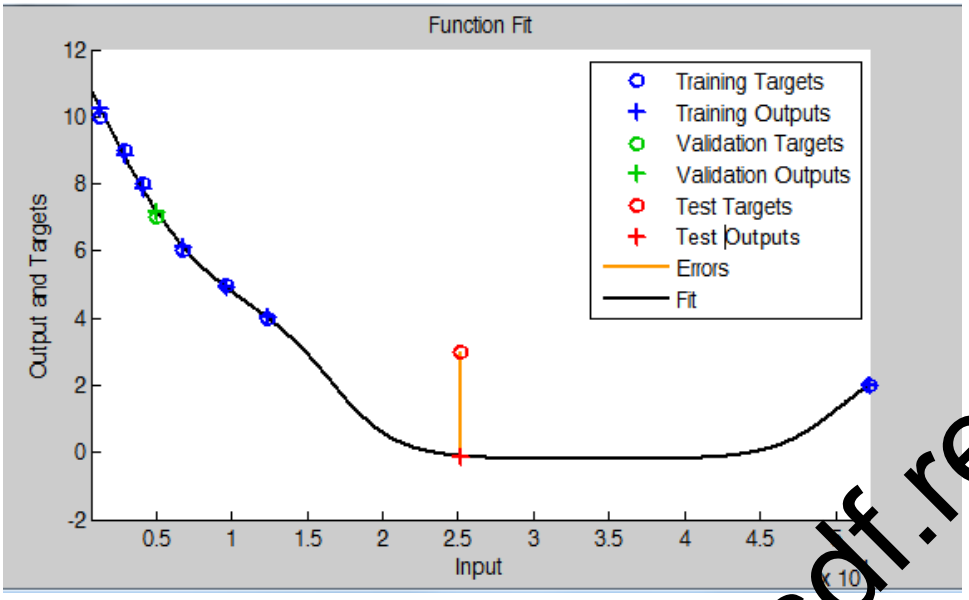

Figure 8. Fitting function for the trained ANN model.

V. EXPERIMENT RESULTS

The proposed method has been tested on sequences acquired in an indoor environment. Several experiments have been executed following a strategy that involves an increasing complexity for the tests. The robot is equipped with a vision sensor and an on board PC. The tests have been performed on a laptop equipped with an Intel processor and 1 GB of RAM. In this section, we verify the feasibility of the object tracking.

We designed a testing environment to verify the feasibility of object tracking as shown in Fig. 9. During the experiment, a target object starts moving from a starting point in the upper left corner. The target object moves according to a predefined route (dotted line in Fig. 9) with speed 20 cm/s and stops walking when it arrives at the starting point again. PTR starts to follow the target object until the distance between itself and the target object becomes longer than 1m. The traces of SSPR are represented by black circles in Fig. 10. We observed that PTR tried to follow the target object and kept the distance between itself and the target object within 1m. As long as the distance was less than 1m, SSPR stopped moving. Furthermore, SSPR kept correcting its direction according to the measured distance of ultrasonic sensors. As a result, PTR made the right turns at each corner. This test result verified the feasibility of object tracking by using vision sensor on a mobile robot.

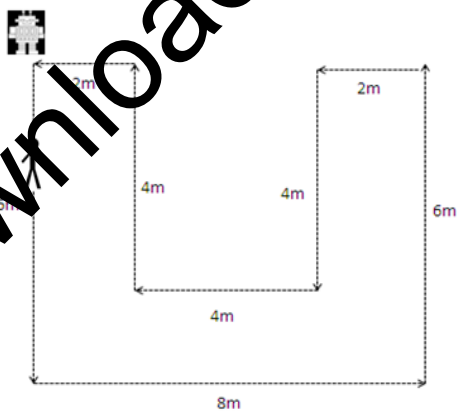

Figure 9. Testing environment for object tracking.

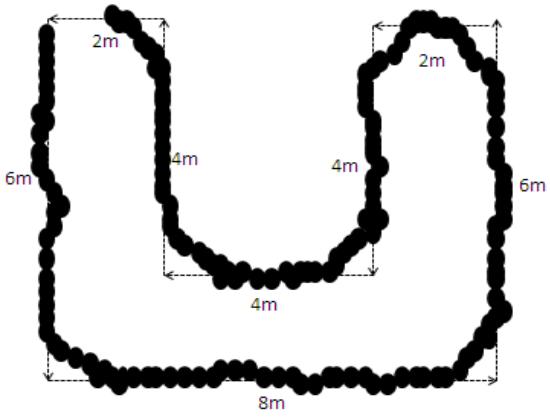

Figure 10. Experiment result of object tracking.

CONCLUSION

We have proposed a practical design of PTR for home environment. PTR can trace a moving object actively and establish audio/video streams between itself and the target. The PTR has been specifically designed to help a operator in monitoring wide indoor areas. For such purposes, it is able to move around a specific indoor environment (e.g., a building) and to track moving people. The selection of the target object to be tracked can be decided by the remote operator or autonomously by the PTR itself in the case that a suspicious behavior has been detected (e.g., a person entering a forbidden area, etc.).

Several experiments [3-5] on indoor sequences have demonstrated that the proposed PTR performs a robust detection of the motion inside the monitored scene. Then, to achieve a good identification of the mobile objects and to track them with enough accuracy, the ASV maintains the objects inside the field of view. Finally, the system shows a good object-detection rate for further person identification. When tracking a moving object, the robot may not keep the moving object in its viewing range due to a sudden change of motion direction. In this case, the robot follows a circular path in search of the moving object so that the mobile robot can detect the moving object as quickly as possible

Although the results appear promising, the constraints imposed still limit the exploitation of such a vehicle as a fully autonomous surveillance system. In particular, a single vision sensor cannot provide the complete scene of the area. Thus, the use of a multiple camera network could be useful to improve the understanding of the monitored scene. Moreover, there is a limit to the maximum speed of the vehicle. Indeed, when the speed remarkably increases, the parallax error cannot be addressed by the proposed method. This problem can be solved using additional information on the scene depth that could be acquired by range sensors mounted on board. Hence, future works will certainly give the robustness required to efficiently adopt the proposed PTR for the surveillance of indoor environments.

REFERENCES

[1]SeoKH, ShinJH,KimW,LeeJJ.Real-timeobjecttrackingandsegmentation using adaptive color snake model. International Journal of Control, Automation, and Systems 2006; 4(2):236–46.

[2]SongKT, ChenWJ, Face recognition and tracking for human-robot interaction. In: IEEE international conference on systems, man and cybernetics. The Hague, Netherlands: Vol.3.2004.p.2877–82.

[3] KwonH, YoonY, ParkJB, KakAC. Person tracking with a mobile robot using two uncalibrated independently moving cameras.In: IEEE international conference on robotics and automation. Barcelona, Spain:2005.p.2877–83.

[4] FodA, HowardA, aMatarićMJ. Laser based people tracking. In:proc of the IEEE international conference on robotics & automation (ICRA). Vol3. Washington, DC, UnitedStates: 2002.p.3024–9.

[5] MontemerloM, Thun S, WhittakerW. Conditional particle filters for simultaneous mobile robot localization and people-tracking. In: Proceedings of the IEEE international conference on robotics & automation (ICRA). Washington, DC, USA: Vol.1.2002.P.695–701.

[6] ScheutzM, McRavenJ, CsereyG.Fast, reliable, adaptive, bimodal people tracking for indoor environments. In: Proceedings of the 2004 IEEE/RSJ international conference on intelligent robots and systems (IROS'04). Vol.2. Sendai, Japan: 2004.p.1347–52.

[7] FritschJ, KleinehagenbrockM, LangS, FinkGA, SagererG. Audio visual person tracking with a mobile robot. In: GroenF, editor. Proceedings of the international conference on intelligent autonomous systems. Amsterdam: IOS Press; 2004.p.898–906.

[8] Chen CY. Obstacle avoidance design for a humanoid intelligent robot with ultrasonic sensors. Journal of Vibration and Control 2011;17(12):1798–804.

[9] Boykov Y, Huttenlocher D. Adaptive bayesian recognition in tracking rigid objects. In: Proceedings of IEEE conference on computer vision and pattern recognition. Vol. 2. Hilton Head, SC: 2000. p. 697–704.

[10] Isard M, Blake A. Visual tracking by stochastic propagation of conditional density. In: Proceedings of the fourth European conference computer vision. Cambridge, UK: 1996. p. 343–56.

[11] Merwe R, Doucet A, Freitas N, Wan E. The unscented particle filter, Technical Report CUED/F-INFENG/TR 380, Cambridge University Engineering Depart- ment, 2000.

[12] Song KT, Chang CC. Ultrasonic sensor data fusion for environment recogni- tion. In: Proceedings of international conference on intelligent robots and systems 1, 1993. p. 384–90.
[13] T. Kanade, R. Collins, A. Lipton, P. Burt, and L. Wixson, "Advances in cooperative multisensor video surveillance," in Proc. DARPA Image Understanding Workshop, Monterey, CA, Nov. 20–23, 1998, pp. 3–24.
[14] G. L. Foresti, "Object recognition and tracking for remote video surveillance," IEEE Trans. Circuits Syst. Video Technol., vol. 9, no. 7, pp. 1045–1062, Oct. 1999.
[15] Z. Zhu, G. Xu, B. Yang, D. Shi, and X. Lin, "VISATRAM: A realtime vision system for automatic traffic monitoring," Image Vis. Comput., vol. 18, no. 10, pp. 781–794, Jul. 2000.
[16] S. Dockstader and M. Tekalp, "Multiple camera tracking of interacting and occluded human motion," Proc. IEEE, vol. 89, no. 10, pp. 1441–1455,Oct. 2001.
[17] S. Haykin, Neural Networks, A Comprehensive Foundation. Prentice Hall, New Jersey, 1999.
[18] Zhen, B., Wu, X. and Chi, H., "On the Importance of Components of the MFCC in Speech and Speaker Recognition", Center for Information Science, Peking University, China, 2001.
[19] N. K. Kasabov, Foundations of Neural Network, Fuzzy Systems, and Knowledge Engineering. The MIT Press Cambridge, London, 1996.
[20] C. Solomon and T. Breckon.: Fundamentals of Digital Image Processing: A Practical Approach with Examples in Matlab.,New York, NY, USA: Wiley, 2010.
[21] E. D. Davies, Machine Vision—Theory, Algorithms, Practicalities, 3rd ed.Amsterdam, The Netherlands: Elsevier, 2005.
[22] 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.
[23] T. Komuro and M. Ishikawa, "A moment-based 3D object tracking algorithm for high-speed vision," in Proc. Int. Conf. Robot.