Moving Object Detection using Lab2000HL Color Space with Spatial and Temporal Smoothing

Muhammet Balcilar[1,*], *M. Fatih Amasyali*[1] *and A. Coskun Sonmez* [2]

[1] Department of Computer Engineering, Yildiz Technical University, EEF Block, D124, Davutpasa, 34220, Istanbul, Turkey
[2] Faculty of Computer and Informatics, Istanbul Technical University, Maslak, 34469, Istanbul, Turkey

Abstract: In order to detect moving objects such as vehicles in motorways, background subtraction techniques are commonly used. This is completely solved problem for static backgrounds. However, real-world problems contain many non-static components such as waving sea, camera oscillations, and sudden changes in daylight. Gaussian Mixture Model (GMM) is statistical based background subtraction method, in which values of each pixels features are represented with a few normal distributions, partially overcame such problems at least. To improve performance of GMM model, using spatial and temporal features in Lab2000HL color space which have linear hue band, is proposed in this study. The spatial and temporal features performed by using spatial low-pass filter and temporal kalman filter respectively. As a performance metric, the area under the Precision Recall (PR) curve is used. In addition to videos existing in the I2R dataset, a new dataset which images gained from traffic surveillance cameras placed over the entrance of the Istanbul FSM Bridge at different times of the day used for compare proposed method against other well-known GMM version. According to our tests proposed method has been more successful to the other methods in most cases.

Keywords: Moving object detection, Background subtraction, Lab2000HL, GMM, Kalman smoothing

1 Introduction

Identifying moving objects from video images is a basic and critical issue in computer vision applications. A well-known approach to perform this is the background subtraction technique, which determines significantly different regions of the background model as foreground and deletes the remaining parts. There exist two basic common steps of background subtraction techniques except preprocessing and post processing, which are, forming the up-to-date background model that represents the background image and determining the foreground regions which significantly differ from the background model. Background subtraction methods can be classified up to background initialization, methods used for background modeling, features used in modeling and foreground detection techniques. Although in some applications some extra post processing procedures are applied in order to reduce foreground detection errors, these methods are considered indifferent.

Many successful background subtraction algorithms are maintained in the literature. The first problem faced has been light changes. Initially, methods independent from light changes were introduced. Then methods finding non stationary objects in the background such as leaves, sea waves, rain and snow, and finally methods distinguishing temporary movements of objects such as stopping and moving again were developed [1]. In terms of models used, background subtraction methods introduced up to present can be classified as, Deterministic, Statistical, Filter-Based Estimation and Fuzzy methods [2]. Up to image processing methods, these methods can be grouped into two as pixel based and region based. In addition to these, there are methods which use one or more features such as shape, texture, color, gradient, histogram and motion.

The method in which for each pixel, mean and median values of past n pixels data are assigned as background is introduced in [3]. This method requires extremely large memory with respect to the variable n. On the other hand, running average method does not save previous values and a new value affects the background as much as the learning coefficient [4]. Therefore, memory requirement is reduced. Using Gaussian distributions is the most

* Corresponding author e-mail: muhammet@yildiz.edu.tr

popular approach for statistical modeling of each pixel. The simplest method is to detect an average of past values of the pixel, subtract this from the pixel value and obtain a binary form from the result up to a determinate threshold. The adaptive application of this method consists of repetitive updates of model parameter. A single Gaussian distribution model in which background is considered stationary is also introduced in [5]. However, this method is not applicable for modeling backgrounds consisting of periodic movements. Expression of the background in terms of more than a single Gaussian distribution is called Gaussian Mixture Model [6], and required parameter values of the model is computed with the online k-means method. Many researchers added different extensions to the GMM method afterwards [7]. The weakest parts of these methods were the assumption that the distribution of past values of the pixel would fit the Gaussian distribution, and the need for estimation of these distribution parameters. A solution to this problem was proposed in [8] which is background modeling with the KDE (Kernel Density Estimator) technique. After that, there has been many background subtraction methods proposed which use KDE such as [9]. However, this method has also a very huge memory requirement.

Another method in background modeling is clustering method. The background subtraction models that use k-means clustering and sequential clustering are given in [10] and [11] respectively. But, the most remarkable one of clustering based models is the codebook method [12]. This method uses vector quantization technique which is an unsupervised learning algorithm. Subspace learning based background subtraction methods, which is studied especially in recent years, has been developed with many versions. Recently, incremental robust PCA (Principal Component Analysis) which computes the transformation matrices considering outliers is also stated in [13]. The background subtraction methods which use Independent Component Analysis (ICA) are given in [14]. Incremental Maximum Margin Criterion (IMMC) is on the other hand presented in [15]. Background subtraction, in its nature, involves many uncertainties. To get rid of them, the method which reconstruct the classical GMM method with type-2 fuzzy membership functions is stated in [16], as applications in which background modeling is done by many fuzzy similarities are in [17]. Background modeling with fuzzy logic rules on the other hand is presented in [18].

Reviewing the literature according to the features used in GMM method, it is observed that researchers have used different color space and different features together or separately. The study in [19], has used the YUV color space values. Other study, RGB color components in 8x8x3 blocks obtained from three frames in time of the 8x8 neighborhood of the processed pixel are chosen as features [20]. The method introduced in [21] is called Local Patch GMM. In this study, the RGB color space values of all pixels in a fixed neighborhood of the pixel are used in GMM for modeling. The study in which the

normalized RGB (nRGB) color values of pixels are used in GMM is cited in [22]. The study in which HSV color space is used is in [23]. In [24], HSI color space is used, but HS and I bands are modeled with different GMMs. In the study [25] background modeling is done with Luv color space values of the background. In [26], improved HSL (IHSL) color space, in which chromatic plane is determined with circular statistics, is preferred in background modeling. In [27], the successes of other color spaces in background modeling are compared according to RGB and it is reported that the most succeeded one is YCbCr color space. How much the gradient of the corresponding pixel of incoming frame fits the gradient of the background is computed and thus background/foreground is decided. In [28], the features selected to be used in GMM were the main location and the main angle values of the gradients of pixels in the nxn neighborhood of the pixel. Thus, in this application both location and gradient information are used in background modeling. On the other hand, in [29], background modeling was made using RGB values, gradient magnitude, direction values and 8 Haar features.

The purpose of this study is to improve performance of GMM background subtraction method, by the way of using Lab2000HL color space with spatial and temporal features. Lab2000HL color space, which is an improved version of CIELAB color space, is recently introduced and is thought to perform a better modeling of human perception. Having linear Hue channel is the most advantages of this color space. To perform spatial and temporal feature, adding spatial neighbor and temporal neighbor pixels values to feature vector have heavy computational cost. Beside this situation, spatial filter is used for spatial consistency; kalman filter is used for temporal consistency.

The remaining parts of the paper are organized as following. The second section includes background subtraction methods, and the third section explains features in detail. Performance measure is described in section 4. Section 5 is reserved for application and test results at each datasets. Finally, conclusion is presented in section 6.

2 Background Subtraction

Background modeling methods, trying to model the background with statistical distributions of values of a fixed feature of the pixel, have a significant place in the literature. These methods, unlike background estimation methods, are concerned with the distribution of the values of a fixed feature of the pixel rather than ordering them. The method we used in this study differs from the background subtraction method proposed in [6], in terms of features, but they are similar in basic terms. This method is based on GMM, which is indeed an unsupervised learning method, for clustering data by minimizing the variance. The most significant difference

from the classical GMM is that past values of data are not evaluated with expectation maximization technique; they are evaluated online, step by step, as they are produced and the centers of the clusters, variances and weights are updated stepwise. In this method, each pixel will be modeled with K Gaussian distributions. Each distribution has a weight, mean vector and a covariance matrix. These values of distributions will be updated as soon as a new value of the pixel has arrived. The position of the pixel in the image is expressed with $s = (x, y)$, i represents the number of the distribution, and t the time. Let $w_s^{i,t}$ denote the weight of the ith distribution of the pixel at position s and time t, $\mu_s^{i,t}$ denote the mean vector, and $\Sigma_s^{i,t}$ denote the covariance matrix. In these expressions, weights are in scalars and the mean vector is a row vector with the same size as the pixels feature. The covariance matrix, up to the hypothesis that none of the features has a correlation with each other, is a diagonal matrix, having all entries except the diagonal zero, equivalent number of rows and columns to the mean vector, and is defined in (1).

$$\Sigma_s^{i,t} = \left(\sigma_s^{i,t}\right)^2 I.$$ (1)

I denote the identity matrix, and σ denote the standard deviation of the pixel feature. With respect to the above equation, dimensions of the pixel must all have equal variances. When a new frame arrives, the background models of the processed pixel are arranged in a descending order according to the $w_s^{i,t}/\sigma_s^{i,t}$ rates. Then the minimum of the distributions with total weights over the threshold are decided to be the background and the remaining distributions are considered as foreground distributions. The mathematical expression of this procedure is defined in (2).

$$B = \arg\min_b \left(\sum_{i=1}^{b} w_s^{i,t} > T\right)$$ (2)

In the arrival frame, the Mahalanobis differences to all distributions of the pixels feature vector are computed. If the pixel is far from all distributions, more than k times of the standard deviation of the distribution, then it means that the pixel could not be classified. If otherwise, then it means that it is nearer to the first class satisfying the difference boundary and the parameters of this distribution will be updated. At time $t + 1$, if we denote by $X_s^{i,t}$ the feature vector of the pixel in the arriving frame, the question whether its difference to the ith distribution is in acceptable limits or not, is expressed in the (3).

$$\sqrt{\left(X_s^{t+1} - \mu_s^{i,t}\right)^T \left(\Sigma_s^{i,t}\right)^{-1} \left(X_s^{t+1} - \mu_s^{i,t}\right)} < k\sigma_s^{i,t}$$ (3)

If the vector X_s^{t+1} is included in the ith class, as α to be the learning coefficient, parameters for the ith class are

updated as (4).

$$w_s^{i,t+1} = (1-\alpha)w_s^{i,t} + \alpha$$
$$\mu_s^{i,t+1} = (1-\rho)\mu_s^{i,t} + \rho X_s^{t+1}$$
$$\left(\sigma_s^{i,t+1}\right)^2 = (1-\rho)\left(\sigma_s^{i,t}\right)^2 + \rho\left(X_s^{t+1} - \mu_s^{i,t+1}\right)\left(X_s^{t+1} - \mu_s^{i,t+1}\right)^T$$ (4)

The temporary variable ρ, represents in (5).

$$\rho = \alpha N\left(X_s^{t+1}, \mu_s^{i,t}, \Sigma_s^{i,t}\right)$$ (5)

The function N in the above expression is the probability density function of the multi-dimensional Gaussian distribution. The inputs for the function N are, respectively, the vector for which the probability will be computed, the mean of distribution center and the covariance matrix of the distribution. Since the ith distribution is chosen, the mean vector and the variances of all other distributions will be fixed. But the weights need to be reduced with respect to (6).

$$w_s^{i,t+1} = (1-\alpha)w_s^{i,t+1} \quad j = i, j = 1..K$$ (6)

If none of the distributions has the distance to the corresponding vector of the arriving frame, in this case the class with the smallest weight must be deleted and replaced with a new class with mean as the feature vector of the arriving pixel, having the weight of the lowest deleted weight, and the largest variance value of the remaining distributions as (7).

$$w_s^{i,t+1} = Low\,Weight$$
$$\mu_s^{i,t+1} = X_s^{t+1}$$
$$\left(\sigma_s^{i,t+1}\right)^2 = Large\,Variance$$ (7)

If for each arriving frame, every pixel is updated with these update procedures, then an up to time model of the background can be obtained.

3 Features

Lab2000HL color space is better than other color spaces for background modeling shown in the previous study [30,31]. This section explains, Lab2000HL color space and how spatial and temporal consistency taken into account, instead of using the color value directly.

3.1 Lab2000HL Color Space

Because the information gained directly from the camera sensor is in RGB color space, these color bands are commonly used as a feature. However, these bands are widely affected by the light since because they cannot represent the chromaticy feature of the segment independently. There is a lot of research in the literature on distinctive properties of color spaces. According to
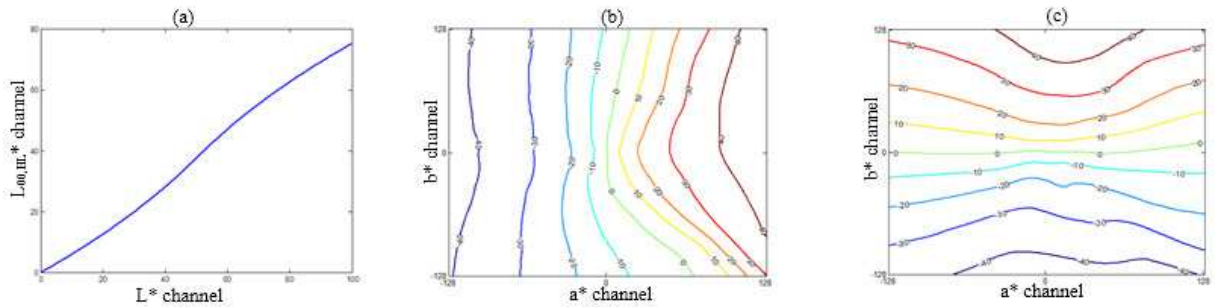
Fig. 1: Transformation map from CIELAB to Lab2000HL, a) Transformation of channel $L^*_{00,HL}$, b) Transformation of channel $a^*_{00,HL}$, c) Transformation of channel $b^*_{00,HL}$

this research, YCbCr and HSV color spaces were more accurate since they represent the amount of light and the chromaticy separately. CIELAB color space, on the other hand, is a color space which indicates proper changes in the direction of human color perception. Its components are the lightness of the color (L^*) and two color-opponent dimensions (a^*, b^*). Lab2000HL (L^*_{00}, $a^*_{00,HL}$, $b^*_{00,HL}$) color space, which is an improved version of CIELAB color space, is recently introduced and is thought to perform a better modeling of human perception [32].

In [32], a numerical method to determine a transformation of a color space into a hue linear color space with a maximum degree of perceptual uniformity proposed. According to [32], the transformation takes place in two steps. To obtain Lab2000HL, CIELAB is determined as initial color space. Because it is well known and used in many industrial standards. The transformations can be summarized as (8) [32].

$$\Psi_{00} : CIELAB \rightarrow Lab2000$$
$$\Psi_{00,HL} : Lab2000 \rightarrow Lab2000HL \qquad (8)$$

In the first step, the transformation provides perceptual uniformity. The transformation of the CIELAB ($L^*a^*b^*$) color space into a Euclidean space with respect to the Lab2000 (L^*_{00}, a^*_{00}, b^*_{00}) color-difference formula has been described in [33]. According to this paper, Ψ_{00} transformation performed by one-dimensional look-up table for $L^* \rightarrow L^*_{00}$ transformation, two-dimensional lookup table for each $a^* \rightarrow a^*_{00}$ and $b^* \rightarrow b^*_{00}$ transformation. Required look-up tables are also presented in that research.

The second step of transformation maps the curves of constant hue to straight lines while preserving the perceptual uniformity as far as possible. Because L^*_{00} represents lightness coordinate and lightness do not effect to hue information, the lightness values are not changed. $\Psi_{00,HL}$ transformation performed only on a^*_{00} and b^*_{00} channels. A mapping of the resulting constant hue curves to straight lines with hue angles $\varphi = \{\varphi_0, \varphi_1, ..., \varphi_{359}\}$ can be performed by a two-dimensional color lookup table. This lookup table must be minimizing the mean

disagreement between CIEDE2000 color differences in CIELAB and Euclidean distances in the new space. The lookup table maps φ to $\varphi^{opt} = \{\varphi_0^{opt}, \varphi_1^{opt}, ..., \varphi_{359}^{opt}\}$. φ^{opt} must be monotonically increasing array. In [35] optimization problem described above is solved. By using lookup table which is results of optimization problem, the transformation $a^*_{00} \rightarrow a^*_{00,HL}$ and $b^*_{00} \rightarrow b^*_{00,HL}$ can be possible. The transformations can be turned into a single transformation by composition: $\Phi = \Phi_{00} \circ \Phi^*_{00,HL}$. The single transformation maps, which presented in [33] is shown as Fig 1. Detailed information about this color space and calculation procedure can be found in [32,33]. RGB to Lab2000HL conversion code is available at [31].

3.2 Spatial Smoothing

The images involve spatial consistency at high rates. When each pixel is processed separately consistency will be disregarded and the performance of the system reduces. In order to take consistency into account, not only the value of the processing pixel in the background subtraction model but also the values of the neighboring pixels may be given. However in this case the length of the feature vector increases at a rate of square the neighborhood length and hence it causes rise in computational cost. Alternatively, in this study, the approach of considering the pixel with pixels in its fixed length neighborhood and producing values for each color channel separately is preferred. When determining the representing value of the pixel group, features of pixel values such as entropy, edge, texture, mean, etc. are used in literature. In this study, low-pass filter response for the pixel group is used as feature.

The filters preferred in this study are the filters which approximately calculate the human perceiving color difference between two images and are used in S-CIELAB metric. In this method, the received image is first transformed into opponent color space (AC_1C_2) in order to make it independent of the input device. AC_1C_2 color space has one luminance channel and two
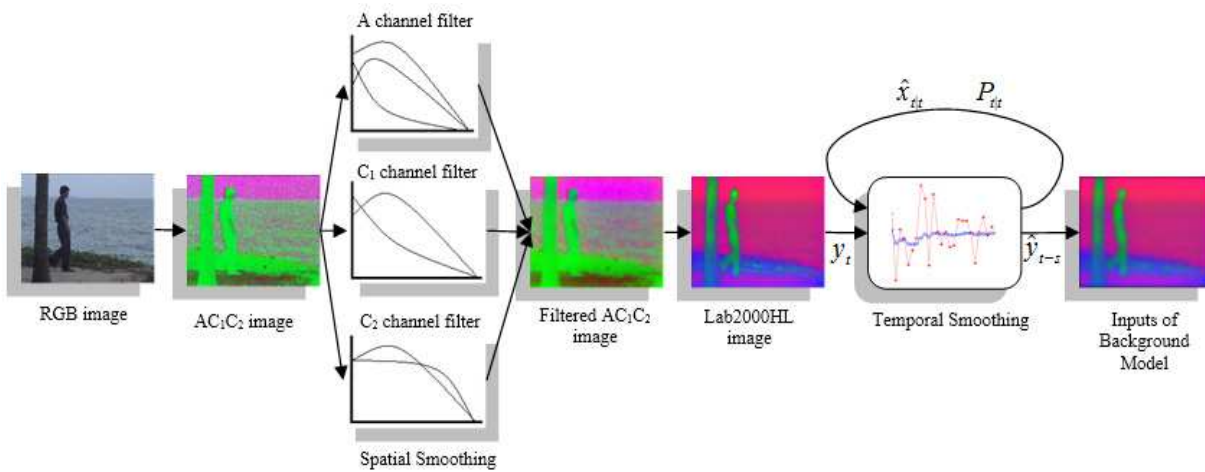
Fig. 2: Diagram of obtaining the background subtraction inputs.

chrominance channels. These channels were determined through a series of psychophysical experiments testing for pattern color separability [34]. Without being completely orthogonal, all channels of this color space are obtained with linear transformations from CIEXYZ color space. After the color space transformation is made, spatial filters special to each channel are applied using lters that approximate the contrast sensitivity functions (CSF) of the human visual system [34]. 3 filters for Luminance channel and 2 filters for each chrominance channel are applied. After filter responses are gathered, the process lifts to the desired color space by applying reverse color space transformation again. Each Gaussian filter has different spread and different weight from others; these values are given in Table.1 [34]. The place of spatial consistency in the procedure of obtaining input for background subtraction is shown in Fig.2

Table 1: Weight and spread of filters

Filter	Weight	Spread
Channel A	1.00327	0.0500
Channel A	0.11442	0.2250
Channel A	-0.11769	7.0000
Channel C_1	0.61673	0.0685
Channel C_1	0.38328	0.8260
Channel C_2	0.56789	0.0920
Channel C_2	0.43212	0.6451

3.3 Temporal Smoothing

The values of a pixel in the time axis are consistent with each other. If pixel's temporal feature is not given to the

background model then temporal consistency happens to be disregarded and accordingly much error occurs. In Literature, in order to take temporal features in to account, pixel's past values of a fixed time interval are added to the feature vector. This process increases the computational cost since it will increase the dimension of the feature vector for a discrete time interval. Alternatively, rather than using directly the value of the pixel, it is a better solution to use its values which are filter processed in the time axis. Moving average filter, since it evaluates all values in its filter mask by multiplying with same coefficient, lacks off adapting to sudden signal changes at the same time with achieving desired rate of noise removal. To handle this problem, first-order low-pass filter for time series can be used. Up to this method, the smallest coefficient is applied to the oldest pixel value in the filter mask and the largest coefficient is applied to the newest pixel value. However in this method, filter coefficients are fixed, they do not change with respect to time. Moreover it is not concerned about how the signal changes or what its dynamics are. It makes filtering for each signal with constant coefficients during the signal. Therefore it is almost well adapted to sudden changes. Kalman filtering, for time series filtering process, solving problems stated above, is a widely used filtering method [35]. It is used in engineering for time series filtering, for prediction of the next value and for fusion of information gathered from several sources [35]. The most significant advantage of using Kalman filter is that it processes the signal with considering its dynamics. Therefore, it is possible to gain more proper filtering and prediction. With Kalman filter, the process of determining the position variable at time t using measurements up to time t is called estimation, determining the position variable at time $t+1$ is called prediction, and determining the position variable at time $t-s$ is called smoothing. While estimation and prediction can be done real-time
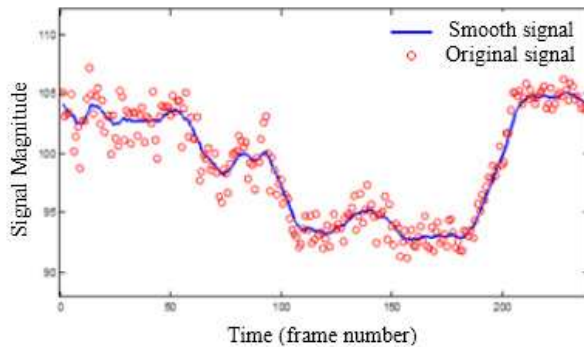
Fig. 3: Effect of Temporal Smoothing Process.

Table 2: Temporal Smoothing Algorithm

Algorithm: Kalman Smoothing with Time Delay

Inputs: $\hat{x}_{0|0}, P_{0|0}, Q, R, A, H, s$

for $t = 0$ **to** $s - 1$ **do**

$\quad \hat{x}_{t+1|t} = A\hat{x}_{t|t}$

$\quad P_{t+1|t} = AP_{t|t}A^T + Q$

$\quad K_{t+1} = P_{t+1|t}C^T \left(CP_{t+1|t}C^T + R \right)^{-1}$

$\quad \hat{x}_{t+1|t+1} = \hat{x}_{t+1|t} + K_{t+1} \left(y_{t+1} - C\hat{x}_{t+1|t} \right)$

$\quad P_{t+1|t+1} = P_{t+1|t} - K_{t+1}CP_{t+1|t}$

endfor

for $index = 0$ **to** end **do**

$\quad T = index + s$

$\quad$ **for** $t = T - 1$ **to** $index$ **step** -1 **do**

$\quad\quad L_t = \left(P_{t|t}A^T P_{t+1|t} \right)^{-1}$

$\quad\quad \hat{x}_{t|T} = \hat{x}_{t|t} + L_t \left(\hat{x}_{t+1|T} - \hat{x}_{t+1|t} \right)$

$\quad\quad P_{t|T} = P_{t|t} + L_t \left(P_{t+1|T} - P_{t+1|t} \right) L_t^T$

$\quad$ **endfor**

$\quad \hat{y}_t = C\hat{x}_{t|T}$

$\quad t = T$

$\quad \hat{x}_{t+1|t} = A\hat{x}_{t|t}$

$\quad P_{t+1|t} = AP_{t|t}A^T + Q$

$\quad K_{t+1} = P_{t+1|t}C^T \left(CP_{t+1|t}C^T + R \right)^{-1}$

$\quad \hat{x}_{t+1|t+1} = \hat{x}_{t+1|t} + K_{t+1} \left(y_{t+1} - C\hat{x}_{t+1|t} \right)$

$\quad P_{t+1|t+1} = P_{t+1|t} - K_{t+1}CP_{t+1|t}$

endfor

online, smoothing, since it requires the feature values of measurements, can be done only with time delay during the background subtraction application. An example for a signal subjected to Kalman filtering and its smooth version is given in Fig 3. Application of Kalman Filter as a smoothing process involves two steps up to the method which is also known as Rauch-Tung-Striebel Smoother [36]. Forward pass step is a Kalman filter which makes estimation filter. Smoothing process on the other hand is made in backward pass step. In the original case, the method requires all data which is going to be smoothed. However this case is not suitable for background subtraction. Instead of all data, processing exactly s many feature data, although it is not completely real time, the method can be applied with time delay. The recommended time delayed Kalman smoothing algorithm is as in Table.2.

The subscript t in the algorithm represents time variable, $\hat{x}$ represents position vector, A denotes the state transition matrix which represents the signal dynamics, P represents the error covariance matrix, Q represents the covariance matrix for model error K represents the Kalman gain value, C stands for state-to-measure matrix which translates state vector to measurements, R represents the covariance matrix of measure error, y denotes the value of the measurement, and L denotes the gain coefficient of smoothing.

The initial values for $\hat{x}_{0|0}$, $P_{0|0}$, Q, R in the method should be given. Also, matrices for constants A and H, related to the system model and the time delay s should be defined initially.

The smooth measurement values the method has produced are denoted by $\hat{y}$. Since each channel of the color space changes in different rates with respect to time, for each channel of each pixel a different Kalman filter is applied.

4 Performance Measure

In this study, the performance measurements are performed pixel-based. After the background/foreground segmentations of all pixels of all frames in the dataset are made, they are compared with ground truth segmentation results. Receiver Operator Characteristic (ROC) curves are commonly used to present results for binary decision problems in segmentation. However, when dealing with highly skewed datasets, Precision-Recall (PR) curves give a more informative picture of an algorithm's performance [37]. For that reason, area under the PR curve is determined for performance metrics.

Suppose we have a large collection of pixels, C, of which only a fraction π $(\pi/C << 1)$ is foreground. An algorithm detects foreground pixel respect to t [0 1] from C, of which $h(t)$ [0 t] is later confirmed to be foreground. The function $h(t)$ is called a hit curve. Two numeric performance measures often considered are the recall, defined as the probability of detecting an item given that it is foreground, and the precision, defined as the probability that an item is foreground given that it is detected by the algorithm. At the particular detection level of t, the recall and precision are simply in (9).

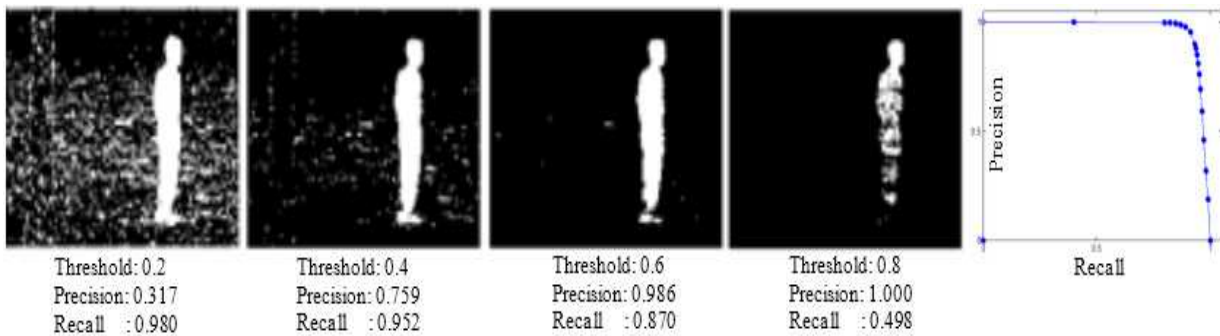$$r(t) = \frac{h(t)}{\pi} \quad p(t) = \frac{h(t)}{t} \qquad (9)$$

Fig. 4: Foreground detection results of a sample test image obtained under different variances. Changes in Precison and Recall rates with respect to variance and the PR graph.

The PR curve realized by the $\{r(t), p(t)\}$ points in the two dimensional space. Therefore, the area below curve can be defined as average value of prediction over the entire interval from $r = 0$ to $r = 1$ as shown (10).

$$\bar{p} = \int_0^1 p(r)dr \qquad (10)$$

Fig.4 shows segmentation results for the test image of a video and the PR curve drawn with respect to these results.

5 Application and Results

In the scope of the research, among the known features that used GMM in literature, tested against the proposed method. These features are gray level, RGB, HSV, YCbCr, Lab and Lab2000HL without any smoothing. In order to increase the success rate in a generic background subtraction, some preprocessing applications such as noise removal and some post processing applications such as blob analysis and shadow detection are used. However, in this application in order to determine the contribution of the proposed method, except the required filters in the method, neither a preprocessing nor a post processing process is applied.

There exist four parameters to be fixed during the background subtraction method introduced in Section.2. The number of mixture models is set 5, learning coefficient is set 0.001, and the background threshold rate is set 0.6. The default value of the class variance required for the model is set by applying the method to 16 different monotone increasing values separately. Therefore, the precision and recall values of each variance are obtained. The change in the obtained precision values with respect to the recall values form the PR graph. The area under the curve is calculated with numerical integration. These calculations are shown schematically in Fig.4. The time delay frame number of the temporal filter is set 10. The matrices representing the signal dynamics belonging to the system model and the initial values of the parameters required for the smoothing process are set as follows;

$$A = \begin{bmatrix} 1 & 0.1 \\ 0 & 1 \end{bmatrix} \quad C = \begin{bmatrix} 1 & 0 \end{bmatrix} \quad Q = \begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix}$$
$$R = 10 \qquad \hat{x}_{0|0} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad P_{0|0} = \begin{bmatrix} 50 & 0 \\ 0 & 50 \end{bmatrix}$$

5.1 I2R Dataset

There exist 9 different videos in the I2R dataset to which tests are applied [38]. The names of these videos are WaterSurface, ShoppingMall, Bootstrap, Campus, Curtain, Escalator, Fountain, Hall, and Lobby respectively. In this dataset, there are foreground images of 20 images for each video, which are manually labeled.

In the WaterSurface video, there is one person as a foreground object and this object holds 8% of the whole image. There is a wavy sea as a non-stationary background and the tree in the image is also slowly but blowing in the wind. Using 20 ground truth images existing in this video, the evaluations of the area under the PR curve indicate that 0.901 success rate is obtained when RGB color space is used, 0.979 when Lab2000HL color space is used without any smoothing process applied, 0,988 when Lab2000HL is used with the proposed method. The PR graph of the video is in Fig.5.

Another video in the dataset is the video called Campus. In this video, there are people and vehicles passing in front of trees swaying strongly in the wind. The foreground objects occupy between 1% and 6% of the whole image.

The results belonging to this video are in Fig.6. The area under the PR curve becomes 0.368 when only RGB color space is used, 0.750 when Lab2000HL is used, and 0.811 when proposed method is applied. The whole results for other videos in the dataset together with these are stated in Table.3.

5.2 FSM Dataset

FSM dataset, which is for the first time introduced in this article, consists of traffic surveillance camera images

Table 3: Area of under PR curve for all video

Methods	I2R1	I2R2	I2R3	I2R4	I2R5	I2R6	I2R7	I2R8	I2R9	FSM1	FSM2	FSM3	FSM4	FSM5
Gray Level	.906	.666	.643	.343	.853	.589	.605	.783	.222	.763	.772	.698	.641	.630
RGB	.901	.708	.700	.368	.861	.600	.614	.796	.237	.792	.810	.711	.668	.726
HSV	.926	.764	.757	.638	.650	.340	.551	.651	.397	.820	.840	.483	.684	.665
YCbCr	.914	.723	.706	.371	.863	.613	.614	.799	.243	.796	.814	.706	.674	.725
Lab	.923	.732	.714	.381	.858	.615	.630	.798	.245	.790	.803	.704	.664	.725
Lab2000HL	.979	.842	.744	.750	.797	.584	.584	.824	.376	.852	.837	.673	.680	.712
Proposed	.988	.863	.756	.811	.864	.648	.648	.842	.438	.874	.890	.735	.719	.757
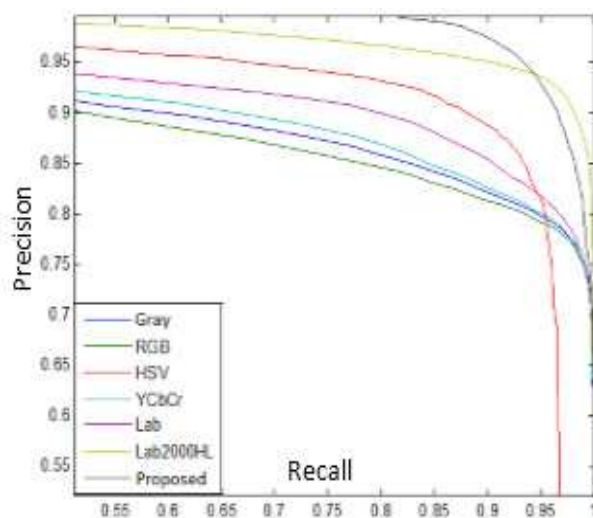
Fig. 5: PR Curve of WaterSurface Video in I2R Dataset.
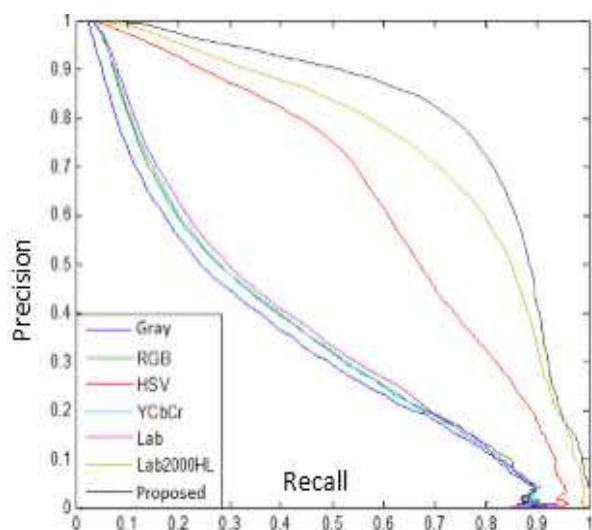
Fig. 6: PR Curve of Campus Video in I2R Dataset.

including different environments for evaluation of background subtraction methods. This dataset is obtained from cameras placed at Etiler entrance of FSM Bridge in Istanbul within the project in [39]. Including totally 5 different videos, this dataset has 300 images per each video. Ground truth images are obtained from 10 images per each video by labeling them manually. A mask which determines the region to be analyzed (region of interest) is prepared and added to the dataset. The images obtained from the videos are saved in .jpg format and they are in 576x720 resolution. The image saving frequency is 25 fps. There exist camera oscillations, shadows, reflections, sudden and gradual light changes at high rates.

The images of FSM1 video are captured in a bright whether at an early morning time. There is not much traffic. The light amount is less and therefore there is less shadow. FSM2 videos on the other hand are captured during a bright day and belong to a lighter environment. Traffic and shadow are a little more than the previous videos. FSM3 is also captured in outdoors but belongs to an environment with so much sun light. The vehicle shadows in this video are so long because of horizontal sun light and almost equal to the area of vehicles themselves. The video FSM4 is on the other hand, captured during morning time in a rainy day. In this video, there is much vehicle traffic but since it is raining the image quality is bad. Wet ground also makes vehicle images reflect. Vehicle reflection from the ground is labeled as the background when the ground truth values are being generated. The problem of image reflections is crucial for especially accuracy of real time traffic analysis. However, in this study there is no solution for this problem. FSM5, which is the last video of the dataset, has images of a rainy day at an evening time. There is so much vehicle traffic and because of the rain there is some reflection from the ground and some noise such as disturbance of camera images. Sample images of the videos in the dataset, the ground truth values and the segmentation results of the recommended method are shown in Fig.7. FSM dataset is an open-source dataset which is available online. [1]

Tests applied using the FSM dataset indicate that precision values are not convergent to 1 despite reduced recall values. In FSM1 the precision value converges to 1
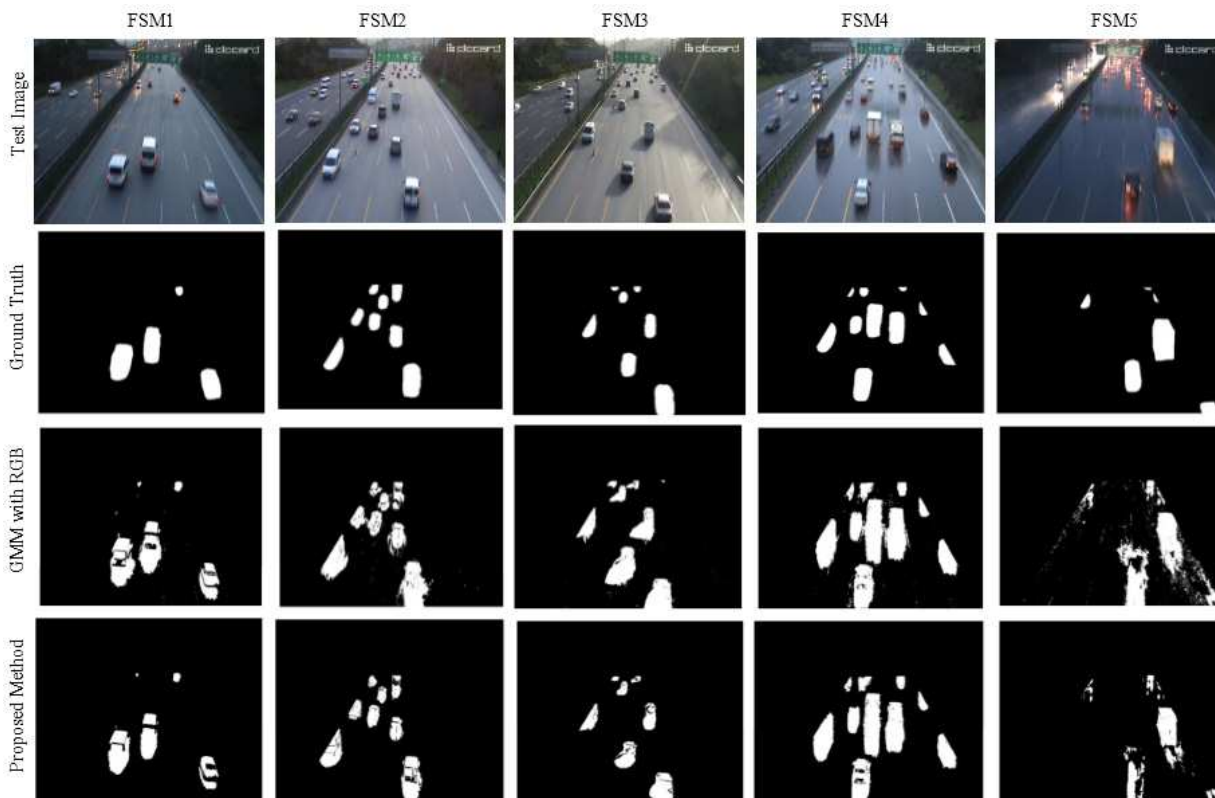

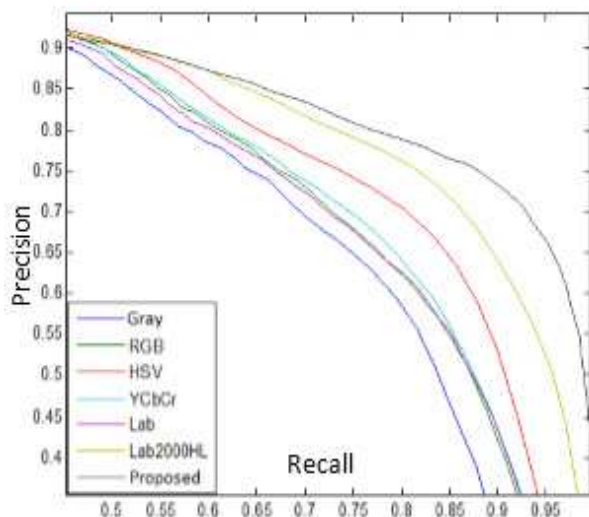**Fig. 7:** Sample image, ground truth and segmentation result of FSM dataset.

Fig. 8: PR Curve of FSM1 video in FSM Dataset.

almost earlier whereas in FSM3 convergence happens quite lately. This is because of the systematical classification errors depending on shadows. In this study shadow detection is disregarded since the study investigates only the performance increase in background subtraction due to using Lab2000HL color space together with the defined smoothing processes.

In FSM1 video, the performance values of RGB, only Lab2000HL, and proposed method are calculated as 0.7299, 0.792, 0.850 and 0.894 respectively. The PR graph obtained from this video is in Fig.8.

In FSM3 video, performance values of RGB, only Lab2000HL and proposed method are obtained as 0.711, 0.673, and 0.735 respectively. The PR graph of this video is also in Fig.9. The areas under the PR curves of both 9 videos of the I2R dataset, and 5 videos of FSM dataset, under 7 different applications are introduced in Table.3.

With respect to the computational costs, the methods are evaluated for images of four different resolutions. In order to perform evaluations, for the features gray level, RGB, HSV, YCbCr, and Lab, color space conversion function of OpenCV library is used. Lab2000HL color space and proposed method are coded in the scope of this study. Computational cost tests are performed over dataset involving 1000 images of the corresponding resolution. Methods are run on a single core i5 processor, 2.67 GHz PC. The results of the tests are given in Table 4.
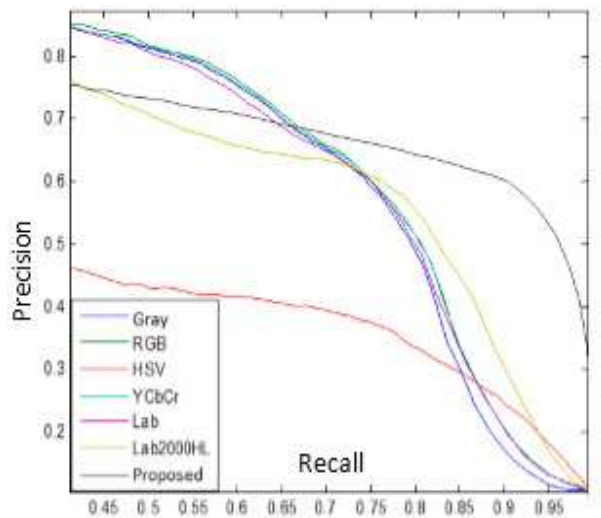
Fig. 9: PR Curve of FSM3 video in FSM Dataset.

Table 4: Average processing times under different resolutions in terms of milliseconds per frame

Feature	128x160	144x176	256x320	576x720
Gray Level	2.2	2.5	7.9	46.0
RGB	4.2	5.3	16.8	84.2
HSV	4.5	5.6	17.6	90.1
YCbCr	4.6	5.6	17.5	90.0
Lab	5.4	6.7	20.2	110.5
Lab2000HL	12.4	16.2	49.4	248.8
Proposed	20.6	27.2	83.2	412.0

6 Conclusion

Lab2000HL color space, having two color axes as much independent as possible of each other, and having color value differences of ideal linearity with respect to human perception, has many advantages and has proved its success in various applications [30, 31, 32, 33]. In this study, in order to increase the success rate, spatial and temporal consistency features are used in background subtraction with GMM. For spatial features, in AC1C2 space 7 different filters are used for each channel. For temporal features, Kalman filtering based Kalman smoothing with time delay which is a different application of Rauch-Tung-Striebel method is applied.

It is shown that the recommended method within tests using I2R and FSM datasets is more successful than the classical GMM using well-known color space. The results indicate that proposed method has increased the performance almost in every sample. If only temporal smoothing been applied, on the other hand, it was more efficient for high performance in samples where objects were large and moving slowly, whereas in samples where foreground objects were small and moving quickly it

affects the performance worse since it blurs the object borders.

Besides performance improving, the calculations of Lab2000HL color space with respect to numerical values with interpolation can be considered to have increased the complexity. In addition to applying convolution process during the spatial filtering, applying different temporal smoother for each color channel of each pixel increases the computational costs significantly. According to Table.4, proposed method needs nearly 0.5 second for processing high resolution image. It means that proposed method do not process high-resolution image in real time at single core processor. It will be possible using special hardware such as GPU or FPGA. However, improves the moving object detection performance at high rates so can be preferred for some applications.

Future studies aim to develop shadow detection application using Lab2000HL color space and to use Lab2000HL color space together with methods such as Kalman filter based background estimation, Kernel Density background estimation, Subspace Learning based background estimation and Fuzzy background estimation.

Acknowledgement

References

[1] S. Cheung and C. Kamath, Robust techniques for background subtraction in urban traffic video in Proc. of Video Comm. and Image Proc., SPIE Electronic Imaging, (2004).

[2] T. Bouwmans, F. El Baf, B. Vachon: Background Modeling using Mixture of Gaussians for Foreground Detection - A Survey, Recent Patents on Computer Science, **1**, 219-237 (2008).

[3] B. P. L. Lo and S. A. Velastin, Automatic congestion detection system for underground platforms, Proc. ISIMP2001, 158-161 (2001).

[4] R. Cucchiara, C. Grana, M. Piccardi, and A. Prati,Detecting moving objects, ghosts, and shadows in video streams, IEEE Tram on Paftem Anal. and Machine Infell, **25**, 1337-1442 (2003).

[5] C. R. Wren, A. Azarbayejani, T. Darrell, and A. P. Pentland, Pfinder: Real-Time Tracking of the Human Body, IEEE Trans. Pattern Analysis and Machine Intelligence, **19**, 780-785, (1997).

[6] C. Stauffer and W. E. L. Grimson, Adaptive Background Mixture Models for Real-Time Tracking, Proc. IEEE CS Conf. Computer Vision and Pattern Recognition, **2**, 246-252 (1999).

[7] Zivkovic Z. Improved adaptive Gaussian mixture model for background subtraction, International Conference Pattern Recognition, **2**, 28-31 (2004).

[8] A. Elgammal, R. Duraiswami, D. Harwood, and L.S. Davis, Background and Foreground Modeling Using Nonparametric Kernel Density Estimation for Visual Surveillance, Proc. IEEE, **90**, 1151-1163 (2002).

[9] J. Gu; Z. Liu; Z. Zhang; Novel moving object segmentation algorithm using kernel density estimation and edge information, Journal of Computer-Aided Design and Computer Graphics, **21**, 223-228 (2009).

[10] J. Wu; J. Xia; J. Chen; Z. Cui. Adaptive Detection of Moving Vehicle based on On-line Clustering, Journal of Computers, **6**, 2045-2052 (2011).

[11] M. Benalia.; A. Ait-Aoudia. An improved basic sequential clustering algorithm for background construction and motion detection, International Conference on Image Analysis and Recognition, ICIAR, Aveiro, (2012).

[12] H. Hu; L. Xu; H. Zhao. A Spherical Codebook in YUV Color Space for Moving Object Detection, Sensor Letters, **10**, 177-189 (2012).

[13] T. Zhou, T. R. Godec: Randomized low-rank & sparse matrix decomposition in noisy case, Proceedings of the 28th International Conference on Machine Learning (ICML-11), (2011).

[14] H. Jimenez. Background Subtraction Approach based on Independent Component Analysis, Sensors, **10**, 6092-6114, (2010).

[15] D. Farcas; C. Marghes.; T. Bouwmans. Background Subtraction via Incremental Maximum Margin Criterion: A discriminative approach, Machine Vision and Applications, (2012).

[16] T. Bouwmans; F. El Baf. Modeling of Dynamic Backgrounds by Type-2 Fuzzy Gaussians Mixture Models, MASAUM Journal of Basic and Applied Sciences, **1**, 1-10 (2009).

[17] M. Balcilar, A.C. Sonmez, Region Based Fuzzy Background Subtraction Using Choquet Integral, In Adaptive and Natural Computing Algorithms. Springer Berlin Heidelberg, 287-296 (2013)

[18] M. Sivabalakrishnan; D. Manjula. Adaptive Background subtraction in Dynamic Environments Using Fuzzy Logic, International Journal on Computer Science and Engineering, **2**, 270-273 (2010).

[19] T. Feldmann, Spatio-temporal optimization for foreground/ background segmentation, In Computer VisionACCV 2010 Workshops. Springer Berlin Heidelberg, 113-122 (2011).

[20] D. Pokrajac, L. Latecki, Spatiotemporal Blocks-Based Moving Objects Identification and Tracking, VS-PETS 2003, 70-77 (2003).

[21] S. Wang, T. Su, S. Lai. Detecting moving objects from dynamic background with shadow removal, ICASSP 2011, Prague, (2011).

[22] M. Xu, T. Ellis,Illumination-Invariant Motion Detection Using Colour Mixture Models, British Machine Vision Conference BMVC 2001, 163-172 (2001).

[23] Y. Sun, B. Yuan, Z. Miao, C. Wan, Better Foreground Segmentation for Static Cameras via New Energy Form and Dynamic Graph-cut, In ICPR '06, Vol. 4. IEEE Computer Society, Washington, DC, 49-52 (2006)

[24] W. Wang, R. Wu, Fusion of luma and chroma GMMs for HMM-based object detection, Pacific Rim Symposium on Advances in Image and Video Technology, pages 573-81, Hsinchu, Taiwan, (2006).

[25] S. Yang, C. Hsu, Background Modeling from GMM Likelihood Combined with Spatial and Color Coherency, ICIP, (2006).

[26] N. Setiawan, S. Hong, J. Kim, C. Lee, Gaussian Mixture Model in Improved HLS Color Space for Human Silhouette Extraction, ICAT 2006, Hangzhou, China, 732-741 (2006).

[27] F. Kristensen, P. Nilsson, V. wall, (Background Segmentation Beyond RGB), ACCV 2006, Hyderabad, Indian, 602-612 (2006).

[28] V. Jain, B. Kimia, J. Mundy: Background Modeling Based on Subpixel Edges, ICIP, 321-324 (2007).

[29] B. Klare, S. Sarka, Background Subtraction in Varying Illuminations Using an Ensemble Based on an Enlarged Feature Set, OTCBVS 2009, Miami, Florida, (2009).

[30] M. Balcilar, A.C. Sonmez, The Effect of Color Space and Block Size on Foreground Detection, SIU2013, Girne, Cybrus, (2013).

[31] M. Balclar, F. Karabiber, A.C. Sonmez, Performance Analysis of Lab2000HL Color Space for Background Subtraction, INISTA, Albena, Bulgaria, (2013).

[32] I. Lissner, P. Urban, How Perceptually Uniform Can a Hue Linear Color Space Be?, 18th Color Imaging Conference Final Program and Proceedings, 97-102 (2010).

[33] I. Lissner, P. Urban, Toward a Unied Color Space for Perception-Based Image Processing, IEEE Transactons On Image Processing, **21**, 1153-1168 (2012).

[34] G.M. Johnson and M.D. Fairchild, A top down description of S- CIELAB and CIEDE2000, Col Res Appl, **28**, 425- 435 (2003).

[35] R. Cheng, J.S. Liu: Mixture Kalman filters, Journal of the Royal Statistical Society: Series B, **62**, 493-508 (2000).

[36] H.E. Rauch, F. Tung, C.T. Striebel, Maximum likelihood estimates of linear dynamic systems, AIAA J, **3**, 1445-1450 (1965).

[37] J. Davis, M. Goadrich, The relationship between Precision- Recall and ROC curves, In Proceedings of the 23rd international conference on Machine learning , 233-240 (2006).

[38] L. Li, W. Huang, I. Y. H. Gu, and Q. Tian, Foreground object detection from videos containing complex background. In ACM International Conference on Multimedia, 210 (2003).

[39] H. B. Celikoglu, H.K. Cigizoglu, G.E. Gurcanli, Hybrid models of neural network method for road safety regulations: Safety index calibration and Intelligent Transportation Systems based safety control (in Turkish). Final report of the project with no. 108M299 supported by the Scientific and Technological Research Council of Turkey (TUBITAK). TUBITAK, Ankara, (2010).

1766

M. Balcilar et. al. : Moving Object Detection using Lab2000HL Color...

Muhammet Balcilar graduated from Yildiz Technical University Computer Engineering Department, in 2004, and finished M.Sc. and Ph.D. in the same department in 2007 and 2013, respectively. He is curently working in the same department as a research assistant. He is a member of Stochastic Robotic Research Group and Computational Intelligence Research Group. Mathematical Modelling, Signal & Video Processing, Robotic and Optimization are his main research areas.

M Fatih Amasyali is Assistant Professor of Computer Engineering Department at Yildiz Technical University, Istanbul. He received the PhD degree in Computer Science at the same university. He is a member of Stochastic Robotic Research Group and Linguistic Research Group. His main research interests are: Natural Language Processing, Expert Systems, Artificial Intelligence and Robotics.

A Coskun Sonmez is Professor of Computer Engineering Department at Istanbul Technical University, Istanbul. He received the PhD degree in Computer Science at Cambridge University (UK). His main research interests are: Real Time Computers, Microprocessors, Artificial Inteligence, Expert Systems and Robotics.