A REGION BASED APPROACH TO BACKGROUND MODELING IN A WAVELET MULTI-RESOLUTION FRAMEWORK

Ainhoa Mendizabal, Luis Salgado

Grupo de Tratamiento de Imágenes, Universidad Politécnica de Madrid

ABSTRACT

In the field of detection and monitoring of dynamic objects in quasi-static scenes, background subtraction techniques where background is modeled at pixel-level, although showing very significant limitations, are extensively used. In this work we propose a novel approach to background modeling that operates at region-level in a wavelet based multi-resolution framework. Based on a segmentation of the background, characterization is made for each region independently as a mixture of K Gaussian modes, considering the model of the approximation and detail coefficients at the different wavelet decomposition levels. Background region characterization is updated along time, and the detection of elements of interest is carried out computing the distance between background region models and those of each incoming image in the sequence. The inclusion of the context in the modeling scheme through each region characterization makes the model robust, being able to support not only gradual illumination and long-term changes, but also sudden illumination changes and the presence of strong shadows in the scene.

Index Terms— Region analysis, wavelet transform, Background modeling, Mixture of Gaussians.

1. INTRODUCTION

Many applications such as traffic monitoring or surveillance systems are based on the unsupervised analysis of quasi static video sequences. This analysis is focused mainly in the detection and monitoring of elements of interest appearing in the scene. In this context, background subtraction techniques are extensively used: a description and characterization of the scene in the absence of elements of interest is carried out (background modeling), and objects are detected through the analysis of the scene deviations with respect to this model. As the scene background evolves along time, due to illumination changes or new static objects incorporated or subtracted from the scene, models should be adaptive.

Most background modeling techniques proposed in the literature operate at pixel level: The color of each background pixel is modeled independently. A first approach was proposed in [1], where the intensity value of each background pixel is modeled by a Gaussian distribution. But this model does not support the presence of moving objects in cluttered areas like swaying trees. To overcome this drawback, Stauffer et al. [2], model each pixel value as a mixture of Gaussians, supporting then bimodal background pixels. Non Gaussian probability density function was considered in [3]: Kernel Density Estimation techniques are incorporated to the characterization. Although these strategies base the decision whether one pixel belongs to the background or not on the probability of its color value, other non-parametric methods predict the value of each background pixel based on its recent history values. If a new value differs enough from its prediction, it is considered as a foreground pixel. In [4] the prediction of a pixel value is obtained as a weighted sum of its *p* previous values; [5] proposed a cluster scheme where each cluster has its representative value and a weighting factor that represents its similarity to the background; and in [6] a codebook is constructed for each background pixel that registers each pixel value appearance frequency.

Although all these pixel-based models support gradual and long-term illumination changes, they underperform in the presence of sudden ones. To reduce their impact, some authors incorporate the idea of context in the characterization, introducing the neighborhood in the analysis. For example, in [7] it is used not only the intensity value of the pixel but also its gradient magnitude. Sudden illumination changes alter meaningfully the intensity but not the gradient information, fact that makes the characterization stronger. In general, border information is less sensitive to sudden illumination changes. In this sense, [8] and [9] propose using this information in the characterization, and detect the changes in a "map of borders". In [10], based on Pattern Recognition techniques, a vector of characteristics using Binary Local Patterns is obtained for each pixel; and in [11] the changes on *NxN* pixel blocks are analyzed, based on the idea that neighborhood pixels change similarly throughout time.

Nevertheless, the pixel oriented model of all these methods bound the usefulness of context information. Attempts to tackle sudden illumination changes obtain limited results, and no solutions are given to the presence of shadows. As a result, accuracy in the detection suffers from model instabilities, thus requiring ad-hoc post-processing strategies.

To overcome the abovementioned limitations, in this paper we present a novel approach to background modeling that operates at region-level in a wavelet based multi-resolution framework. Based on a segmentation of the background, characterization is made for each region independently as a mixture of K Gaussian modes, considering not only the model of the approximation coefficients but also the model of the detail coefficients at the different decomposition levels. Background region characterization is updated along time, and the detection of elements of interest is carried out computing the distance between background region models and those of each incoming image in the sequence. The inclusion of the context in the modeling scheme through each region characterization makes the model robust, being able to support not only gradual illumination and long-term changes, but also sudden illumination changes and the presence of strong shadows in the scene.

2. STRATEGY OVERVIEW

A block diagram of the proposed strategy is shown in Fig. 1.

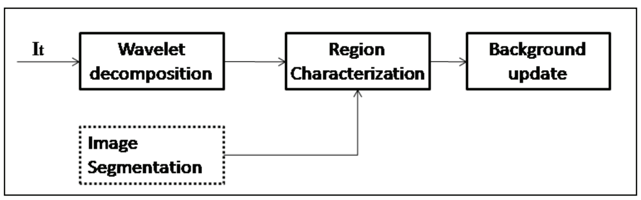

Figure 1: Block diagram of the proposed strategy

2.1. Wavelet Decomposition

The proposed strategy works at region level, so a partition of the background is assumed to have been previously computed based on any color or texture similarity criterion. This partition provides the topology of the regions that form the background, and its modeling and update is carried out through the analysis of the incoming images.

For each incoming image, a wavelet transformation is applied. This decomposition produces a set of coefficients which provide information at different resolution levels. "Approximation" coefficients are related with a low pass filtered version of the original image meanwhile "detail" coefficients are related with the high frequencies present in the scene. Detail coefficients also provide information about the orientation of the details in three directions: vertical (v), horizontal (h) and diagonal (d).

2.2. Region characterization

For each region of the background partition, a set of approximation and detail coefficients are associated – those lying within the region boundaries. Parametric models are applied to these coefficients thus resulting in a set of parameters at different resolution levels. This vector of parameters characterizes an instantiation of the region at a specific point in time (see Section 3). Whenever a new incoming image is analyzed, a new instantiation of the region is characterized.

2.3. Background modeling and update

In this strategy, a background region is modeled as a mixture of K Gaussian modes (MoG) in which one of them represents the region when it is considered to be part of the background. In our approach, each mode of the mixture models the deviations of the different instantiations from the prediction of the region throughout the time. Therefore, the standard deviation of each Gaussian mode represents the stability of its associated region. Besides, each mode has an associated factor that weights its relevance in the recent history of the region model. As a consequence, the mode holding the lowest value of the standard deviation and the higher weighting factor is likely to be the background representation of the region.

As it was mentioned in the introduction, the background model has to be updated in order to support long term and illumination changes. Based on the strategy proposed by Stauffer at al. [2], for each instantiation of the region, the parameters associated to the closest mode in the MoG are updated (parametric model of the estimated region and its standard deviation), and the weighting factors of all the modes are modified (see Section 4).

3. REGION CHARACTERIZATION

Parametric models are applied to the approximation and detail coefficients. The distribution of the approximation coefficients in the lowest resolution level for each region is modeled as a Gaussian distribution:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-1/2\left(\frac{x-\mu}{\sigma}\right)^2} \tag{1}$$

In this case, the parameters involved in the characterization process are the mean and the standard deviation values of the distribution, $\{\mu_c,\sigma_c\}_{c=[R,G,B]}$, where R, G, B represents the different color bands.

Based on the work by Mallat [12], for the detail coefficients and at each resolution level a generalized Gaussian distribution is assumed:

$$f(x) = Ke^{-\left(|x|/\alpha\right)^{\beta}} \tag{2}$$

In this case, the parameters involved in the characterization process are $\{(\alpha_{hl}, \beta_{hl}), (\alpha_{vl}, \beta_{vl}), (\alpha_{dl}, \beta_{dl})\}_{i=[1,2\ldots n]}$, where l represents the resolution level and h, v and d, the horizontal, vertical and diagonal detail coefficients respectively.

Some characterization results for a textured region are shown in Fig. 2. As it can be observed, the parameterized distribution for approximation and detail coefficients follows the real distribution tendency.

In order to measure the similarity between two regions, the Kullback-Leibler distance is proposed [13]: for each band and at each resolution level, the similarity between the two distributions involved is computed.

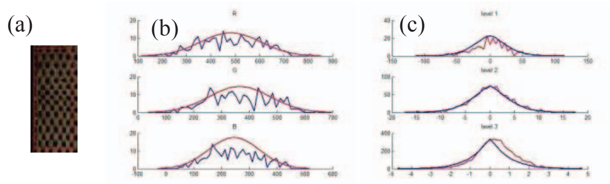

Figure 2: (a) Analyzed Region (b) In blue, RGB approximation coefficients distribution, in red, characterized distribution. (c) In blue, vertical detail coefficients distribution (level 1, 2 y 3), in red, characterized distribution.

4. BACKGROUND MODEL UPDATE

In this strategy, a background region is modeled as mixture of K gaussian modes. Each mode models the region deviations from the region estimation throughout the time.

$$P(R(i;t)) = \sum_{k=1}^{K} \omega_k(i;t) * \eta\left(d_b^k(i;t), 0, \Sigma_{k,t}\right) \tag{3}$$

where $R(i;t)$ represents the i-th region of the image in the instant t, $\omega_k(i;t)$ represents the weight of the k-mode, $\Sigma_{k,t}$ is the covariance matrix of the mixture, $d_b^{\,k}(i;t)$ is the distance of the instantiation of the region $R(i;t)$ in the instant t with respect to the estimation of the k-mode $\hat{R}_b^{\,k}(i;t)$ and η is the gaussian probability of belonging to each mode associated to the instantiation.

The distance for the approximation and detail coefficients is computed separately as an addition of the partial distances obtained for each set of coefficients at different scales, directions and bands. This approximation does not affect the effectiveness of the algorithm.

$$d_{app}^k(i;t) = \sum_{c=[R,G,B]} d_c^k(i;t)$$

$$d_{det}^k(i;t) = \sum_{l=[1,2,\dots,n]} \left(d_{hl}^k(i;t) + d_{vl}^k(i;t) + d_{dl}^k(i;t)\right) \tag{4}$$

We expect the detail coefficients to be less sensitive to sudden illumination changes so, approximation and detail coefficients are combined through a weighting constant γ as:

$$d_b^k(i;t) = \gamma d_{app}^k(i;t) + (1-\gamma) d_{det}^k(i;t) \tag{5}$$

When a new instantiation of a region arrives, it is compared with the K-gaussian modes. A region matches with one of the modes if the distance between the region estimation and its instantiation is below a threshold value:

$$d_b^k(i;t) < T_h \sigma_b^k(i;t) \tag{6}$$

Afterwards, all weighting factors are updated as:

$$\omega_k(i;t) = (1-\alpha)\omega_k(i;t-1) + \alpha(M_k(i;t)) \tag{7}$$

where α is the learning constant; $M_k(i;t)$ is 1 if the region matches the k-mode and 0 otherwise. If the region matches the k-mode, the standard deviation of this mode is updated:

$$\left(\sigma_b^k(i;t)\right)^2 = (1-\alpha)(\sigma_b^k(i;t-1))^2 + \alpha(d_b^k(i;t))^2 \tag{8}$$

and the parameters of the region estimation, $\hat{R}_b^{\,k}(i;t)$ as:

$$\begin{aligned} \hat{\mu}_c^k(i;t) &= (1-\rho)\hat{\mu}_c^k(i;t-1) + \rho\mu_c(i;t) \\ (\hat{\sigma}_c^k(i;t))^2 &= (1-\rho)(\hat{\sigma}_c^k(i;t-1))^2 + \rho(\sigma_c(i;t))^2 \\ \hat{\alpha}_l^k(i;t) &= (1-\rho)\hat{\alpha}_l^k(i;t-1) + \rho\alpha_l(i;t) \\ \hat{\beta}_l^k(i;t) &= (1-\rho)\hat{\beta}_l^k(i;t-1) + \rho\beta_l(i;t) \end{aligned} \tag{9}$$

where ρ is a weighting constant computed as:

$$\rho = \alpha\eta\left(d_b^k(i;t), 0, \sigma_b^k(i;t)\right) \tag{10}$$

As we operate at region-level, only one mode, B, is representative of the background, where:

$$B = \max_{k}\left(\frac{\omega_k(i;t)}{\sigma_b^k(i;t)}\right) \quad k = 1, \dots, K \tag{11}$$

5. RESULTS

To test the response of the proposed background modeling approach, different indoor video surveillance sequences have been used. Illumination changes have been introduced in the scene covering from gradual to sudden ones, new static objects are included (objects entering and becoming part of the background), and shadows cast either from objects and from moving objects are analyzed.

As expected, in long-term and gradual illumination changes, the proposed approach behaves adapting smoothly the model of the background regions. In case of changes in the scene configuration (appearance of new static objects), those are accurately detected, as pixel-based oriented approaches do, thus forcing a new mode to appear in the region model that after some time is identified as the background one (when its ratio B becomes the highest).

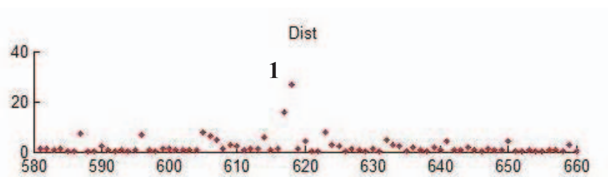

Figure 3: Behavior in a sudden illumination change (γ=0.1). The distance to the principal mode increases with the change, but no new mode appears in the mixture.

In the presence of sudden illumination changes, detail coefficients show significantly less sensitiveness than the approximation ones. Thus, the significance of detail coefficients in the distance computation (modeled through the value γ in (5)) can be used to tackle this situation efficiently. Low values of this parameter prevent the distance computation to recognize an illumination change as a reconfiguration in the scene. Fig. 3 shows the distance computation in case of γ=0.1, and how when the illumination changes abruptly (time indicated as 1 in the Fig.3), the distance changes but fast converges again to very low values. In pixel-based approaches, this situation would have been always identified as new object. Obviously, an inadequate use of this parameter (i.e. γ=0.5 in Fig. 4) would lead to the identification of a new mode in the MoG (in green) which would be initially identified as a new object in the scene.

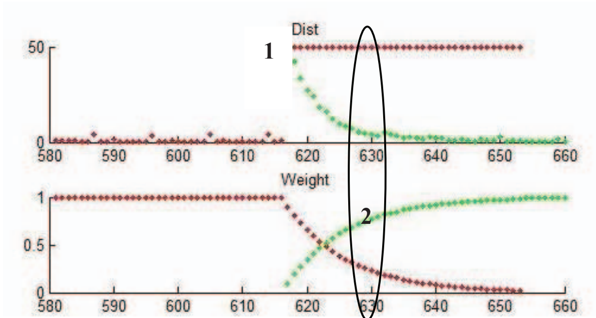

Figure 4: Behavior in a sudden illumination change (γ=0.5). In red, behavior of the initial principal mode. (1) Illumination change, appearance of a new mode (green). (2) The principal mode of the mixture changes.

In the presence of shadows, also the response of the detail information is more stable absorbing them in the model. Fig. 5 shows the distance evolution for a region which is affected by them (Fig. 6): although the distance to the representative mode increases, the model update strategy adapts to the new situation, thus preventing from the creation of a new mode in the MoG. Therefore shadows are not detected as new objects (as pixel-based approaches would do) and the proposed update strategy allows adapting the background model to its presence.

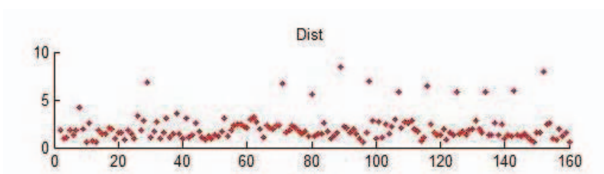

Figure 5: Distance to the principal mode in presence of shadows.

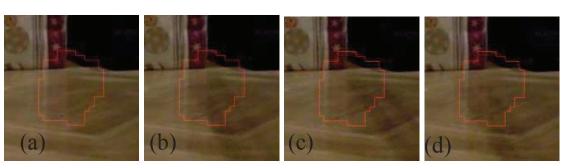

Figure 6: Video Sequence of a region affected by shadows. (a) Region free of shadows (b),(c),(d) Region in presence of shadows.

6. CONCLUSIONS

A novel approach to background modeling that operates at region-level in a wavelet based multi-resolution framework has been presented. The joint consideration of approximation and detail coefficients to model the regions as a MoG allows handling efficiently illumination changes (including sudden ones) the appearance of new objects becoming part of the background and shadows cast either by background objects or by moving objects in the scene. The information gathered in the proposed framework shows high potential not only for background modeling but also for intelligent analysis of the scene evolution.

ACKNOWLEDGEMENTS

This work has been partially supported by the Ministerio de Ciencia e Innovación of the Spanish Government under projects TEC2007-67764 (SmartVision) and TEC2010-20412 (Enhanced 3DTV).

7. REFERENCES

[1] C.R Wren, A. Azarbayejani, T. Darrel, A.P. Pentland, “Pfinder: real-time tracking of the human body” *IEEE Trans. on PAMI* , Vol. 19, pp. 780-785, 1997.

[2] C. Stauffer, W.E. Grimson, “Adaptive background mixture models for real-time tracking” *IEEE Conf. on Computer Vision and Pattern Recognition*, Vol. 2, 1999.

[3] A. Elgammal, D. Harwood, “Non-parametric model for background subtraction” *Proc. of the 6th. European Conf. on Computer Vision*, Vol. 2, pp. 751-767, 2000.

[4] K. Toyama, J. Kurm, B. Brummit, B. Meyers, “Wallflower: principles and practice of background maintenance” *Proc. of the 7th Internat. Conf. on Computer Vision*, Vol. 1, pp. 255-261, 1999.

[5] D. Butler, S. Sridharan, V.M. Bove, “Real-Time adaptive background segmentation” *Proc. of the IEEE Internat. Conf. on Acoustic, Speech & Signal Processing*, Vol. 3, pp. 349-352, 2003.

[6] K. Kim, T.H. Chaldabhongse, D. Harwood, L. Davis, “Background modeling and subtraction by codebook construction” *IEEE Intern. Conf. on Image Proc.*, Vol. 5, pp. 3061-3064, 2004.

[7] O.Javed, K.Shafique, M.Shah “A hierarchical approach to robust background subtraction using color and gradient info.” *Workshop on Motion and Video Computing*. pp. 22-27, 2002.

[8] J Zhang, L. Zhang, Tai Heng-Hing “Efficient video object segmentation using adaptive background registration and edge-based change detection techniques” *IEEE Internat. Conf. on Multimedia and Expo*, Vol. 2, pp. 1467-1470, 2004.

[9] Yee-Hong Yang, M.D. Levine, “The background primal sketch: An approach for tracking moving objects” Machine Vision and Applications, 2005.

[10] M. Heikkila M. Pietikainen, “A texture-based method for modeling the background and detecting moving objects” *IEEE Trans. on PAMI*, Vol. 28, pp. 657-662, 2006.

[11] M. Seki, T. Wada, H. Fujiwara, K. Sumi, “Background subtraction based on coocurrence of image variation” *Proc. IEEE Conf Computer Vision & Pattern Recog.* Vol. 2, pp. 65-72, 2003.

[12] G. Mallat, “A theory for multiresolution signal decomposition: the wavelet representation” *IEEE Trans. on PAMI*, Vol. 2, no. 7, pp. 674-693, 1989.

[13] G.V. Wouwer, P. Scheunders, D. V. Dyck, “Statistical texture characterization from Discrete Wavelet Representations” *IEEE Trans. on Image Processing*, Vol. 8, no. 4, pp. 592-598, 1999.