ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL

-

REAL TIME PEDESTRIAN TRACKING USING ADAPTIVE KALMAN FILTER

M.Sc. THESIS

Coşkun Orkun Vural

Department of Control and Automation Engineering

Mechatronics Engineering Programme

February 2022

ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL

REAL TIME PEDESTRIAN TRACKING USING ADAPTIVE KALMAN FILTER

M.Sc. THESIS

Coşkun Orkun Vural
518181042

Department of Control and Automation Engineering

Mechatronics Engineering Programme

Thesis Advisor: Prof. Dr. Mustafa DOĞAN

February 2022

ISTANBUL TEKNİK ÜNİVERSİTESİ ★ LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

ADAPTİF KALMAN FİLTRESİ KULLANIMI İLE GERÇEK ZAMANLI YAYA TAKİBİ

YÜKSEK LİSANS TEZİ

Coşkun Orkun Vural
518181042

Kontrol ve Otomasyon Mühendisliği Anabilim Dalı

Mekatronik Mühendisliği Programı

Tez Danışmanı: Prof. Dr. Mustafa Doğan

Şubat 2022

Coşkun Orkun Vural, a M.Sc. student of İTU Graduate School student ID 518181042, successfully defended the thesis entitled "REAL TIME PEDESTRIAN TRACKING USING ADAPTIVE KALMAN FILTER", which he prepared after fulfilling the requirements specified in the associated legislations, before the jury whose signatures are below.

Thesis Advisor :	**Prof. Dr. Mustafa DOĞAN** İstanbul Technical University	.....
Jury Members :	**Prof. Dr. Müjde GÜZELKAYA** Istanbul Technical University	.....
	Assoc. Prof. Dr. Janset DAŞDEMİR Yıldız Technical University	.....

Date of Submission : 14 January 2022
Date of Defense : 01 February 2022

To all the special ones who were there when I needed most,

FOREWORD

Accurate and fast detection of objects in real time is a fundamental research subject. The computing simplicity and adaptation to changing environmental variables are crucial for the tasks in real time. One of the applications of real time object detection is pedestrian tracking in moving cars.

In this work, a fast and robust system for pedestrian tracking using an adaptive kalman filter is developed. The algorithm predicts the next state of the pedestrian and aims to warn the driver beforehand.

I would like to thank my advisor Prof. Dr. Mustafa Doğan for always guiding me when I need guidance, being always the most understanding and best mentor possible. If it wasn't him who guided me throughout this process it was not possible to conclude this work. I also would like to thank my wonderful family who were always there for me since the day 1. They are always my inspiration and role models. I also would like to thank my dear girlfriend Burcu Akad who was there in my best and worst. Last but not least, I would like to thank my friends Emrah Kumru and Berk Başaran for their help during the experimentation phase of the project.

February 2022

Coşkun Orkun VURAL
(Electrical and Electronics Engineer)

TABLE OF CONTENTS

Page

ABBREVIATIONS

IDE	**:** Integrated development environment
AKF	**:** Adaptive Kalman Filter
KF	**:** Kalman Filter
IR	**:** Infrared
IAE	**:** Innovation based adaptive estimation
MMAE	**:** Multiple model adaptive estimation

SYMBOLS

P **:** Initial state estimation error

Q **:** Process noise

R **:** Measurement noise

K **:** Kalman Gain

LIST OF TABLES

Page

LIST OF FIGURES

Page

REAL TIME PEDESTRIAN TRACKING USING ADAPTIVE KALMAN FILTER

SUMMARY

Image processing has been one of the hot topics since early 1950's and it has developed from simple researches on the images to real time video processing. This development brought more challenges to the researches as the environment around became a complex variable to process. Thus new techniques with low computing time and flexible algorithms that adapt to changes in the environment emerged. Human detection and estimation of their movements is one of the research areas described before. For security and emergency situations, applications of real time human detection is fundamental.

In this work it was aimed to develop a robust system that can predict human movement on real time video. The motivation behind the work is to design a system that can detect pedestrians and their movement to warn the driver beforehand. The system needed to do the lowest amount of computation time as possible and it needed to adapt to changes of the environment in order to work in real time with unstable surroundings. Thus an Adaptive Kalman algorithm is developed for prediction of the next steps as only the previous steps information was needed.

For the development of the project Matlab is used as the programming language. The reason behind choosing Matlab is due to the fact that Matlab IDE includes many libraries and toolboxes useful for this work. As the first step of the project the camera input is evaluated in real time by taking snapshots and further algorithms are applied on the frames taken. For human detection the built in Matlab library which gets data from the Caltech Dataset is used. The data gathered from the human detection algorithm is used to find the centroids in the region of interest. Centroids are fed to the Adaptive Kalman Filter as the input data with corresponding error parameters. The term Adaptive Kalman Filter comes from the responsive error tuning from the previous input. P, Q and R parameters are tuned with the error values of the previous frame. It was observed that all P, Q and R parameters and K value is converged with

the increasing number of frames. Thus the learning process is also included in the work.

In the 1000 test frames the system is tested with single and multiple pedestrians in the frame. Out of total 1678 samples 1418 of them are the right classifications which results in a %84.5 success rate. In the test process, it was observed that the error is mainly caused when two objects move towards each other. From that observation it can be concluded that the system works better in areas where the crowd is less dense.

ADAPTİF KALMAN FİLTRESİ KULLANIMI İLE GERÇEK ZAMANLI YAYA TAKİBİ

ÖZET

Görüntü işleme, 1950'lerin başından beri önde gelen araştırma konulardan biri olmuştur ve fotoğraf üzerinde yapılan basit araştırmalardan gerçek zamanlı video işlemeye kadar gelişmiştir. Çevredeki ortamın işlenmesi karmaşık bir değişken haline geldiğinden, bu gelişme araştırmacıların önüne aşılması gereken daha çok engel getirdi. Böylece ortamdaki değişikliklere uyum sağlayan düşük hesaplama süresine ve esnek algoritmalara sahip yeni teknikler ortaya çıktı. İnsan hareketinin tespiti ve tahmini, daha önce öne çıkan araştırma alanlarındandır. Güvenlik ve acil durumlar için yapılan çalışmalarda gerçek zamanlı insan tespiti uygulamaları esastır.

Bu çalışmada gerçek zamanlı video üzerinde insan hareketini tahmin edebilen tutarlı bir sistem geliştirilmesi amaçlanmıştır. Çalışmanın arkasındaki motivasyon, sürücüyü önceden uyarmak için yayaları ve hareketlerini algılayabilen bir sistem tasarlamaktır. Sistemin mümkün olan en düşük hesaplama süresinde çalışması ve gerçek zamanlı olarak sabit olmayan ortamlarla çalışabilmesi için ortamdaki değişikliklere uyum sağlaması gerekiyordu. Bunu göz önünde bulundurarak sadece önceki adımlar bilgisine ihtiyaç duyulduğu ve datanın tamamını işleme gerekliliği bulunmaması nedeniyle sonraki adımların tahmini için Adaptif Kalman Filtresi kullanılmıştır.

Projenin geliştirilmesi için programlama dili olarak Matlab tercih edilmiştir. Matlab'ın seçilmesinin nedeni, Matlab IDE'sinin bu iş için faydalı birçok kütüphane ve araç kutusu içermesidir. Projenin ilk adımı olarak, video anlık görüntü alınarak gerçek zamanlı olarak değerlendirildi ve alınan kareler üzerinde ilerleyen bölümlerde açıklanan algoritmalar uygulandı. İnsan tespiti için Caltech Veri Kümesinden veri alan yerleşik Matlab kütüphanesi kullanıldı. İnsan algılama algoritmasından toplanan veriler, ilgilenilen bölgedeki merkezlerin bulunması için kullanıldı. Merkez noktalar, karşılık gelen hata parametreleriyle giriş verileri olarak Adaptif Kalman Filtresine beslendi. Adaptif Kalman Filtresi terimi, önceki giriş parametrelerinden gelen değişken hata ayarından gelir. P, Q ve R parametreleri bir önceki karedeki hata değerleri ile ayarlandı. Tüm P, Q ve R parametrelerinin ve K değerinin artan işlenen kare sayısı arttıkça sabit bir değere yakınsadığı gözlendi. Böylece öğrenme süreci de bu çalışmanın içinde kullanılmış oldu. Her bir kare üzerinde işlem yaparken önceki karede yapılan tahmin ile o anki karedeki gerçekte oluşan merkez noktaları kıyaslanıp buna göre ilgili hata parametreleri güncellendi. Bu durumda hareket veya ölçüm yapılan cihazdan kaynaklı olan hataların minimuma indirgenmesi hedeflenmiştir. Bir sonraki kareye geçilirken ise ilk olarak insan algılama

algoritmasındaki algılanan insan sayısı hesaplanıp eğer kareden çıkan bir kişi var ise bu kişinin kaydı silinip gereksiz işlem yükünü azaltmak istendi. Ayrıca bulunan her veri için ayrı filtre kullanılması değişken boyutlarda ve hareket yönlerine sahip yayalar için ortak tahminler yapılıp hatalı bir tahmin yapılmasını engellemiştir.

Yapılan denemeler sonucu gerçek zamanlı videoda önerilen sistemin olumlu sonuçlar verdiği gözlemlenmiştir. 1000 test görüntüsünde, sistem görüntüdeki tekli ve çoklu yayalarla test edildi. Toplam 1678 örnekten 1418'i doğru sınıflandırma olup %84.5 başarı oranı ile sonuçlanmıştır. Bu gözlemden, kalabalığın daha az yoğun olduğu alanlarda sistemin daha iyi çalıştığı sonucuna varılabilir. Ayrıca hata değerlerinin sabit bit değere yakınsaması ve hata oranının kare sayısı arttıkça azalması da sistemin yaptığı hatalardan öğrenerek daha stabil bir hale geldiğinin göstergesidir.

Bu çalışmada, Adaptif Kalman Filtresi kullanılarak görüntüdeki yayaları başarılı bir şekilde algılayabilen ve hareketlerini tahmin edebilen sağlam bir sistem tasarlamak amaçlanmıştır. 1000 test görüntüsünde sistem, çerçevedeki tekli ve çoklu yayalarla test edildi. Toplam 1678 örnekten 1418'i doğru sınıflandırma olup %84.5 başarı oranı ile sonuçlanmıştır. Test sürecinde, hatanın esas olarak iki nesnenin birbirine doğru hareket etmesinden kaynaklandığı gözlemlendi. Bu gözlemden, kalabalığın daha az yoğun olduğu alanlarda sistemin daha iyi çalıştığı sonucuna varılmıştır.

Sistemin performansını karşılaştırmak için iki öğrenme yöntemi daha (Bayesian Regularization ve Levenberg-Marquardt) uygulandı. Doğruluk ve kesinlik değerleri karşılaştırıldığında Adaptif Kalman Filtresinin diğer yöntemlere göre daha iyi sonuçlar verdiği görülmüştür. Ayrıca Adaptif Kalman Filtresi herhangi bir eğitim aşaması gerektirmediği için gerçek zamanlı uygulamalar için daha uygundur.

Kalman kazancı ve hata parametreleri gözlendiğinde, bu parametrelerin bir aralığa yakınsadığı, yani filtrenin kendisini ortama adapte ettiği, durdurma kriterlerinin başarıyla sağlandığı görülmektedir. Adaptif Kalman Filtresi yerine Kalman Filtresi kullanılmış olsaydı, ortam değişiklikleri nedeniyle sistemin doğru şekilde ayarlanması mümkün olmaz ve bu da tahminlerde tutarsızlıklara neden olurdu.

Bu çalışma ile bir dereceye kadar tek bir kamera kullanılarak yaya algılama ve hareket tahmininin yapılabileceği gösterilmiştir. İzlenen yöntemi kalabalıkları birbirinden ayırabilen bir algoritma ile birleştirmek, sistem performansını önemli ölçüde artıracaktır.

Bu çalışma için, Adaptif Kalman Filtresi kullanılarak, tek bir kamera ile gerçek zamanlı video üzerinde insan hareketi tahmini yapılabileceği varsayılmıştır. Kurulan sistemin %84.5 başarı oranı ile hipotezin doğrulandığı görülmüştür. Ayrıca gerçek zamanlı videonun değişen çevre koşullarının zorluklarını ortadan kaldırmak için Adaptif Kalman Filtresinin tercih edilmiştir. Kalman kazancı ve hata parametrelerinin sistemin öğrenme sürecini kanıtlamak için gerekli olan küçük bir değer aralığına yakınsadığını gözlemlemek de önemli bir faktör olup, sonuçlarda gözlemlenmesi sistemin gerçek zamanda çalışabileceğini kanıtlamıştır. Sonuç olarak, bu tez çalışması, tek bir kamera ile gerçek zamanlı video üzerinde insan hareketi tahmininin yapılabileceğini göstermiştir.

Başarı oranı, bunun gerçek zamanlı yaya takibinde uygulanacak etkili bir yöntem olduğunu kanıtlayacak kadar yüksek olmasına rağmen, sistem için bazı geliştirme alanları bulunmaktadır. Birbirine doğru hareket eden birden fazla yayada sistem performansı önemli ölçüde azaldığı gözlemlenmiştir. Daha iyi sonuçlar için nesneler için daha iyi bir ayırma yöntemi geliştirilmelidir. İkinci olarak sistem, yayanın tüm vücudu çerçevede bulunduğunda çalışır. Kameraya çok yakın olan yayaları sınıflandırmak için bir gövde algılama algoritması uygulanabilir. Son olarak, daha iyi bir işlemci ve kamera ile test etmek, saniyedeki kare sayısını artırabilir ve bu da daha hassas bir sınıflandırma ile sonuçlanacaktır.

1. INTRODUCTION

Image processing has been one of the hot topics since early 1950's and it has developed from simple researches on the images to real time video processing. This development brought more challenges to the researches as the environment around became a complex variable to process. Thus new techniques with low computing time and flexible algorithms that adapt to changes in the environment emerged. Human detection and estimation of their movements is one of the research areas described before. For security and emergency situations applications of real time human detection is fundamental.

In this work it was aimed to develop a robust system that can predict human movement on real time video. The motivation behind the work is to design a system that can detect pedestrians and their movement to warn the driver beforehand. The system needed to do the lowest amount of computation time as possible and it needed to adapt to changes of the environment in order to work in real time with unstable surroundings. Also it was aimed to do the work without using any sensors other than a single camera. There are many examples of multiple sensor usage besides the camera like IR sensors and usage of multiple cameras. It was aimed to design an alternative system that worked good enough in real time while keeping the complexity level in minimum. Kalman Filter usage also serves the purpose of low complexity and low time cost. As Kalman Filter only takes two inputs, previous and current location, it is significantly faster(computation time wise) compared to other methods like neural networks and also adapts well the changing environmental uncertainties. For the adaptation part to work P, Q and R values of the Kalman Filter are updated according to the miscalculations made in the previous prediction. Thus the system improved itself to work better in the new environmental conditions in each iteration. After designing a real time human pedestrian detection system using Adaptive Kalman Filter, it was observed that real time pedestrian movements could be tracked with a single camera accurately.

1.1 Purpose of Thesis

In this work, it is aimed to design a robust system that works in real time which can detect and estimate the location of pedestrians in the next frame. There are many different approaches to the real time human tracking problem. In this work, it is aimed to approach the problem in a simple but accurate way. The approach is to only use a single camera as the only sensor for the system. The next frame location estimation is evaluated with the state estimation. An Adaptive Kalman Filter is used for the state estimation phase of the project. Adaptive Kalman Filter is chosen over Kalman Filter for the real time adaptation of the state estimation process. The changing environmental factors needed to be eliminated, which made the Kalman Filter inconvenient as the process noise and measurement noise parameters are not constant with the changing environmental factors of the real time video. Update of process and measurement noise in each frame allows Adaptive Kalman Filter to adapt to new frames. Also as Adaptive Kalman Filter, like Kalman Filter, only uses the data from previous and present frames fast adaptation of the system to the changes is beneficial. For the input parameters of the Adaptive Kalman Filter only the centroids of the detected pedestrians boundary boxes is used which gave a computational advantage(as only a single point per person is evaluated). The boundary box is reconstructed around the predicted centroid output of the Adaptive Kalman Filter.

1.2 Literature Review

Digital images and video sequences have gained importance since the start of multimedia information era. The objects of need can be extracted from digital images or frames of videos. Object detection can be linked with image processing and computer vision which includes segmenting of the images and extracting the ROI of the particular image. In certain cases this can be human, face, animals or trees etc [1]. Machine recognition of objects from digital images and video frames has increased significantly over the last 40 years [2]. An efficient tracking algorithm that can work on real time emerged with the increased demand for surveillance in the

developing world conditions [3]. A digital image can be represented by a function of two variables which corresponds to the x and y coordinates of the pixels [4]. The field that works on increasing the quality of the image and getting the information of need from an image by using operations like removing noise and limiting the frame to the ROI's is named as digital image processing [5]. Segmentation of an image, is one of the fundamental processes of digital image processing which is mainly used for separating the images into parts to work on ROI's, not the whole frame. Segmentation also allows to annotate the image and present more information on the object of interest [6-7].

1.2.1 Human detection

The size and location of humans in a digital image or video frame is uncertain in daily applications. Thus the whole frame is needed to be evaluated for finding human figures and each separate object needs to be classified whether it is human or not [8].

For video input, foreground extraction is commonly used for subtracting the stable objects from the moving ones. However this method can be used only in stable cameras where background information is presented to the system [8].

Features are one of the most common areas of study for human classification. Features of humans can be obtained by examining edges, textures, color or motion. Combining multiple features of humans can be used for increasing the performance of the system as it some features can be compensators to the disadvantages of the other features [8].

Object representation is another method for human detection. An object can be divided into parts to make the features work. Sampling the frame to grids or extending the regions on a local point of interest are two methods to divide the frame into ROI's [8].

Binary classifiers and learning methods are widely used in human detection researches. Usage of classifiers aim to increase the performance(efficiency and accuracy of the algorithm) of the learning algorithms. An example for this case is Adaboost framework. Adaboost framework speeds up the classification process by minimizing the number of weak classifiers in its cascade [8].

For supporting the learning process there are multiple datasets dedicated to classify humans in an image. These datasets include images from different occasions. Resolution, format, pose and distance are some parameters effecting the scope of the dataset [8].

Pedestrian detection and tracking is one of the key problems in computer vision on the fields like automative industry, surveillance and robotics. Public dataset updates enabled development in pedestrian detection in recent years [9].

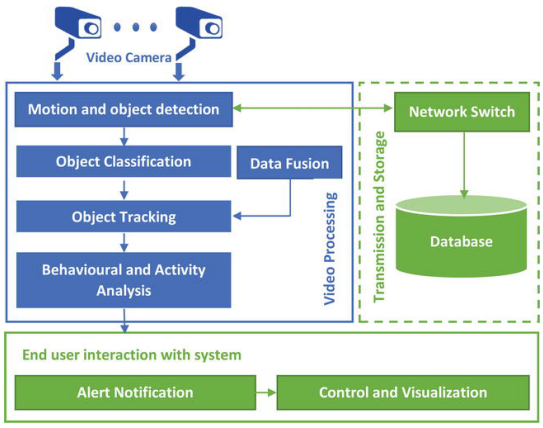

Figure 1.1 : General framework for an automated surveillance system [10].

Overview of an example of a surveillance system is represented on Figure 1.1. The traditional version included steps like object detection, motion detection, classification, tracking, behavior estimation and activity analysis steps as shown in Figure 1.1.

1.2.2 Pose estimation

There is a significant progress in the field of human pose estimation in recent years. The field is still in progress due to the limitations in the limited dataset numbers. The overall approach to pose estimation problem is to train the system with the existing datasets and evaluate the results [11]. Complex appearance models are evaluated on learning algorithms to model from the provided datasets. Performance of the system directly relies on the dataset used. Range of annotated images with different surroundings, clothing, occlusions and truncation at the corners effect the performance of the system directly. For pose estimation there are multiple approaches like taking full body in consideration or splitting the body in parts(head,

upper body, legs, arms) and those approaches rely in on the datasets on their own. In a comparison research it was proved that all of these methods have their benefits on the accuracy of pose estimation with their own datasets and estimation models [12].

1.2.3 Kalman Filter

The usage of Kalman Filter is for state estimation. Kalman Filter is one of the most famous techniques for state estimation [13]. Kalman Filter is a filter used for estimating the instantaneous state of a dynamic system which has noisy observations on its states [14]. There are two models that Kalman Filter rely on. System behavior over time is described by plant dynamic model. Process and observation noise properties are described by the stochastic model [15-16].

Noise statistics influence the Kalman gain which represents the error between the measurement and process information. The best estimate is projected on the process information with the Kalman Filter gain. Regarding this information it can be concluded that noise parameters have a direct effect on the performance of the Kalman Filter [13]. The discrete time linear state space model for Kalman Filter is represented in equations 1.1 and 1.2.

$$x_k = Ax_{k-1} + Bu_{k-1} + \upsilon_{k-1} \tag{1.1}$$

$$y_k = Hx_k + v_k \tag{1.2}$$

The $x_k \in R^n$ is an n-dimensional state vector with initial state value x_0 which has a mean m_0 and covariance P_0 (i.e. $x_0 \sim N(m_0, P_0)$), $A \in R^{n \times n}$ is the state matrix, $B \in R^{n \times m}$ is the input matrix, $u \in R^m$ is the system input, $\upsilon \in R^n$ is the Gaussian process noise with zero mean and constant covariance Q (i.e. $\upsilon \sim N(0, Q)$), $v \in R^d$ is the Gaussian measurement noise with zero mean and constant covariance R (i.e. $v \sim N(0, R)$) , $y \in R^d$ is a d-dimensional measurement vector, $H \in R^{d \times n}$ is the output matrix, and k is the time index [13]. A, B and H matrices are known and

m_0, P_0 are feeded to the system before the application of Kalman Filter. The assumptions are explained in equations 1.3-1.5.

$$E(\upsilon_k) = E(v_k) = E(\upsilon_k v_i^T) = 0 \tag{1.3}$$

$$Q = \delta_{k\,i}\, E\,(\,\upsilon_k\, \upsilon_i^T\,)\;;\; R = \delta_{k\,i}\, E\,(\,v_k\, v_i^T\,) \tag{1.4}$$

$$\delta_{k\,i}=1 \text{ if } i=k;\ \delta_{k\,i}=0 \text{ if } i\neq k \tag{1.5}$$

Also it is assumed that inputs are constant over the sampling time, the noise covariances are constant and process and measurement share the same sampling time [13]. The equations in 1.1 becomes the previous state in the prediction step and is represented like in 1.6 and 1.7 respectively.

$$x_k^- = Ax_{k-1} + Bu_{k-1} + \upsilon_{k-1} \tag{1.6}$$

$$P_k^- = AP_{k-1}A^T + Q_{k-1} \tag{1.7}$$

In measurement update steps values of kalman gain(K), estimation error covariance(P), and estimated state(x) is updated. Measurement update steps are explained in equations in 1.8-1.10.

$$K = P^-H^T(HP^-H^T+R)^{-1} \tag{1.8}$$

$$x=x_k^-+K_k\,(z_k-Hx_k^-) \tag{1.9}$$

$$P_k=(I-K_kH)P_k^- \tag{1.10}$$

1.2.4 Adaptive Kalman Filter

As a result of the uncertainty of stochastic model parameter values Kalman Filter performance diverges or degrades[17-18]. Due to the importance of the noise characteristics on the performance of the Kalman Filter, the researches continued on improving the performance by adjusting the static noise parameters and reducing their effect on systems. The result of these researches is called an Adaptive Kalman Filter [13]. There are three categories of AKF's which can be found in literature; innovation based adaptive estimation, multiple model adaptive estimation and scaled state covariance method [19].

Process noise covariance matrix Q and/or the measurement covariance R is estimated in IAE method. The methods used in IAE are maximum likelihood, correlation and covariance matching. Estimated Q is evaluated by calculated sample covariance and sample covariance of the state error is computed by using the past data. R is computed similarly using sample covariance of KF innovation [20-23]. Covariance matrices are estimated on correlations of samples between the steps with the relations of innovation covariance and estimation error covariance [24-27]. However this method does not be sure that the matrices are positive definite and the estimates of covariances are biased [28]. Also as these methods use a large window of data, it is not efficient to use them. The methods that use maximum likelihood, maximizes likelihood function of the innovation to estimate the covariance [29]. Drawback of these methods are they can not be implemented online due to the computation load [13].

Different models are considered in MMAE method with the assumption of one of the models is correct. The probability of the correctness is calculated for each model, Bayesian probability, and either the one with highest probability or the weighing sum of models is taken to process [30]. For the dynamic systems that possess uncertainty assumption of a single model is not suitable [31]. To overcome the uncertainty in the model an optimization based adaptive estimation was developed by Karasola and Hu [32].

The method proposed by Hashlamon and Erbatur is explained in equations 1.11-1.26 [13].

Initial values $\bar{\omega},\bar{e},\hat{x},P,N,N,Q>0,R>0,$ (1.11)

$$\hat{x}_k^- = A\hat{x}_{k-1} + Bu_{k-1} \tag{1.12}$$

$$P_k^- = AP_{k-1}A^T + Q_{k-1} \tag{1.13}$$

$$e_k = z_k - H\hat{x}_k^- \tag{1.14}$$

$$\alpha_1 = (N_R-1)/N_R \tag{1.15}$$

$$\bar{e}_k = \alpha_1\bar{e}_{k-1} + (1/N_R)e_k \tag{1.16}$$

$$\Delta R_k = (1/(N_R-1))(e_k-\bar{e}_k)(e_k-\bar{e}_k)^T - (1/N_R)(HP^-H^T)_k \tag{1.17}$$

$$R_k = |\mathrm{diag}(\alpha_1 R_{k-1} + \Delta R_k)| \tag{1.18}$$

$$K = P_k^- H^T (HP_k^- H^T + R_k)^{-1} \tag{1.19}$$

$$\hat{x}_k = \hat{x}_k^- + K_k e_k \tag{1.20}$$

$$P_k = (I - K_k H)P_k^- \tag{1.21}$$

$$\hat{\omega}_k = \hat{x}_k - \hat{x}_k^- \tag{1.22}$$

$$\alpha_2 = (N_Q-1)/N_Q \tag{1.23}$$

$$\bar{\omega}_k = \alpha_2\bar{\omega}_{k-1} + (1/N_Q)\omega_k \tag{1.24}$$

$$\Delta Q_k = (1/N_Q)(P_k - AP_{k-1}A^T)_k + (1/(N_R-1))(\hat{\omega}_k-\bar{\omega}_k)(\hat{\omega}_k-\bar{\omega}_k)^T \tag{1.25}$$

$$Q_k = |\mathrm{diag}(\alpha_2 Q_{k-1} + \Delta Q_k)| \tag{1.26}$$

An example flow chart for Adaptive Kalman Filter is presented on the flowchart in Figure 1.2.

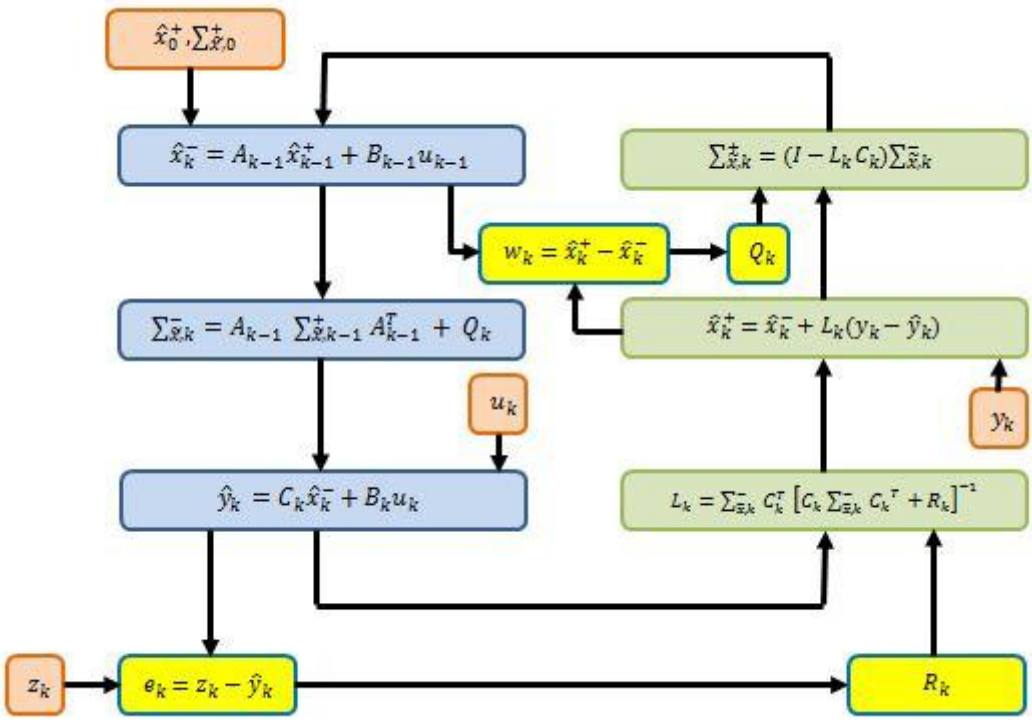

Figure 1.2 : Flowchart of Adaptive Kalman Filter [33].

1.2.5 Backpropagation neural network

Backpropagation is one of the widest applied architectures of the artificial neural networks. As complex multidimensional mappings can be learned by backpropagation networks, backpropagation neural networks are popular in the field [34].

Backpropagation architecture design consists of layers that are interconnected to each other. Gradient of the loss function with respect to the weights of the nodes is computed by the chain rule(layer by layer), iterating in backward direction [34].

1.2.6 Levenberg-Marquardt method

Levenberg-Marquardt algorithm was designed for solving the nonlinear least squares problems. The least squares objective is quadratic in parameters if the model is linear in parameters and can be solved in one step by solving the linear matrix equation. If not(non-linear) an iterative approach is needed to solve the problem. Sum of square errors, data points and model function, are reduced via updating the model parameters iteratively. Gradient descent method and Gauss-Newton methods are used in Levenberg-Marquardt algorithm. If the parameters are close to the optimal value then the system acts more like Gauss-Newton. If the parameters are far from the optimal value then the system acts like gradient descent [35].

1.2.7 Bayesian Regularization method

Bayesian regularization is a robust method that reduce or eliminate the costs cross validation process of standart backpropogation neural networks. A nonlinear regression is converted into a statistical problem with Bayesian Regularization method. In Bayesian Regularization method network parameters are trained and calculated according to their effectiveness. The irrelevant network parameters are given less weight and the relevant ones are boosted for improvement in performance. This process allows network to evaluate the importance of each input [36].

1.3 Hypothesis

It is hypothesized that by using Adaptive Kalman Filter, human movement estimation can be made on real time video with a single camera.

2.MATERIALS AND METHODS

2.1 Materials

For this work it is aimed to detect pedestrians and predict their movement with a single camera. For this purpose the hardware used is a single integrated MacBook camera. The camera has 1280x720 resolution and its model is 'FaceTime HD Camera (Built-in)'. Both for process speed and cost efficiency the built in camera of the computer which processes the code is used.

For the software part MATLAB is used for coding the system. Selection of MATLAB is due to the wide range of libraries that it includes for both control and image processing applications.

2.2 Methods

The system is implemented on MATLAB and the only hardwares used in the system is the computer for processing and a single built in webcam. Overall architecture of the code is as follows:

First the webcam object is created on the source code and then the preliminary values of P, Q and R are defined as the adaptive Kalman needs a previous state. The values are found by trial and error and the equations for initial values are defined as in Equation 2.1.

$$P_0=[20,10],\ Q_0=[5,5],\ R_0=[100] \tag{2.1}$$

After defining the initial values of the system the real time video needed to be converted to a digital image to process. Thus the snapshot of the video stream is taken as the first step of the loop. The snapshotting process is done in every iteration

which makes the output stream work as a real time video with little delay. Considering the system works between 10-15 frames, this delay is negligible.

Snapshotted frame is then used as the input for the human detection algorithm. The human detection algorithm uses the default MATLAB algorithm of 'peopleDetectorACF' which uses the trained algorithm on Caltech Pedestrian Dataset. In the pedestrian detection algorithm a thresholding algorithm is used for eliminating the false positive classifications made by detection algorithm.

As a result of the detection algorithm the boundary boxes of human figures are obtained. These boundary boxes are not used directly but they are used to get the centroid values and later on will be used to reconstruct the boundary boxes to be annotated on the newly predicted centroids.

Centroids are feeded to the Adaptive Kalman Filter algorithm with the previous error parameters, estimations and actual values of the previous iteration. The motion model of the Adaptive Kalman Filter is chosen as constant velocity as the constant acceleration case is not applicable for the pedestrian movement case. Calculation of noise parameters Q and R are explained in equations 2.2 and 2.3.

$$Q=[abs((var([predictedCentroids\ postCentroids]))/1024+0.01)\ 1] \quad (2.2)$$

$$R=abs((predictedCentroids(1)-postCentroids(1))/2048) + abs((predictedCentroids(2)-postCentroids(2))/2048)+0.01 \quad (2.3)$$

For Q and R, as stated in the algorithm of Hashlamon and Erbatur [13], an adaptive noise estimation algorithm is developed based on the estimated and present centroid data differences.

After predicting the centroids P and K values are calculated from the prediction of the Adaptive Kalman Filter. Boundary boxes are also reconstructed from the centroid values. Depending on the size of the boundary box, size of the figure directly relates with the distance of the object from the camera, object is annotated with either red(warning) or green (safe).

After all the states are concluded, in order to start the next step it is needed to check whether the number of centroids are the same with the previous one. If there are

more centroids on the next state, the new objects previous state is taken as the same value with the present state to make the algorithm work. If vice versa, there are more centroids on the previous state, the closest centroid values of the present state is matched with the ones on the previous state and the excluded ones are eliminated.

For comparison of the performance of Adaptive Kalman Filter with the learning algorithms, two neural network algorithms are implemented. As the outputs of the next state is known and can be used in the training process, backpropagation neural networks are used for comparison. Deep network toolbox of the MATLAB IDE enables to design backpropagation neural networks that can work on two different algorithms; Levenberg-Marquardt and Bayesian Regularization. For training both of these networks, 1500 frames of images are recorded. The position of detected human is used used as an input and the location of the same person in the next frame is used as the actual output for that input. Trained networks are used as predictors in real time.

3. RESULTS AND DISCUSSION

3.1 Results

For evaluating the results of the system 1000 frames of data is recorded. Comparison of the results of 1000 frames with Bayesian Regularization, Levenberg-Marquardt and AKF 500 frames is shown in table 3.1.

Table 3.1 : Experimentation results of the system.

	True Positive	True Negative	False Positive	False Negative	Training	Accuracy	Precision
Adaptive Kalman Filter 1000 Frames	1418	0	213	47	0	%84.5	%86.9
Bayesian Regularization	362	5	104	29	Around 400 frames	%73.4	%77.8
Levenberg-Marquardt	384	5	86	25	Around 20 frames	%77.8	%81.7
Adaptive Kalman Filter 500 Frames	397	5	63	35	0	%80.6	%85.9

Some of the successful experimentation results can be found in figures 3.1-3.3. In figure 3.1 the frontal view of a successful classification is present, in figure 3.2 side view of a successful classification can be observed and in figure 3.3 successful classification of a rear view is shown.

Figure 3.1 : Successful result for frontal view.

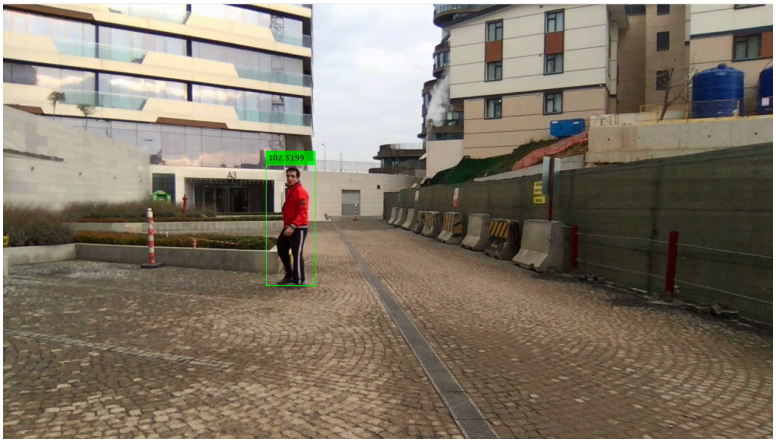

Figure 3.2 : Successful result for side view.

Figure 3.3 : Successful result for rear view.

The system is tested for multiple humans as well. The result of estimation for multiple humans is shown in figure 3.4.

Figure 3.4 : Successful result for multiple objects.

After the completion of the recordings, the noise values(P, Q and R) and Kalman gain are plotted for 500 frames. Plots of Q, R, P and K are shown in figures 3.5-3.8 respectively.

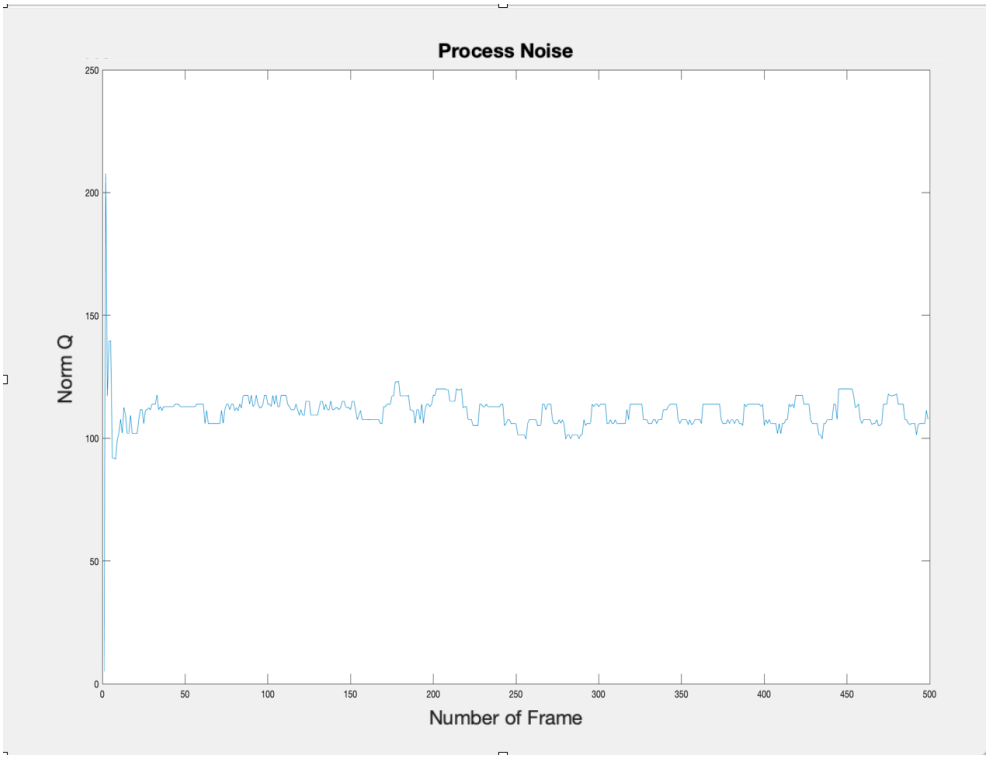

Figure 3.5 : Norm Q vs number of frames plot.

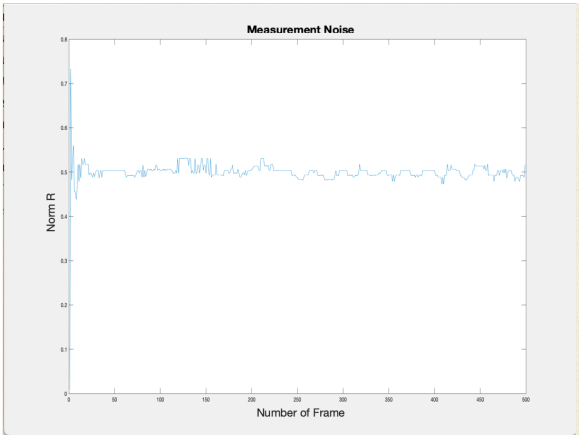

Figure 3.6 : Norm R vs number of frames plot.

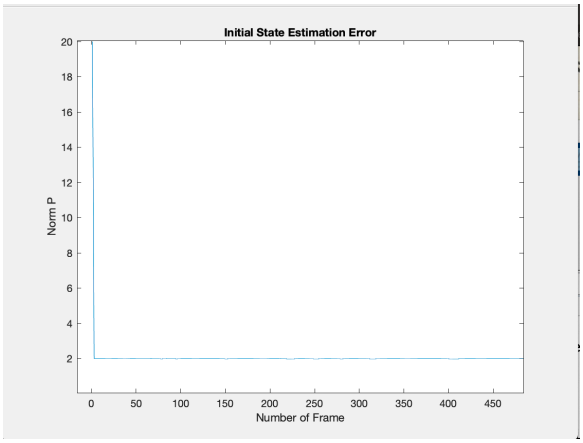

Figure 3.7 : Norm P vs number of frames plot.

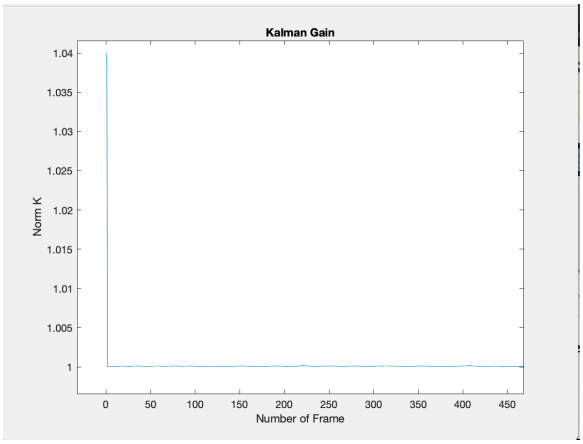

Figure 3.8 : K vs number of frames plot.

The results of location estimates and actual locations are evaluated for 500 frames for all the implemented methods. Plots of estimations vs actual centroids of Adaptive Kalman Filter, Bayesian Regularization and Levenberg-Marquardt methods are

shown in figures 3.9-3.11 respectively. Green lines in figures represent the actual values of the human locations and red stars represent the estimated locations.

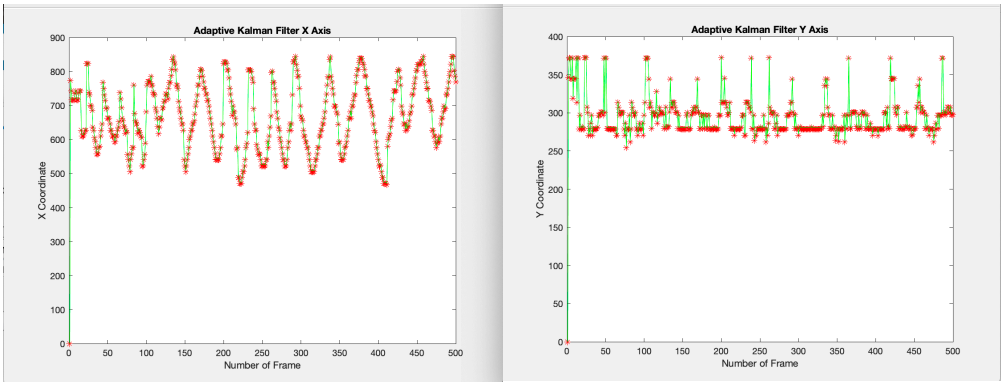

Figure 3.9 : Adaptive Kalman Filter actual vs estimated locations.

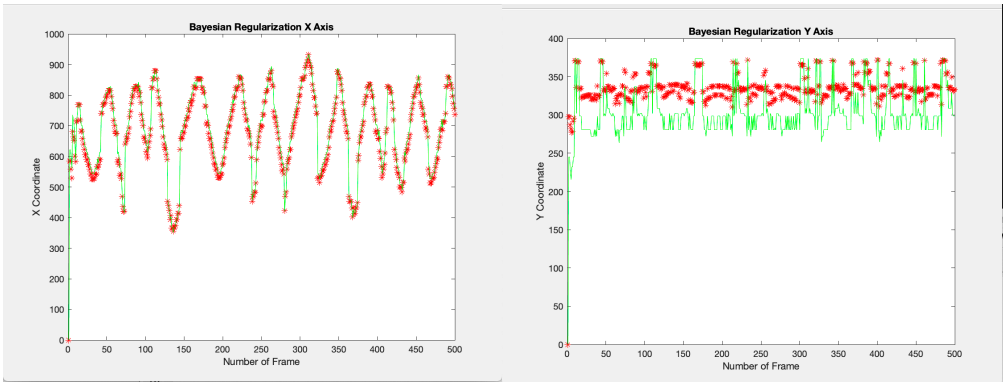

Figure 3.10 : Bayesian Regularization actual vs estimated locations.

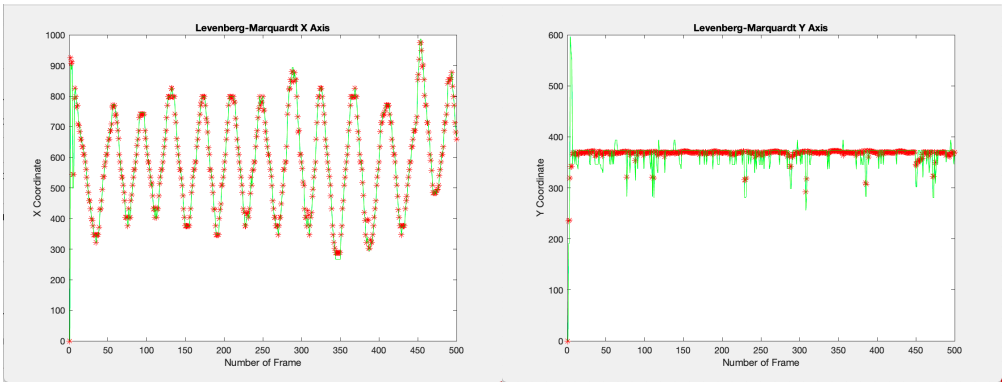

Figure 3.11 : Levenberg-Marquardt actual vs estimated locations

3.2 Discussion

In this work it was aimed to design a robust system that can successfully detect pedestrians in the frame and estimate their movement by using an Adaptive Kalman Filter. In the 1000 test frames the system is tested with single and multiple

pedestrians in the frame. Out of total 1678 samples 1418 of them are the right classifications which results in a %84.5 success rate. In the test process, it was observed that the error is mainly caused when two objects move towards each other. From that observation it can be concluded that the system works better in areas where the crowd is less dense.

In order to compare the performance of the system, two more learning methods(Bayesian Regularization and Levenberg-Marquardt) is implemented. When the accuracy and precision values are compared, it was observed that the Adaptive Kalman Filter gave better results than the other methods. Also as Adaptive Kalman Filter does not require any training phase it is a better fit for the real time applications.

When the Kalman Gain and error parameters are observed, it is visible that these parameters converge to an interval which means that the filter adapts itself to the environment, in other words the stopping criteria is successfully achieved. If the Kalman Filter was used instead of Adaptive Kalman Filter then it would not be possible to correctly tune the system due to the environment changes, which will result in inconsistencies in the predictions.

With this work, it is shown that the pedestrian detection and motion estimation can be done using a single camera to some extend. Combining the method followed with an algorithm that can separate crowds from each other would increase the system performance significantly.

4. CONCLUSIONS AND RECOMMENDATIONS

For this work it was hypothesized that by using Adaptive Kalman Filter, human movement estimation can be made on real time video with a single camera. It was observed that the hypothesis is proved with %84.5 success rate of the system built. It was also observed that choice of Adaptive Kalman Filter for eliminating the challenges of changing environmental condition of real time video . It was also important to observe that the Kalman Gain and error parameters converged to a small interval of values, which was essential to prove the learning process of the system. In conclusion, this thesis work showed that human movement estimation can be made on real time video with a single camera.

Despite the success ratio is high enough to prove that this is an efficient method to apply on real time pedestrian tracking, there are some development areas for the system. First of all, systems performance significantly decreases in multiple pedestrians moving towards each other. A better separation method for objects needs to be developed for better results. Secondly, the system works when the whole body of the pedestrian is present in the frame. A torso detection algorithm can be implemented for classifying pedestrians that are too close to the camera. Lastly, testing with a better computation device can increase the number of frames per second which will result in a more sensitive classification.

REFERENCES

[1] **Hatem H., Beiji Z., Majeed R.** (2015). A Survey of Feature Base Methods for Human Face Detection. International Journal of Control and Automation. 8 (5): 61-78. DOI:10.14257/ijca.2015.8.5.07

[2] **Borah S., Konwar S., Tuithung T., Rathi R.** (2014). A Human Face Detection Method Based on Connected Component Analysis. International Conference on Communication and Signal Processing, 1-4. DOI:10.1109/ICCSP.2014.6950033.

[3] **Chatrath J., Gupta P., Ahuja P., Goel A., Arora S.M.** (2014). Real Time Human Face Detection and Tracking, International Conference on Signal Processing and Integrated Networks, Spin, 1-6. DOI:10.1109/SPIN.2014.6777046

[4] **Avinash R., Shivhare E., Mandloi D.** (2018). Segmentation Based Denoising of Color Images using Morphological Operation. International Journal of Engineering Research and Application. 8(7): 75-79. 10.9790/9622-0807018084.

[5] **Rudnitskii A.G., Rudnytska M.A.** (2017). Segmentation and Denoising of Phase Contrast MRI Image of the Aortic Lumen via Fractal and Morphological Processing, 37th International Conference on Electronics and Nanotechnology (ELNANO), IEEE. 1-5.

[6] **Chudasama D., Patel T., Joshi S., Prajapati G.** (2014). Survey on Various Edge Detection Techniques on Noisy Images, IJERT International Journal of Engineering Research & Technology. 3 (10): 1-5.

[7] **Raman M., Aggarwal H.** (2009).Study and comparison of various image edge detection techniques, International Journal of Image Processing (IJIP). 1 (3): 1- 11.

[8] **Nguyen D.T., Wanqing L., Ogunbona P.** (2015). Human Detection from Images and Videos: A Survey. Pattern Recognition. 51:148-175. 10.1016/j.patcog.2015.08.027.

[9] **Dollar P., Wojek C., Schiele B., Perona P.** (2009). Pedestrian detection: A benchmark. 2012 IEEE Conference on Computer Vision and Pattern Recognition. 304-311. 10.1109/CVPR.2009.5206631.

[10] **Gawande U., Hajari K., Golhar Y. (**2020). Pedestrian Detection and Tracking in Video Surveillance System: Issues, Comprehensive Review, and Challenges DOI: 10.5772/intechopen.90810.

[11] **Ainsworth B.E., Haskell W., Herrmann S., Meckes N., Bassett D., Tudor-Locke C., Greer J., Vezina, Whitt-Glover M., Leon A.** (2011). Compendium of Physical Activities: A Second Update of Codes and MET Values. Medicine and Science in Sports and Exercise . 43 (8): 1575-1581. DOI: 10.1249/MSS.0b013e31821ece12.

[12] **Andriluka M., Pishchulin L., Gehler P., Schiele B.** (2014). Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 3686-3693.

[13] **Hashlamon I., Erbatur K.** (2013). An Improved Real-Time Adaptive Kalman -Filter with Recursive Noise Covariance Updating Rules. Turkish Journal of Electrical Engineering and Computer Sciences. 24: 524-540. DOI: 10.3906/elk-1309-60.

[14] **Grewal M.S., Andrews A.P.** (2001). Kalman Filtering: Theory and Practice Using MATLAB. 2nd ed. New York, NY, USA: John Wiley & Sons.

[15] **Simon D.** (2006). Optimal State Estimation: Kalman, H Infinity, and Nonlinear Approaches. Hoboken, NJ, USA: Wiley & Sons.

[16] **Maybeck P.S.** (1982). Stochastic Models, Estimation and Control. Vol. 3. New York, NY, USA: Academic Press.

[17] **Heffes H.** (1966). The Effect of Erroneous Models on the Kalman Filter Response. IEEE Transactions on Automatic Control. 11: 541–543.

[18] **Fitzgerald R.J.** (1971). Divergence of the Kalman Filter. IEEE Transactions on Automatic Control. 16: 736–747.

[19] **Mohamed A., Schwarz K.** (1999). Adaptive Kalman Filtering for INS/GPS. Journal of Geodesy. 73: 193–203.

[20] **Myers K., Tapley B.D.** (1976). Adaptive Sequential Estimation with Unknown Noise Statistics. IEEE Transactions on Automatic Control . 21: 520–523.

[21] **Ding W., Wang J., Rizos C., Kinlyside D.** (2007). Improving Adaptive Kalman Estimation in GPS/INS Integration. Journal of Navigation. 60: 517.

[22] **Shi E.** (2012). An Improved Real-Time Adaptive Kalman Filter for Low-Cost Integrated GPS/INS Navigation. 2012 International Conference on Measurement, Information and Control (MIC). 1-6. DOI: 10.1109/ MIC.2012.6273443.

[23] **Hide C., Moore T., Smith M.** (2003). Adaptive Kalman Filtering for Low-Cost INS/GPS. Journal of Navigation. 56: 143–152.

[24] **Mehra R.K.** (1970). On the Identification of Variances and Adaptive Kalman Filtering. IEEE Transactions on Automatic Control . 15: 175–184.

[25] **Belanger P.R.** (1974). Estimation of Noise Covariance Matrices for a Linear Time-Varying Stochastic Process. Automatica. 10: 267–275.

[26]**Oussalah M., Schutter J.D.** (2001). Adaptive Kalman Filter for Noise Identification. In: Proceedings of the International Seminar on Modal Analysis. Kissimmee, FL, USA. Leuven, Belgium: KU Leuven. 1225–1232.

[27] **Odelson B.J., Rajamani M.R., Rawlings J.B.** (2006). A new Autocovariance Least-Squares Method for Estimating Noise Covariances. Automatica. 42: 303–308.

[28] **Odelson B.J., Lutz A., Rawlings J.B.** (2006). The Autocovariance Least-Squares Method for Estimating Covariances: Application to Model-Based Control of Chemical Reactors. IEEE Transactions on Control Systems Technology. 14: 532–540.

[29] **Raghavan H., Tangirala A.K., Gopaluni R.B., Shah S.L.** (2006). Identification of Chemical Processes with Irregular Output Sampling. Control Engineering Practice. 14: 467–480.

[30] **Wang X.** (2003). Vehicle Health Monitoring System Using Multiple-model Adaptive Estimation. University of Hawaii at Manoa, Manoa, HI, USA.

[31] **Rong Li X., Youmin Z.** (2000). Multiple-model estimation with variable structure part V: Likely-Model Set Algorithm. IEEE Transactions on Aerospace and Electronic Systems. 36: 448–466.

[32] **Karasalo M., Hu X.** (2011). An Optimization Approach to Adaptive Kalman Filtering. Automatica. 47: 1785–1793.

[33] **Yıldırım K.** (2021). Elektrikli Araçlardaki Lityum-İyon Bataryalar için Modelleme ve Şarj Durumunun Kestirimi. MSc, Istanbul Technical University.

[34] **Hecht-Nielsen R.** (1989). Theory of the backpropagation neural network. International 1989 Joint Conference on Neural Networks, 593-605 vol.1.

[35] **Gavin H.P.** (2013). The Levenberg-Marquardt method for nonlinear least squares curve-fitting problems c ©.

[36] **Burden F., Winkler D.** (2008). Bayesian regularization of neural networks. Methods in molecular biology (Clifton, N.J.), 458, 25–44.

CURRICULUM VITAE

Name Surname **:Coşkun Orkun VURAL**

EDUCATION **:**

- **B.Sc.** **:** 2018, Bilkent University, Faculty of Engineering, Electrical and Electronics Engineering

PROFESSIONAL EXPERIENCE AND REWARDS:

- 2018-2019 Vodafone Turkey — Discover Attendee
- 2019-Present Vodafone Turkey — Solution Designer
- 2019-Present Vodafone Turkey — Scrum Master