FACTA UNIVERSITATIS
Series: **Automatic Control and Robotics** Vol. 12, N° 1, 2013, pp. 43 - 51

KALMAN FILTER AND NARX NEURAL NETWORK FOR ROBOT VISION BASED HUMAN TRACKING*

UDC (004.421KALMAN), (004.032.26), (007.2)

Emina Petrović[1], Žarko Ćojbašić[1], Danijela Ristić-Durrant[2], Vlastimir Nikolić[1], Ivan Ćirić[1], Srđan Matić[1]

[1] University of Niš, Faculty of Mechanical Engineering,
Department of Mechatronics and Control, Niš, Republic of Serbia
[2]University of Bremen, Institute of Automation, Germany

Abstract. *Tracking human is an important and challenging problem in video-based intelligent robot systems. In this paper, a vision-based human tracking system is supposed to provide sensor input for vision-based control of a mobile robot that works in a team helping the human co-worker. A comparison between NARX neural network and Kalman filter in solving the prediction problem of human tracking in robot vision is presented. After collecting video data from a robot, simulation results obtained from the Kalman filter model are used to compare with the simulation results obtained from the NARX Neural network.*

Key words: *robot vision, Kalman filter, neural networks, human tracking*

1. INTRODUCTION

Human tracking and recognition of human actions in the real world environment is a very important perception task in robotics. Following a specific person is a significant task for service robots. A person-following robot finds a specified person using sensory input and follows the person in order to provide various services to humans or to archive different tasks depending on the application [1]. Neural networks have been widely used for robotic control and human tracking because of their ability to approximate arbitrary linear or non-linear systems in a compact set. The control methods based on neural networks were demonstrated through many theoretical and industrial solutions [2].

Received March 10, 2013
Corresponding author: Emina Petrović
Department of Mechatronics and Control, University of Niš, Faculty of Mechanical Engineering, Niš, Republic of Serbia • Email: emina@masfak.ni.ac.rs
* **Acknowledgements.** This research in this paper was supported by the Ministry of Education and Science of Republic of Serbia and DAAD through the project "A Novel Approach for Human Detection and Tracking in Robotics". The research was supported by the project "Research and Development of New Generation of Highly Energy Efficient Wind Turbines" (No. 35005/2011) funded by the Ministry of Education and Science of Republic of Serbia.

Many researchers have applied artificial neural networks into effective human tracking systems and navigational systems. An efficient human action recognition using NN is presented in [2]. Convoluted NN is used for human action recognition in videos. The video frames are treated and this image is used by NN to recognize actions in the individual frame. Also, another efficient tracking system using an algorithm for tracking a human using NN is developed in [3]. The multi layer feed forward NN is used to estimate the distance of the person from the robot's camera, and that estimation is combined and used from Kalman filter to track the position of the person over time. Tracking humans by mobile robots using the Kalman filter is presented in many papers, and different solutions for tracking with robots are reported. Some applications use robot's range sensors, considering people as moving entities [4, 5]. On the other hand, some approaches use robot's onboard camera to detect people [6, 7]. Others make use of multi-sensor systems in which visual data is combined with other robot's range sensors [8, 9]. Existing approaches use color tracking, a mixture of features, or stereo vision to identify single persons and eventually their gestures or shape contour and different filtering techniques to estimate the target states. Mobile robots using 2D or 3D data by any classical Kalman filter can track people. The system implemented in [10] adopts Kalman filters to track people using cameras mounted on mobile robots. The techniques described in [11] are designed to improve tracking performance using Kalman filter and Optical flow. The approach presented in [12] uses a stereo-based camera to detect and track persons using Extended Kalman filter. The system presented in [13] applies Kalman filter method to synchronize the vision based data and the motion control of the robot to prevent losing the tracked target due to robot motion. The approach described in [14] presents the self-tuning Kalman filter technique using echo state networks for mobile robots.

In this paper the vision-based human tracking system should provide sensor input for vision-based control of a mobile robot which operates in a team helping the human coworker. The robot vision system for human tracking for the considered robot working scenarios has to be able to detect the human in camera images. Also it should calculate the 3D coordinates of human's center of mass in camera coordinate system (x, y, z) and to calculate the distance to the human and to track the human keeping a constant distance between the robot and human [15]. Experiments were conducted where a human walking towards the robot was captured by a Point Grey Bumblebee stereo camera. The frame rate of the used camera at full resolution is 15 fps (frames per second). Each pair of stereo frames was processed so to extract information for stereo-vision based reconstruction of human walking with respect to camera coordinate system. This means that the distance between the human and the camera coordinate system, attached to the left stereo camera, was calculated for every moment of capturing the stereo images according to:

$$D = \sqrt{x^2 + y^2 + z^2} \tag{1}$$

where *x, y* and *z* are the 3D coordinates of the human's center of mass that are calculated using a stereo vision based method for 3D robot vision reconstruction of human walking that is described in details in [16].

The collected vision data such as the distance between the robot and the human in time were used for getting the simulation results of KF and NN. The state of the person is presented in Cartesian coordinates *x, y, z* and three-dimensional velocity. Assuming that

the velocity is constant, the normalized state transition matrix can be obtained from basic kinematic equations as:

$$\begin{aligned} S_k &= S_{k-1} + V_{k-1} * \Delta t \\ V_k &= V_{k-1} \end{aligned} \tag{2}$$

where $\mathbf{S}_k$ is position of a human, $\mathbf{V}_k$ is velocity, and Δt is a sampling interval.

The comparison between NARX neural network and Kalman filter in solving the prediction problem of human tracking in robot vision is presented. After collecting video data from a robot, simulation results obtained from the Kalman filter are used to compare with the simulation results obtained from the NARX Neural network. The presented methods are assumed to provide a reliable input for the control of a mobile robot following human co-worker.

2. KALMAN FILTER

In the Kalman Filter approach [17], it is presumed that the following predefined models of motion and measurement would characterize the behavior of a moving object

$$\mathbf{x}_k = \mathbf{A}\mathbf{x}_{k-1} + \mathbf{w}_k \tag{3}$$

$$\mathbf{z}_k = \mathbf{H}\mathbf{x}_k + \mathbf{v}_k \tag{4}$$

and the models can be represented in terms of a state vector $\boldsymbol{x}_k$ that corresponds to image frame k. In (3) **A** represents the state transition matrix which determines the relationship between the present state $\boldsymbol{x}_k$ and the previous one $\boldsymbol{x}_{k-1}$ and the matrix **H** describes the relationship between the measurement vector $\boldsymbol{z}_k$ and the state vector $\boldsymbol{x}_k$. The vectors $\boldsymbol{w}_k$ and $\boldsymbol{v}_k$ are noise terms which are assumed to be independent of each other, Gaussians with zero mean and covariance matrices $\mathbf{Q} = E[\mathbf{w}_k \mathbf{w}_k^{\mathrm{T}}]$ and $\mathbf{R} = E[\mathbf{v}_k \mathbf{v}_k^{\mathrm{T}}]$.

We assume that the velocity of the tracked human is constant between the subsequent video frames so the state vector is simplified and does not include the acceleration term. This three dimensional problem can be expressed in matrix form. Kinematical equation is rewritten as matrix:

$$\begin{bmatrix} S_{x,k} \\ S_{y,k} \\ S_{z,k} \\ V_{x,k} \\ V_{y,k} \\ V_{z,k} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & \Delta t & 0 & 0 \\ 0 & 1 & 0 & 0 & \Delta t & 0 \\ 0 & 0 & 1 & 0 & 0 & \Delta t \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} S_{x,k-1} \\ S_{y,k-1} \\ S_{z,k-1} \\ V_{x,k-1} \\ V_{y,k-1} \\ V_{z,k-1} \end{bmatrix} \tag{5}$$

where the vector $\mathbf{S}_k$ represents measurements on 3D position of the detected human in frame k and the vector $\mathbf{V}_k$ represent 3D velocities of the center of mass of the segmented human region. The relationship between the state vector $\mathbf{x}_k$ and measurement vector $\mathbf{z}_k$ is defined by matrix **H**. Given the fact that the state vector is of length six and the measurement vector is of length three, the matrix **H** must be of length six by three:

$$\mathbf{H} = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{bmatrix} \tag{6}$$

The complete output of human detection is the three-dimensional vector:

$$\mathbf{S}_k = \begin{bmatrix} \mathbf{S}_x & \mathbf{S}_y & \mathbf{S}_z \end{bmatrix}^T \tag{7}$$

The presented Kalman filter based tracking has two stages: the prediction and the correction stage. In the prediction stage, the state vector $\boldsymbol{x}_k$ in the current frame is predicted using the corresponding state vector in the previous frame via matrix **A**. In the correction stage, the measurement vector in the current frame is used to update the state estimate and the error covariance matrix.

The prediction stage can be expressed as follows:

$$\hat{\mathbf{x}}_k^- = \mathbf{A}\hat{\mathbf{x}}_{k-1} \tag{8}$$

$$\mathbf{P}_k^- = \mathbf{A}\mathbf{P}_{k-1}\mathbf{A}^T + \mathbf{Q} \tag{9}$$

where $\hat{\mathbf{x}}_k^-$ is the a priori estimate state and $\mathbf{P}_k^-$ is the a priori estimate error covariance matrix at frame *k*. The correction stage can be expressed through the following three equations:

$$\mathbf{K}_k = \mathbf{P}_k^-\mathbf{H}^T(\mathbf{H}\mathbf{P}_k^-\mathbf{H}^T + \mathbf{R})^{-1} \tag{10}$$

$$\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_k^- + \mathbf{K}_k(\mathbf{z}_k - \mathbf{H}\hat{\mathbf{x}}_k^-) \tag{11}$$

$$\mathbf{P}_k = (I - \mathbf{K}_k\mathbf{H})\mathbf{P}_k^- \tag{12}$$

where $\hat{\mathbf{x}}_k$ is the a posteriori estimate state at frame k, given the measurements up to that time, $\mathbf{P}_k$ is the a posteriori estimate error covariance matrix at frame *k* indicating the accuracy of the estimated state vector, and $\boldsymbol{K}_k$ is the Kalman gain. The tracking process is recursive. The estimates computed from (11) are used as the state vectors on the right hand side of (8) for the next video frame k+1. This is exactly the predicted vector $\boldsymbol{S}_k$.

3. THE NONLINEAR AUTOREGRESSIVE NETWORK WITH EXOGENOUS INPUTS (NARX)

The nonlinear autoregressive network with exogenous inputs (NARX) is a recurrent dynamic network, with feedback connections enclosing several layers of the network. The NARX model is based on the linear ARX model, which is commonly used in time-series modeling. One can implement the NARX model by using a feed forward neural network to approximate the function *f*. The defining equation for the NARX model is [18,19]:

$$\boldsymbol{y}(t)=f(\boldsymbol{y}(t\text{-}1), \boldsymbol{y}(t\text{-}2),\ldots,\boldsymbol{y}(t\text{-}n_y),\boldsymbol{u}(t),\boldsymbol{u}(t\text{-}1),\ldots,\boldsymbol{u}(t\text{-}n_u)) \tag{13}$$

where $\boldsymbol{u}(t)$ and $\boldsymbol{y}(t)$represent the input and output of the network at time *t*, and the function *f* is some nonlinear function, in this case a neural network. The next value of the dependent output signal $\boldsymbol{y}(t)$ is regressed on previous values of the output signal and previous

values of an independent (exogenous) input signal. When the function can be approximated by a Multilayer Perceptron, the resulting system is called a NARX network which is a recurrent neural network [18, 19].

There are many applications of the NARX network. It can be used as a predictor, to predict the next value of the input signal. It can also be used for nonlinear filtering, in which the target output is a noise-free version of the input signal.

The output of the NARX network can be an estimate of the output of some nonlinear dynamic system that is modeled. The output is feedback to the input of the feed forward neural network as part of the standard NARX architecture. Because the true output is available during the training of the network, it is possible to create a series-parallel architecture, in which the true output is used instead of feeding back the estimated output. This has two advantages, the first is that the input to the feed forward network is more accurate and the second is that the resulting network has a purely feed forward architecture, and static back propagation can be used for training.

To forecast the trajectory of *x,y,z* coordinates in real time for every predicted state is estimated based on the current inputs in feed forward neural networks, and the estimation in NARX network is done based on the current inputs, previous inputs and previous states. In order for the parallel response to be accurate, it is important that the NARX network is properly trained so that the errors in the series-parallel configuration are very small. Otherwise, the accumulation of error might occur during estimation.

The purpose of this simulation is to forecast the trajectory of *x,y,z* coordinates in real time. For the simulations we used feed forward topology of the NARX NN. The NN had ten hidden neurons and two delays and is presented in Fig. 1.

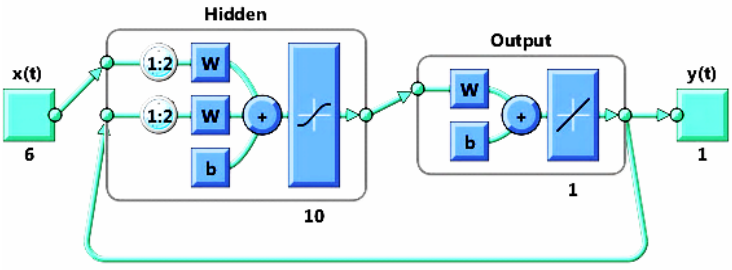

Fig. 1 The used NARX NN topology with weights and biases at each node

The training was done by Levenberg-Marquardt back propagation [18] and regression value was R=0.9685. The network also had additional nodes, so more weighting factors needed to be adjusted and the training of NARX network needed more computational time than it would be needed if conventional neural network was used.

4. Simulation Results

In this paper we present a comparison between NARX neural network and Kalman filter in solving the prediction problem of human tracking in robot vision. After collecting video data from a robot, simulation results obtained from the Kalman filter model are

used to compare with the simulation results obtained from the NARX Neural network. Kalman filter results are obtained from a mobile robot platform equipped with vision sensors that can be used as human co-warker in the examination of environment. The robot is equipped with sensors for environment perception and with sensors for platform navigation as well [15]. One of the robot's working scenarios concerns the vision based human tracking as the robot works as a transportation robot, helping the human to carry containers with collected samples from the investigated environment. Experiments were conducted with a human walking towards the robot. Robot's vision sensor is a Point Grey Bumblebee stereo camera that provides information for human tracking. The camera is mounted on a rack located on the back side of the robot mobile platform.

Figures 2, 3 and 4 presented predicted and measured *x, y* and *z* coordinates of the tracked human with Kalman filter and NARX NN respectively. Some examples of processed images by NARX NN and Kalman filter with superimposed extracted bounding boxes of human are shown in Figure 5. After reviewing these results from the figures below, we can see that both approaches made prediction accurately, and to some extent almost the same. In addition, both approaches were able to reduce noise in measurements, and to predict state when the robot's vision system did not give measurements in some intervals of time. However, in case of Kalman filter we must provide its dynamic model before tracking, because if the provided model is not the actual model the tracking can fail easily, what is the main disadvantage of Kalman filter.

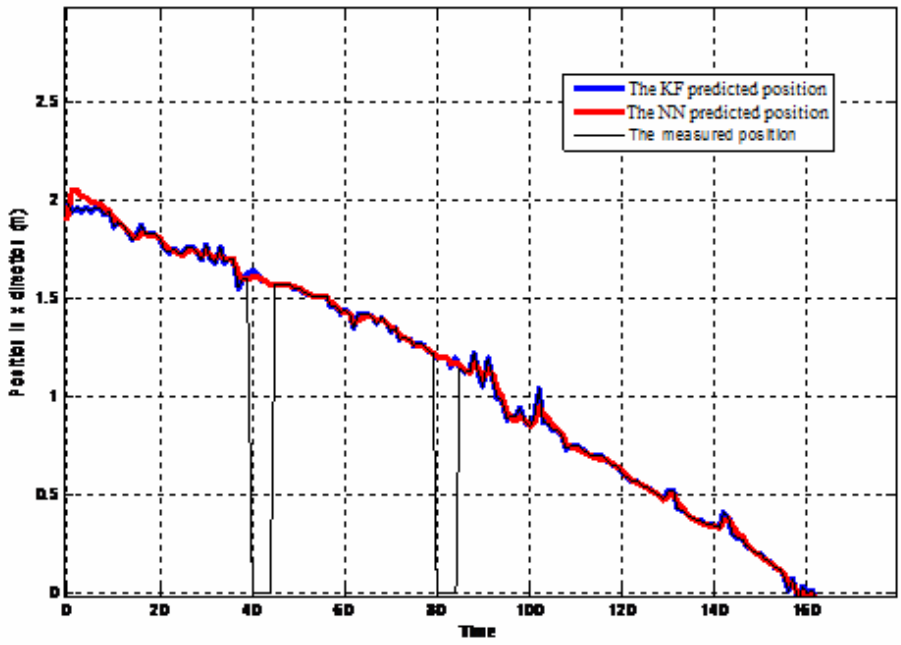

Fig. 2 The predicted and the measured *x* coordinate of the tracked human

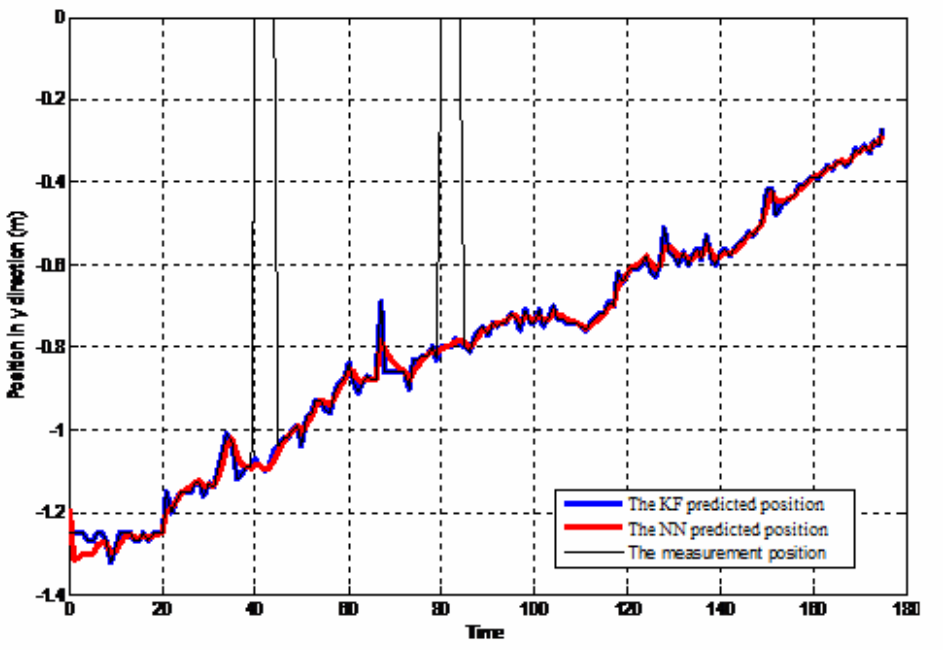

Fig. 3 The predicted and the measured *y* coordinate of the tracked human

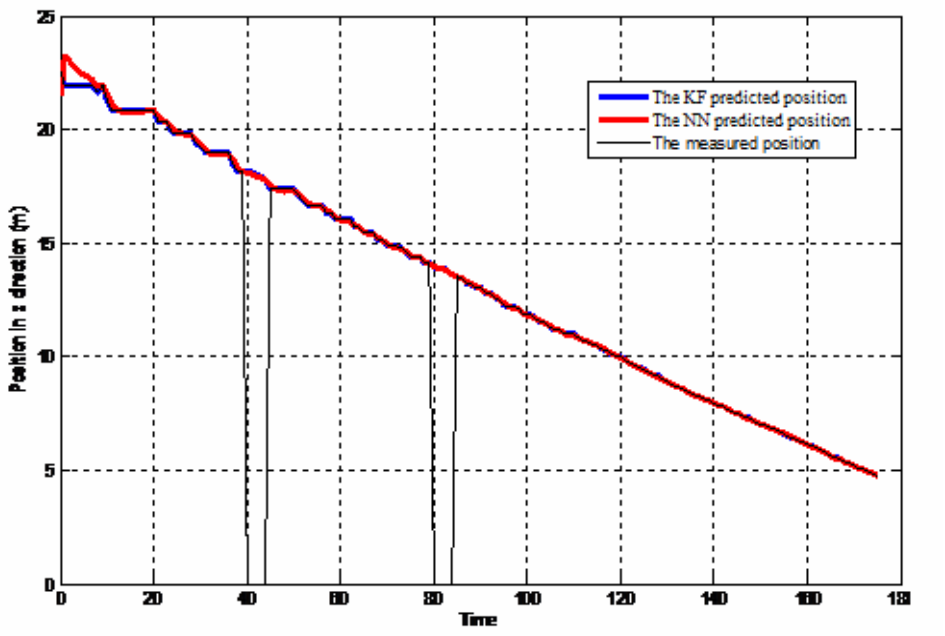

Fig. 4 The predicted and the measured *z* coordinate of the tracked human

Fig. 5 KF predicted position (blue) and NN predicted position (red)

5. CONCLUSION

In this work one solution for human tracking using NN is presented. First, we described the NARX neural network and its advantages in time series prediction, after that we presented the Kalman filter and its contribution in solving the human tracking problem. Finally, we compared the results of human tracking obtained from the NN with the results obtained from the Kalman filter through simulation. Also, we have demonstrated the performance of tracking system using the results obtained from the NARX Neural network which are based on the video sequences and compared with the results obtained from the Kalman filter. Experimental results show us that NARX Neural networks give good results telling us that NN made predictions accurately and reliably as well as the Kalman filter. On the other hand, NN can work directly with raw inputs and data. Using that fact and characteristic of the NN that they are able to learn the dynamic model in real time, NN can be easily implemented in human tracking algorithm in order to have more robust and reliable tracking result.

REFERENCES

[1] J. Borenstein, D. Thomas, B. Sights, L. Ojedo, P. BAnkole, D. Fellars, "Human leader and robot follower team: correcting leader's position from follower's heading," in *Proc. of the SPIE Defense, Security and Sensing*, Orlando, Florida, 2010.

[2] S. Ji, W. Xu, M. Yang, K. Yu, "3D convolutional neural networks for human action recognition," in *Proc. of the 27th International Conference on Machine Learning*, Haifa, Israel, 2010.

[3] G. Cielinak, M. Miladinovic, D. Hammarin, L. Goranson, A. Lilienthal, T. Duckett, "Appearance based tracking of persons with an omnidirectional vision sensor," in *Proc. of Computer Vision and Pattern Recognition Workshop*, 2003. [Online]. Available: http://dx.doi.org/10.1109/CVPRW.2003.10072

[4] J. Bobruk, D. Austin, "Laser motion detection and hypothesis tracking from a mobile platform," in *Proc. of the Australian Conference on Robotics & Automation*, pp. 1–10, 2004.

[5] P. Urcola, L. Montano, "Adapting robot team behavior from interaction with a group of people," in *Proc. Intelligent Robots and Systems*, pp. 2887–2894, 2011. [Online]. Available: http://dx.doi.org/10.1109/IROS.2011.6094961

[6] D. Beymer, K. Konolige, "Tracking people from a mobile platform," Experimental Robotics VIII, pp. 234–244, 2003. [Online]. Available: http://dx.doi.org/10.1007/3-540-36268-1_20

[7] M. Burke, W. Brink, "Gain-scheduling control of a monocular vision-based human-following robot," in *Proc. of 18th World Congress of the International Federation of Automatic Control*, Milano, Italy, 2011, pp. 8177–8182.

[8] N. Bellotto, H. Hu, "A bank of unscented Kalman filters for multimodal human perception with mobile service robots," International Journal of Social Robotics, vol. 2, no. 2, pp. 121–136, 2010. [Online]. Available: http://dx.doi.org/10.1007/s12369-010-0047-x

[9] J. A. Corrales, F. A. Candelas, F. Torres, "Kalman filtering for sensor fusion in a human tracking system," Intech, pp. 59–72, 2010.

[10] N. Bellotto, H. Hu, "Multisensor-based human detection and tracking for mobile service robots," IEEE Transactions on systems, Man, and Cybernetics-part B: Cybernetics, vol. 39, no. 1, pp. 167–181, 2009. [Online]. Available: http://dx.doi.org/10.1109/TSMCB.2008.2004050

[11] R. E. Kalman, "A new approach to linear filtering and prediction problems," Journal of Basic Engineering, vol. 82, no. 1, pp. 35–45, 1960.

[12] J. Satake, J. Miura, "Robust stereo-based person detection and tracking for a person following robot," in *Proc. of the Workshop on People Detection and Tracking*, Kobe, Japan, May 2009.

[13] Y. Zheng, Y Meng, "Real-time people tracking and following using a vision-controlled mobile robot," Robot Vision: New Research, 2009.

[14] C.Y. Tsai, X. Dutoit, K.T. Song, H.V. Brussel, M. Nuttin, "Robust face tracking control of a mobile robot using self-tuning Kalman filter and echo state network," Asian Journal of Control, vol. 12, no. 4, pp 488–509, 2010. [Online]. Available: http://dx.doi.org/10.1002/asjc.204

[15] E. Petrović, D. Ristić-Durrant, A. Leu, V. Nikolić, "A novel approach to human tracking for robotic Follower," in *Proc. XI International SAUM Conference on Systems, Automatic Control and Measurements*, Niš, Serbia, 2012, pp.178–181.

[16] Leu, D. Aiteanu, A. Gräser, "A novel stereo camera based collision warning system for automotive applications," in *Proc of 6th IEEE International Symposium on applied Computational Intelligence and Informatics*, *SACI 2011*, Timisoara, Romania, 2011, pp. 409–414, 2011. [Online]. Available: http://dx.doi.org/10.1109/SACI.2011.5873038

[17] S. Chen, "Kalman filter for robot vision: a survey," IEEE Transactions on Industrial Electronics, vol. 59, no. 11, pp. 4409–4420, 2012. [Online]. Available: http://dx.doi.org/10.1109/TIE.2011.2162714

[18] Pelliccioni, T. Tirabassi, "Air dispersion model and neural network: a new perspective for integrated models in the simulation of complex situations," Environmental Modeling & Software, vol. 21, no. 4, pp. 539–546, 2006. [Online]. Available: http://dx.doi.org/10.1016/j.envsoft.2004.07.015

[19] I. Ćirić, Ž. Ćojbašić, V. Nikolić, P. Živković, M. Tomić, "Air quality estimation by computational intelligence methodologies," Thermal Science, vol. 16, iss. 2, pp. 493–504, 2012. [Online]. Available: http://dx.doi.org/10.2298/TSCI120503186C