2005 WSEAS Int. Conf. on DYNAMICAL SYSTEMS and CONTROL, Venice, Italy, November 2-4, 2005 (pp391-397)

Fast and Accurate Background Subtraction for Video Surveillance, Using an Adaptive Mode -Tracking Algorithm

CODRUT IANAŞI, VASILE GUI, FLORIN ALEXA, CORNELIU TOMA
Department of Electronics and Telecommunications
„Politehnica" University Timisoara, Romania
Bd. V. Parvan No. 2, 30223 Timisoara, Romania

Abstract: − Background estimation and subtraction is a critical and time consuming step in moving object segmentation for video surveillance. Nonparametric kernel density estimation has been successfully used in modeling the background statistics, due to its capability to perform well without making any assumption about the form of the underlying distributions. To obtain real-time performance of the nonparametric estimator, we recently proposed an algorithm based on mean shift mode-tracking and a rough histogram test to fast discard foreground pixels from exact evaluation. In the present work, an improvement of the new algorithm is proposed, leading to faster background change tracking capability and more accurate background estimation.

Keywords — Background subtraction, nonparametric kernel density estimation, video surveillance.

1 Introduction

Video surveillance is a fast growing field with numerous applications including car and pedestrian traffic monitoring, human activity surveillance for unusual activity detection, people counting etc. [1], [2], [3]. Activity is usually associated with motion and motion is related to image change. However, image changes obtained from simple frame differencing are not able to accurately segment the interior of smooth regions, when they move. Since the cameras used in surveillance are typically stationary, a straightforward way to detect moving regions/objects is to compare each new frame with a reference frame, modeling in some optimal sense the scene background. By subtracting the background from the current frame in all regions where the current frame matches the reference frame, a binary segmentation of the foreground/background pixels can be obtained. Despite the extensive research done, background detection remains a challenging problem in applications with difficult circumstances, such as changing illumination, waving trees, water, video displays, rotating fans, moving shadows, inter-reflections, camouflage high traffic etc.

Background modeling is commonly carried out at pixel level. Each pixel is represented by a feature vector, such as intensity or color, disparity, depth etc. The background estimation process has to be done during activity in the scene and has to be updated to follow background changes occurring in time. Moving objects produce samples considerably deviating from the real background. Therefore, background estimation requires robust estimators [4]. A suitable way to model the static background is through a random vector with an associated probability density function (PDF). The estimated background at a pixel is then the feature vector maximizing the estimated PDF. In some cases, like trees waving in the background or a rotating fan, more than just one density mode may be needed for proper background modeling.

The unknown probability density functions can be modeled parametrically, using known statistical distributions. A very popular approach is to fit the real data with per-pixel mixtures of Gaussians [5][6][7]. The strong point of the Gaussian mixture model is that it can work without having to store an important set of input data, as nonparametric methods do. Some known problems with this approach are: the need for good initializations, slow recovery from failures, difficult adaptation to fast illumination changes, dependence of the results on the true distribution law and the need to specify the number of Gaussians to be fitted.

Alternately, the density function modeling the background at each pixel can be obtained through nonparametric kernel density estimation methods [8]. They are known to be able to produce smooth, continuous, differentiable and accurate estimates, without having to assume any particular underlying distribution. The number of modes does not have to be known in advance and adaptation to new data is automatic. Nonparametric methods are less frequently used in visual surveillance applications than the parametric ones, because of the requirement to store a big amount of data samples for estimation and mainly because of their heavier computational load. Several methods have been proposed to reduce the computational burden of these methods. In [9], Girolami is using a reduced data set obtained by an optimized condensation algorithm. The Fast Gauss Transform, data clustering and clever data structures are used in [10],[11] to reduce the computational load to O(2*N*). In [12], we used a recursive implementation of the mean shift algorithm for very fast real-time nonparametric PDF mode tracking, with complexity O(N^0). This work is an extension of [12], aiming to adapt the learning rate of the mode tracker to the speed of change of the background.

2 Kernel based density estimation techniques for background modeling

Given a sample of *N* *d*-dimensional data points, $\mathbf{x}_i$, drawn from a distribution with multivariate probability density function $p(\mathbf{x})$, an estimate of this density at $\mathbf{x}$ can be written as [8]:

$$\hat{p}_{\mathbf{H}}(\mathbf{x}) = \frac{1}{N}\sum_{i=1}^{N} K_{\mathbf{H}}(\mathbf{x}-\mathbf{x}_i) \tag{1}$$

where

$$K_{\mathbf{H}}(\mathbf{x}) = |\mathbf{H}|^{-1/2}\, K_{\mathbf{H}}\left(\mathbf{H}^{-1/2}\mathbf{x}\right) \tag{2}$$

is the kernel function depending on a symmetric positive definite $d\times d$ matrix, called bandwidth matrix. The bandwidth matrix $\mathbf{H}$ may be fixed or may change with both the estimation point $\mathbf{x}$ ($\mathbf{H}=\mathbf{H}(\mathbf{x})$, the so called *balloon estimator*) and with the sample point $\mathbf{x}_i$ ($\mathbf{H}=\mathbf{H}(\mathbf{x}_i)$, the so called *sample-point estimator*). Frequently $\mathbf{H}$ has a diagonal form or even the form $\mathbf{H}= h^2\mathbf{I}$, assuming the same scale h for all dimensions, i.e. a single scale parameter and an isotropic estimator, K_h. A radially symmetric estimator can be generated starting from a 1D kernel function K_1 as:

$$K^R(\mathbf{x}) = \alpha K_1(\|\mathbf{x}\|), \tag{3}$$

with α is a strictly pozitive constant chosen such that the kernel function integrates strictly to 1. Notice that $K_1()$ actually has a scalar argument. The profile of the radially symmetric kernel is defined as:

$$K^R(\mathbf{x}) = c_{k,d}\, k(\|\mathbf{x}\|^2), \tag{4}$$

with $c_{k,d}$ a normalization constant. Common choices for the kernel profile are the rectangular shape, the triangular shape (for the Epanechnikov kernel) and the exponential shape (for the Gaussian or normal kernel):

$$k(x) = rect(x) = \begin{cases} 1, & |x| \le \frac{1}{2} \\ 0, & otherwise \end{cases} \tag{5}$$

$$k_E(x) = \begin{cases} 1-x, & 0 \le x \le 1 \\ 0, & x > 1 \end{cases} \tag{6}$$

$$k_N(x) = \exp(-\frac{1}{2}x), \quad x \ge 0. \tag{7}$$

An efficient way to find local maxima of the estimated PDF is through the mean shift algorithm [13]. Given the PDF estimated with the radially symmetric kernel *K* with profile *k*,

$$\hat{p}_{h,K}(\mathbf{x}) = \frac{c_{k,d}}{nh^d}\sum_{i=1}^{n} k\left(\left\|\frac{\mathbf{x}-\mathbf{x}_i}{h}\right\|^2\right), \tag{8}$$

the mean shift vector is proportional to the normalized gradient of the estimated PDF:

$$\mathbf{m}_{h,G}(\mathbf{x}) = \frac{1}{2}h^2 c\frac{\hat{\nabla}p_{h,K}(\mathbf{x})}{\hat{p}_{h,G}(\mathbf{x})}, \tag{9}$$

$$\mathbf{m}_{h,G}(\mathbf{x}) = \frac{\sum_{i=1}^{n}\mathbf{x}_i\, g\left(\left\|\frac{\mathbf{x}-\mathbf{x}_i}{h}\right\|^2\right)}{\sum_{i=1}^{n} g\left(\left\|\frac{\mathbf{x}-\mathbf{x}_i}{h}\right\|^2\right)} - \mathbf{x}, \tag{10}$$

$$g(x) = -k'(x) \tag{11}$$

The mean shift vector points into the direction of the maximum increase of the PDF, estimated with kernel *K*. Note that normalization is made with respect to the PDF estimated with kernel *G*. The profile of *G* is *g*, the negative of the first derivative of the profile *k*. Using the mean shift vector at a location $\mathbf{y}$, a gradient ascent algorithm can be used to find the location of the maxima of the estimated PDF closest to the starting location. This can be simply done by iterating the equation

$$\mathbf{y}_{j+1} = \frac{\sum_{i=1}^{n} \mathbf{x}_i g\left(\left\| \frac{\mathbf{y}_j - \mathbf{x}_i}{h} \right\|^2\right)}{\sum_{i=1}^{n} g\left(\left\| \frac{\mathbf{y}_j - \mathbf{x}_i}{h} \right\|^2\right)}, \quad j = 1,2,... \qquad (12)$$

until convergence. The proof of the convergence can be found in [14]. More, in practice the convergence is very fast, typically only two or three iterations being needed.

A straightforward application of kernel density estimators for background model estimation implies evaluating equation (1) at each of the N sample points of a frame buffer. This leads to a total number of N^2 operations per pixel. To obtain reliable estimates, N has to be fairly large, of the order of several hundreds, making the estimation prohibitive for real-time operation with today's ordinary processing systems. In [12], we proposed a much faster solution, based on the idea of mode tracking. Note that incremental density estimation has been also adopted recently for fast PDF estimation using a condensation technique for object tracking [14], while our method is theoretically founded on the mean shift density mode estimator and is applied to background modeling.

3 Adaptive learning rate mode-tracking background estimator

3.1 Method definition

We divide our background subtraction task into two stages. The first one is the initial background estimation, while the second one is background tracking. Initial background estimation is carried out when the system is started, according to equation (1). After doing so a mode-tracking approach is used for background model updating, based on recursive data processing and simple heuristics. We exploit the fact that we are only interested in accurately detecting the mode(s) of the PDF and their locations, not the whole PDF. For simplicity, we describe the case of a background modeled by a single feature vector, corresponding to the highest mode of the PDF.

When a new frame is received, a new data point replaces the oldest data point, at each pixel in a frame buffer of length N. For the unchanged $N - 1$ data points, the new densities can be obtained from the old ones recursively, by simply adding the contributions of the new pixel and subtracting the contributions of the old, outgoing pixel:

$$\hat{p}_{new}(\mathbf{x}) = \hat{p}_{old}(\mathbf{x}) + \frac{1}{N} K_h(\mathbf{x} - \mathbf{x}_{new}) - \frac{1}{N} K_h(\mathbf{x} - \mathbf{x}_{old}) \qquad (13)$$

This means only two operations per data point. For data points not belonging to the background distribution, the density is much lower than for the background. An accurate evaluation of the PDF at such points would be a waist of time, once they were identified. A cheap solution to identify such low density points and quickly discard them from exact evaluation is to keep a low resolution histogram for each spatial location of the image frame. Histogram updating can be done with only one increment and one decrement operation per new estimation frame. Low resolution is beneficial for both dealing with data sparseness and memory considerations. In our work, we used a 16×16×16 color histogram. If the newly entered data point is within the active domain of the kernel function centered at the currently estimated background, the density and the location of the background mode are updated in the next code line. Otherwise, the histogram based density estimate at the new data point is checked against a threshold. Only if the density threshold is passed, the new data point is submitted to accurate density evaluation by equation (1) and the result is compared to the current density maxima of the background for possible model replacement. Our experiments with several surveillance sequences have shown that such events occur very infrequently, typically when an object is permanently moved to or removed from the background. The background density is updated from equation (13), while the background model is updated using the following rule:

$$\mathbf{b}_{new} = (1 - \alpha)\mathbf{b}_{old} + \alpha \mathbf{x}_{new} . \qquad (14)$$

This rule has been widely used for mean updating in the Gaussian mixture model parametric approach. Our theoretical motivation behind this option is related to the mean shift paradigm. Suppose the mean shift algorithm converged to $\mathbf{b}_{old}$ and a new sample is acquired in the data buffer. Starting from $\mathbf{b}_{old}$ the mean shift iterations using equation (12), and denoting

$$\hat{p}_G(\mathbf{b}_{old}) = \frac{1}{N} \sum_{i=1}^{N} g\left(\left\| \frac{\mathbf{b}_{old} - \mathbf{x}_i}{h} \right\|^2\right), \qquad (15)$$

after the first iteration we get the result from equation (14), if we make

$$\alpha = \frac{g\left(\left\|\frac{\mathbf{b}_{old} - \mathbf{x}_{new}}{h}\right\|^2\right)}{N\hat{p}_G(\mathbf{b}_{old}) + g\left(\left\|\frac{\mathbf{b}_{old} - \mathbf{x}_i}{h}\right\|^2\right)}. \quad (16)$$

The second term at the denominator can be neglected, since the estimated probability density of the old background in the tracking mode is supposed to be high. The factor α can be thought of as a learning rate, decreasing with the distance of the new sample to the currently estimated background model and with the estimated probability density of the background model. In our implementation, we used a fixed value for the denominator in equation (16). This was mainly motivated not by faster processing reasons, but by the observation that variable denominator in equation (16) results in higher learning rate in regions with more activity, where the background is obscured longer, which is not desirable.

A very stable background estimator needs a big amount of samples. In the case of the tracking estimator, a big N leads to a low learning rate, α. A low learning rate reduces the noise effects on the tracker. However, a lower learning rate also results in slower adaptation of the background model to real changes, like those produced by illumination changes. Apparently, the parameters N and α have to be selected as a compromise between two factors: low error variance in static conditions and low error variance in dynamic situations. The solution we propose in order to break this dilemma is based on the observation that the *learning rate* can changed *adaptively* to cope with both situations.

A good adaptation method has to be able to discriminate between background changes produced by noise and real background changes. The basic idea of the improved tracking algorithm is that real background changes cause fast increasing of the *cumulative difference* $\mathbf{d}_{cum}$ between the estimate $\mathbf{b}$ and the incoming data samples $\mathbf{x}_{new}$, while noise effects on the cumulative difference tend to cancel. If we denote the background estimate at discrete time t with $\mathbf{b}(t)$, then the cumulative error vector at time t is

$$\mathbf{d}_{cum}(t) = \sum_{i=0}^{t} [\mathbf{x}(i) - \mathbf{b}(i)] \ . \quad (17)$$

By comparing the norm of the cumulative difference vector to a threshold $\mathbf{d}_{th}$, we can effectively detect situations when the mode-tracking estimator cannot keep up with the speed of change of the background. When we detect such an event, we simply update the background with the current sample by using learning rate α=1, set to zero the cumulative error signal and resume tracking with the basic learning rate. A pseudo-code description of the adaptive mode tracking background estimator is given in figure 1.

```
if( K_h(x_new — b) != 0 )
        update( b and p̂(b) );
        d_cum = d_cum + x_new — b;
        if( ||d_cum || > d_th )
                b =: x_new;
                d_cum = 0;
                end if
else if( (Hist(x_new) > threshold )
        and ( p̂(x_new) > p̂(b) )
        b =: x_new;
        end else if
end if
```

Fig. 1 . Pseudo-code of the adaptive mode tracking background estimator.

3.2 Comparisons with known adaptive background estimation solutions

Adaptive learning rate has already been used for *parametric* background estimation by several authors. In [15], the learning rate is the product of two functions. The first one is a function of the local confidence, defined as $\exp(-d^2/2\sigma^2)$, where d is the difference between the current sample and the estimated mean. While this factor can be viewed as a generalization of the standard on-line version of the EM algorithm, its presence in our basic (called non-adaptive) mode tracker is theoretically derived from the mean shift estimation paradigm. The second factor of the product used to compute the learning rate in [15] is a function of the global correlation, aiming to detect camera rotation, not considered here. In [6], variable learning rate is made a function of the scene

activity at each pixel, detected by a different processing module, in order to slow down the learning process in regions with high activity, when the real background is obscured by foreground objects. As a result, two people who stopped to have a short conversation would be incorporated in the background slower, as they have a slight motion, detected as an activity by the corresponding processing module. Conversely, a chair moved to the background would be static and therefore faster incorporated into the new background. The idea can be straightforwardly implemented in our nonparametric estimation too. However we use variable learning rate to enhance the tracking speed of our recursive estimator when the real background is visible but changed. Such an event is happening for example when the illumination changes gradually or by a moderate sized step. This happens frequently in outdoor scenes as a result of moving clouds, or in indoor scenes when an additional bulb is switched on or off or when a moving object is partially obscuring the light from some sources. The stated goal leads to a different learning rate adaptation method, complementing the parametric adaptive background estimation solutions discussed above. We include here the possibility of incorporating the basic idea of the adaptive mode tracking estimator from this work into parametric background estimators.

Adaptivity in nonparametric background estimation and subtraction methods is mostly related to finding the appropriate bandwidth or scale parameter for the estimator and an adequate threshold on the estimated PDF for segmenting the foreground objects. Although not described here (the interested reader is referred to our previous paper [12]), in practical implementation we used the adaptive scale based on median of absolute frame differences previously adopted by Elgamal [16]. More recently, adaptive nonparametric kernel density estimation for background subtraction is described in [17].

3.3 Method evaluation

The performances of the basic mode tracking detector have been evaluated in [12]. It has been shown that the mode tracking estimator has significantly lower error variance than the traditional kernel based estimator for a wide range of scales. To assess the performances of the adaptive learning rate mode-tracking estimator, we carried out tests with both static and dynamic background. As theoretically expected, static tests revealed asymptotically identical results. Therefore, we next report only the results of tests with dynamic background.

In our first experiment, we generated a 1D unit step edge of 400 samples. White Gaussian noise with 10% standard deviation was added to obtain a noisy step edge. The signal was tracked with the mean shift mode-tracking estimator and with the adaptive learning rate mean shift mode-tracking estimator. Both estimators used truncated Gaussian kernels with scale parameter $h = 2\sigma^2 = 1$ and $\alpha = 0.02$ in update equation (16). The threshold on the accumulative difference was set at the level 3h = 3. The results of one such experiment are plotted in figure 2. Only a few samples are needed for the fast tracker to switch to fast tracking after the edge. Theoretically, there should be three samples, in the absence of noise.

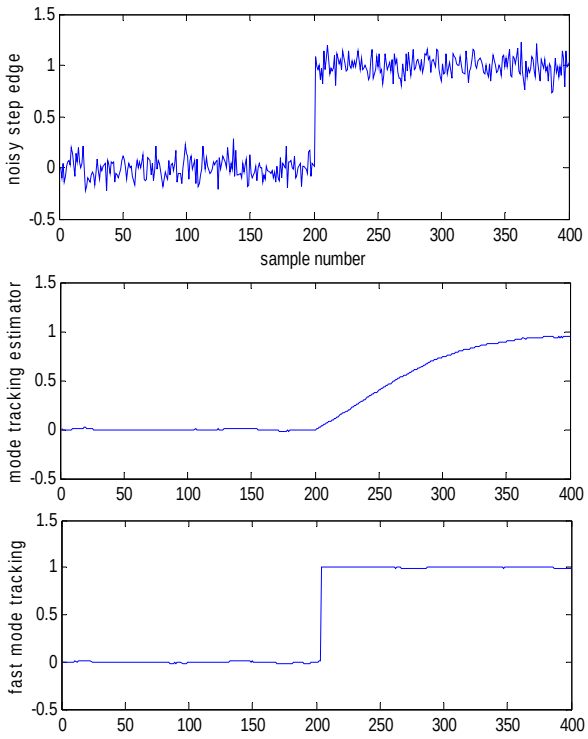

Fig. 2. Noisy step edge response of the mode-tracking estimator and of the adaptive mode tracking estimator.

In the second experiment, we tested the standard deviation of the estimation error of the two mode-tracking estimators as before, as a function of the amplitude of the noisy step edge. The signal was corrupted with Gaussian noise of standard deviation

10% and then 20%. The results are summarized in figure 3. As theoretically expected, the adaptive mode-tracking estimator outperforms the non-adaptive estimator when significant background changes occur. For very small changes or constant background and moderate noise level, the estimators have nearly the same (small) error level. The adaptive estimator has slightly higher error at very high noise level but only for static background.

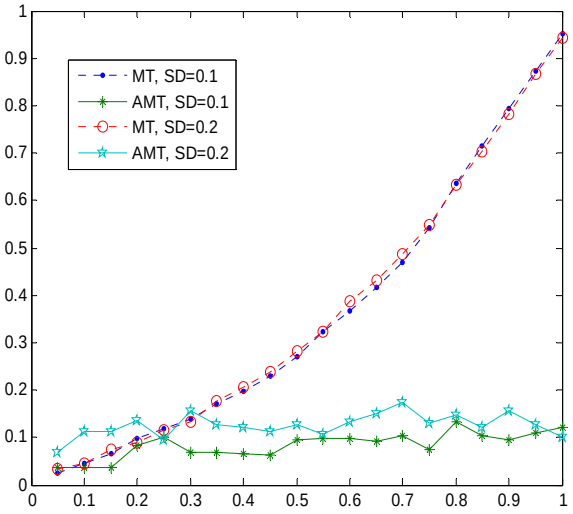

Fig.3. Standard deviation of the estimation error of the mode tracking estimator (MT) and of the adaptive mode tracking (AMT) estimator as a function of the step edge amplitude, for noise standard deviations SD = 0.1 and 0.2.

4 Conclusion

In this work, we proposed an adaptive mode-tracking density estimator for background modeling in video surveillance. While benefiting from the advantages of nonparametric methods, the estimator has a very low computational complexity, as compared to other more general solutions for nonparametric density estimation, like the Fast Gauss Transform. The estimator is able to cope well with fast background change, due, for example, to sudden illumination change. This is obtained by switching the learning rate to the maximal value 1, when significant cumulative error test indicates systematic error build up. Experiments indicate very low estimation error in the presence of noise and dynamic backgrounds. In principle, the adaptive learning rate of the mode-tracking background estimator can be computed in a more general way, to include adaptation criteria already proposed for other (parametric) adaptive estimators, for example detected activity computed from frame differences. Also, the adaptation method proposed here can be incorporated in a straightforward manner in parametric estimation methods. In this sense, the present work is complementing known adaptive background estimation solutions. More, we believe that the proposed solution may be useful in other applications using incremental density estimation as well and that the basic idea can be reshaped to fit specific needs.

References

1. C.R.Wren, A. Azarbayejani, T. Darel and A. Pentland, “Pfinder: Real-time tracking of human body”, *IEEE Trans. Pattern Anal. Machine Intell.* Vol. 19, No. 7, 1997, pp. 780-785.
2. K. Toyama, J. Krumm, B. Brumitt and B. Meyers, “Wallflower: principles and practice of background maintenance”. In *IEEE Conference on Computer Vision*, Kerkyra, Greece, 1999, pp. 255-261.
3. D. Harwood, I. Haritaogly and L.S. Davis, “W^4: Real-time surveillance of people and their activities”. *IEEE Trans. Pattern Anal. Machine Intell*. Vol. 22, No. 8, 2000, pp. 809-830.
4. P. Meer, “Robust techniques for computer vision”, *Emerging Topics in Computer Vision*, G. Medioni and S. B. Kang (Eds.), Prentice Hall, 2004, pp. 107-190.
5. W.E.L. Grimson, C.Stauffer, R. Romano and L. Lee, “Using adaptive tracking to classify and monitor activities in a site”, *IEEE Conference on Computer Vision and Pattern Recognition*, Santa Barbara, CA.,1998, pp. 22-29.
6. M. Harville, G. Gordon and J. Woodfill, “Adaptive video background modeling using color and depth”, *International Conference on Image Processing ICIP 2001*, Tessaloniki, Greece, Vol. 3, 2001, pp90-93.
7. P.Withagen, K. Schutte, F. Groen, “Object detection and tracking using a likelihood

based approach", *Proc. ASCI 2002 Conference*, Lochem, The Netherlands, 2002, pp. 248-253.
8. M.P.Wand and M.C.Jones, "*Kernel Smoothing*", Chapman & Hall, 1995.
9. M. Girolami, C. He, "Probability density estimation from optimally condensed data sets". *IEEE Trans. Pattern Anal. Machine Intell*. Vol. 25, No. 10, 2003, pp. 1253-1264.
10. A.Elgamal, R.Duraiswami, L.S.Davis, "Efficient kernel density estimation using the Fast Gauss Transform with applications to color modeling and tracking", *IEEE Trans. Pattern Anal. Machine Intell*. Vol. 25, No. 11, 2003, pp. 1499-1504.
11. J. Yang, R. Duraiswami, N. Gumerov, L. Davis, "Improved Fast Gauss Transform for efficient kernel density estimation", *IEEE Intl. Conference on Computer Vision, ICCV*, 2003, pp. 464-471.
12. C. Ianăşi, V. Gui, C.I. Toma, D. Pescaru, "A fast algorithm for background tracking in video surveillance using nonparametric kernel density estimation", *Facta Universitatis (Niš)*, Vol. 18, No. 1, 2005, pp 127-144.
13. D.Comaniciu, P.Meer, "Mean shift: A robust approach toward feature space analysis", *IEEE Trans. Pattern Anal. Machine Intell,* Vol. 24, No.5, 2002, pp. 603-619.
14. B. Han, D. Comaniciu, Y. Zhu, L. Davis, "Incremental Density approximation and Kernel-Based Bayesian Filtering for Object Tracking", *IEEE Conf. Computer Vision and Pattern Recognition (CVPR'04)* , Washington DC, 2004.
15. M.Pic, L. Berthouze, T. Kurita, "Adaptive background estimation: Computing a pixel-wise learning rate from local confidence and global correlation values", *IEICE Trans. Inf & Syst.*, Vol. E87-D, No.1, 2004, pp.1-7.
16. A. Elgamal, R. Duraiswami, D. Harwood, L. Davis, "Background and foreground modeling using nonparametric kernel density estimation for visual surveillance", Proceedings of the IEEE, Vol. 90, No.7, 2002, pp 1151-1162.
17. A. Mittal, N. Paragios, "Motion-based background subtraction using adaptive kernel density estimation", *IEEE Conf. Computer Vision and Pattern Recognition (CVPR'04)* , Washington DC, 2004.