A Study of Background Modeling

Ana M. Herrera,José A. Palacios, Hugo Jiménez, Alejandro Romero, Sandra L. Canchola, Carlos A. Olmos

Abstract.

Background modeling is used in different applications to model the background and to detect the moving objects in the scene in video surveillance. Several of background modeling methods have been proposed to identify foreground objects in a video. In this paper we review the background model methods: parametric and non-parametric methods. Parametric methods building the parameter estimation probability distribution based on the color distribution of images. In non-parametric methods, a kernel function is introduced to model the distribution. In this paper, we address the problem of some representative background modeling. Results shown numerous improvements of this some methods developed to background modeling.

Background Subtraction; Optimize; Mixture of Gaussian; Dynamic Adaptation.

I. BACKGROUND MODEL

Background modeling is often used in different applications to model the background and to detect the moving objects in the scene in video surveillance [1],[2]

Image analysis usually involves linear approaches. However, the use of these approaches are not enough robust when the scene has complicated behaviors, such as external factors, which affects common conditions of the stage; just to mention someone we have the light variations [3]. These factors usually have not linear behaviors and common approaches figure out as an approximation to the real model [4]. The main challenge, of those kind of approximations, consists on velocity of moving objects, and luminance perturbation. First ones is related with sampling ratio; i.e. objects to be detected with a common camera, might be slow and distant to the sensors at least be sampled twice times over camera field [5]. But in real application this constraints increased to be almost 10 times sampled [6], avoiding softer images and fuzzy contours of moving objects. In practical situations it means, an object to be measured need to be sampled by the sensors several times; in other situations, they are not detected and changes produced in sensors are miss classify as disturbances on scene. In response to this situation, in the literature has been developed approaches that continuously adapt to changing scenarios [5], [6]. However adaptable approaches need extra knowledge of the scenario to find out the better parameter combination for a particular model [7];

II SUBTRACTION OF THE BACKGROUND

This approximation consists of determining sections of the image that remain constant from which not. Sections that remain constant are called the background; this approximation employs a set of training images$\{I_{n+1}, \ldots\}$, The image is subtracted$I_i - I$, where the result represents the moving sections of the constant sections. The model $\bar{I}$ is updated to better adapt to the atmosphericconditions. As an example, the simplest approximation is described, which consists of the average of the intensities for each pixel that forms the image that is to say with the set of initial images$\{I_{1 \ldots} I_n\}$. The average image I is estimated, denotes the pixel i, j of the image I, we have: $\overline{I =} \sum_{k=1}^{n} I_k(i,j)$for all (i,j). In this way the image formed of the average of each pixel is a good approximation of the parts that remain fixed. The detection of the moving parts and the parts that are background is obtained for the images$k \geq n$, by subtracting the average image I from the imageI_k; where the background is constant for all the last few images, however, there are disturbances due to the acquisition and a threshold λ is considered, so that all the absolute regions of said difference are considered as background, And those older as' tasks in movement.

Inideal situations the background remains constant, but there are situations in which the background may change, some of they are: a) Inclusion / elimination of objects that belong to the fund. b) Luminous climatic changes (intensity of light, dusk, dusk, cloud shadows). c) Luminal projections of objects (reflections and shadows). d) Speed of movement of objects.

It is necessary to continuously modify the background, to adapt to the changes that occur in the scene. These approximations are very useful in constant scenarios, or where it is easy to create a background update model. Particularly it is necessary to continuously modify the background, to adapt to the changes that occur in the scene. These approximations are very useful in constant scenarios, or where it is easy to create a background update model. Variants of this approach have been developed, these variants consist of estimating in different ways. One of the most used variants is [7], where$I$ is modeled at pixel level where each of the pixels is modeled as a random variable with one or multiple modes, so that each $I_{i,j}$ is represented as a mixture

$I_{I,J} = f(\mu_1, \sigma_1) + f(\mu_2, \sigma_2) + \cdots + f(\mu_k, \sigma_k)$ so that they represent the possible values that the pixel can take, Or some object that remains constant for a certain period. The performance of this approximation depends on the reliability of the construction of the Gaussian behavior of the pixel, in this sense, work has been developed that by the application of EM variants of this approach have been developed, thesevariants consist of estimating in different ways.

In this paper, we analyze several approaches in order to explore its strengths and weaknesses in order to create a new robust model of background.

III MIXTURE OF GAUSSIANS

Mixture of Gaussians is an approach for background modeling to detect moving objects. The original method developed by Stauffer and Grimson [8].
Mixture of Gaussian (MOG) approach is a powerful estimation and prediction background subtraction model. Nevertheless, although it has been improved by using several algorithms such as Expectation Maximization (EM); it is still susceptible to sudden changes in light conditions effects.

For dynamically adapting a Mixture of Gaussian, each pixel dynamic is expressed as summation of Gaussian $I(x) = \sum_{i=1}^{n} \mathcal{M}$ such that $\mathcal{M}_i = \alpha_I \alpha_I G_i(\mu_i, \sigma_i)$. Each Gaussian represents a stable pixel colour intensity; where most probable Gaussian represents pixel background intensities. Parameters of model $\mathcal{M}_i$ are expressed as;$\{\mu_i, \sigma_i\}$; which are updated over time. The initial values$\mathcal{M}_i(\mu_0, \sigma_0)$ for the initial set of models $\Phi = \{\mathcal{M}_1, \mathcal{M}_2, \dots \mathcal{M}_n\}$ are fixed with first acquired image. Ratio convergence ρ defines the adaptability and number of frames to reach an stable model. The convergence index takes values from [0,1]which are directly related with the speed of convergence in EM algorithm.

IV MODEL BASED ON ADAPTABLE MOG

In [9] the authors show an algorithm based on a dynamic selection of convergence ratio, which use the expected proportion between movement and fixed zones of scene. This proportion is used as an extra criterion to detect the maximum direction of Entropy in EM algorithm. A criterion basedontheexpectedmotionproportioninscene,provides information about the validity of current ρt. Whenρt results worst enough, it is updated. This change allows to reach faster a stable model expressed by Gaussian parameters $[\mu, \alpha]$ for a particular pixel x. Next, motion detection is performed by building a binary map B(x) mxn, indexed by x for a particular instant t. This map represents pixels which fall into the Gaussian as follows:

$$B(x) = PrI(x) \epsilon G(\mu_i, \alpha_i) \leq \lambda Tn \tag{1}$$

whereP_r means the probability of a measure $I(x)$for a particular pixel x of belongs to the Gaussian (μ_i, α_i) and Th is the confidence degree[9][10].

In this case, that MOG approach converges to the optimum, i.e, the number that has the most likely to be following value (see eq, 2)

$$f_i^*(x \setminus \phi) = maxP(f_i(x \backslash \phi)) \tag{2}$$

Where $f_i^*(x \setminus \phi)$ is the expected value given a $f_i(x \backslash \phi$ and ϕ are the parameters of distribution $f_{i.}$

Since the scheme is interactive, the equation 2, is rewritten as:

$$f_i^*(x \setminus \phi) + H; \qquad 3$$

Where H is an entropy function that usually is defined as $H = (1 - \rho)x$ which indicate the maximum variability of data given evidence x and it usually is represented as the learning adjustment, which depends on the percentage given by ρ. Remember that a Gaussian is givenby ρ. As is appreciated in Fig. 1 the standard MOG (above) began to adapt to changing light conditions, however the difference is very clear, the detection is not good enough since it gives many false positive regarding the optimized MOG (below).

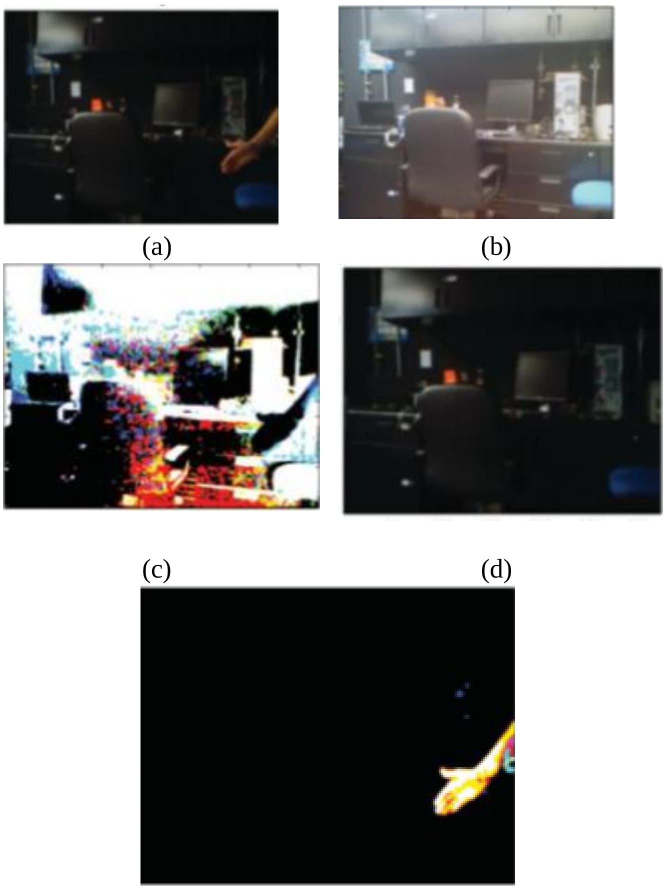

Figure 1. Results and comparison between MoG and optimized in subsequent images.

V .CODEBOOK MODEL

This algorithm adopts a clustering technique, to construct a background model. For each pixel, it builds a codebook, which may contain of one or more codewords. Samples at each pixel are clustered into the set of codewords based on a color distortion metric together with brightness bounds. The clusters represented by codewords do not necessarily correspond to single Gaussian or other parametric distributions. The background is encoded on a pixel-by-pixel basis. Detection involves testing the difference of the current image from the background model with respect to color and brightness differences. If an incoming pixel meets two conditions, it is classified as background (1) the color distortion to some codewords is less than the detection threshold, and (2) its brightness lies within the brightness range of that codeword. Otherwise, it is classified as foreground. [11], [12].

VII. LOCAL BINARY PATTER(LBP)

Local Binary Pattern (LBP) was proposed by M. Heikkila[13], [14]. It is a texture descriptor in gray level images. It generates the texture of one region calculating the threshold difference value between the center pixel and the pixels in its neighboring region. The basic LBP descriptor works in 8 connectivity where it is extended into a region of a circle.

$$LBP_{P,R}(x_{c,}y_c) = \sum_{p=0}^{p-1} s(g_p - g_c)2^p \qquad (4)$$

where$s(x)$ is defined as follows:

$$s(x) = \begin{cases} 1: x \geq 0 \\ 0: x < 0 \end{cases}$$

Whereg_c is the gray value of the center pixel (x_c, y_c), and g_p is the gray value of the pixels on the circle.Particularly, LBP features are very fat to compute in real –time application.

In [13],[14] was proposed a new online dynamic texture extraction operator, named Spatio-temporal Local Binary patterns (STLBP). This method combine spatial texture and temporal motion information together. The dynamic background model of a pixel is built using a group of STLBP histograms.

This modelhave three advantages: 1) it is robust to monotonic gray-scale changes; 2) it is online and very fast to compute, 3) it can extract spatial texture and temporal motion information of a pixel.

(a)MOG (b)MOG
Fig. 2 K Kim et al.(2005)

IV KERNEL DENSITY ESTIMATOR (KDE)

This method was proposed by Elgammal et al. [15]. This introduced Kernel density estimation on N recent sample of intensity values $\{x_{1,}x_{2\ldots,}x_{2\ldots,}\}$ to model the background distribution.
The probability of the intensity of each pixel is defined as follows:

$$P(x_t) = \frac{1}{N}\sum_{i=1}^{N}\prod_{j=1}^{d}\frac{1}{\sqrt{2\pi\sigma_j^2}}e$$

Where N is the number of samples and σ is the kernel function for each color channel.

In [16] a method based on variable bandwidth kernels to determine σ is used.

VIII. PIXEL BASED ADAPTIVE SEGMENTER (PBAS)

Pixel-based adaptive segmenter was introduced by Hofmann et al. [17], [18]It is a non-parametric background model. Every pixel is modeled as follows.

$$B(x_i) = \{B_1(x_1), B_2(x_2) \ldots B_N(x_i)\}$$

A pixel x_i belong to the background if its value $(I(x_i)$ is closer to at least value in terms of a decision threshold $R(x_i)$

IXA UNIVERSAL BACKGROUND SUBTRACTION ALGORITHM (ViBe)

This background subtraction algorithm is based on random substitution and spatial diffusion. Particulary, it uses observed color values of pixels of background training sequences as samples of observed backgrounds. Vibe has only uses color values of pixels to build the background but color values are usually sensitive to noise and illumination changes[19],[20].

Some representative background modeling methods are classified in Table 1.

Table 1. Classification of background modeling methods.

Category	Background modeling methods
Parametric	MOG,
	Model Based on Adaptable MOG
No parametric	Vibe
	CodeBook
	PBAS
	KDE
	LBP

V. CONCLUSION

Background modeling methods should take into accounttheir robustness against illumination changes. In particular,many computer vision systems are used in outdoor scenes.So it is necessary to continuously modify the background, to adapt to the changes that occur in the scene. These approximations are very useful in constant scenarios, or where it is easy to create a background update model.

ACKNOWLEDGMENT

The work was supported by PRODEP (Programa para el DesarrolloProfesionalDocente).

REFERENCES

[1]S. Cheung , C. Kamath. Robust background subtraction with foreground validation for Urban Traffic Video. J Appl Signal Proc, Special Issue on Advances in Intelligent Vision Systems: Methods and Applications (EURASIP 2005), New York, USA, ; 14: 2330-2340. 2005.

[2] Toyama K, Krumm J, Brumitt B, Meyers B. Wallflower: Principles and practice of background maintenance. Int Conf on Computer Vision, (ICCV 1999), Corfu, Greece, September 255 -26,1999.

[3] Stauffer C, Grimson W. Adaptive background mixture models for real-time tracking. Proc IEEE Conf on Comp Vision and Patt Recog (CVPR 1999) 246-252, 1999.

[4] T. B. Moeslund and E. Granum, "A survey of computer vision-based human motion capture," Computer Vision and Image Understanding, vol. 81, no. 3, pp. 231 – 268, 2001.

[5] C. Stauffer and W. Grimson, "Adaptive background mixture models for real-time tracking," in Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on., vol. 2. IEEE, 1999.

[6] H. Nyquist, "Certain topics in telegrph transmission theory," Transsaction of AIEE, vol. 47, no. 2, pp. 617–644, in: Proc. IEEE, 2002.

[7] M. Unser, "Sampling-50 years afer shannon," Proccedings of IEEE, vol. 88, no. 4, pp. 569–587, April 2000.

[8] Z. Wei, "A pixel-wise local information-based background subtraction approach," in Multimedia and Expo, 2008 IEEE International Conference on, June 2009, pp. 1501–1504.

[9] V. Mahadevan, "Background subtraction in highly dynamic scenes," in Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on, 2008, pp. 23–28.

[10] C. Poppe, G. Martens, P. Lambert, and R. Van de Walle, "Improved background mixture models for video surveillance applications," Computer Vision–ACCV 2007, pp. 251–260, 2007.

[11] D.M and Gavrila, "The visual analysis of human movement: A survey," Computer Vision and Image Understanding, vol. 73, no. 1, pp. 82 – 98, 1999.

[12] M. Hofmann, P. Tiefenbacher, G. Rigoll, Background segmentation with feedback: the pixel-based adaptive segmenter, in: Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Workshops, Jun. 2012, pp. 38-43.

[13] M. Heikkila, M. Piettikamen, IEEE Trans. Pattern Anal. Mach. Intell. 28 (4) 657-662, 2006.

[14] T. Ojala, M. Pietikainen, and D. Harwood, "A comparative study of texture measures with classification based on feature distributions," Pattern Recognition, vol. 29, pp. 51–59, 1996.

[15]A. Elgammal, D. Hanvood, L. S. Davids, Non-parametric model for background subtraction in: Proc. ECCV pp. 751-767, 2000.

[16] A. Mittal, N. Paragios, Motion-based background subtraction using adaptive kernel density estimation, in: Proc. Int. Conf. On Computer Vision and Pattern Recognition, IEEE, Piscataway, NJ, pp. 302-309. 2004.

[17] M. Hofmann, P. Tiefenbacher, G. Rigoll, Background segmentation with feedback: the pixel-based adaptive segmenter, in: Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Workshops, pp. 38-43, 2012.

[18] Yong Xu, J. Dong, B, Zhang, D. Xu, Background modeling methods in video analysis: A review and comparative evaluation, CAAI Transaction on Intellilgence Technology, 1, 43-60 , 2016.

[19] O. Barnich, M. Van Droogenbroeck, Vibe: a powerful random technique to estimate the background in video sequences, in: IEEE Int. Conf. on Acoustics, Speech and Signal Processing, pp. 945-948, 2009.

[20] O. Barnich, M. Van Droogenbroeck, IEEE Trans. Image Process. 20 (6), 1709-1724, 2011.

Ana M. Herrera Navarroreceived the B.S. degree in Computer Engineering and her M.S. degree in Computer Science from Informatics Faculty of the Queretaro Autonomous University, Mexico. Her research interests include morphological image processing and computer vision.

Jose Abel Palacios González. Students of M.S. Engineering from Informatics Faculty of the Queretaro Autonomous University, Mexico. His research interests are image processing and computer vision.

Julio Alejandro Romero received the B.S degree in Mechatronics Engineering and her M.S degree in Mechatronics Engineering. His research interests are image processing, automatization and computer vision.

Hugo Jiménez Hernández Computer Science Engineering at the Queretaro Regional Technological Institute; his M.S. in Computer Science at the Center. He received his Ph.D. degree from CICATA Queretaro-IPN.

Carlos Alberto Olmos Trejoreceived the B.S. degree in Software Engineering and her M.S. degree in Software Engineering from Informatics Faculty of the Queretaro Autonomous University, Mexico. His research interests include mathematics, educations.

Sandra Luz Canchola Magdalenoreceived the B.S. degree in Informatics and her M.S. degree in Software Engineering from Informatics Faculty of the Queretaro Autonomous University, Mexico. Her research interests include morphological image processing and computer vision.