

MULTI-RESOLUTION ILLUMINATION COMPENSATION FOR FOREGROUND EXTRACTION

Xavier Suau Josep R. Casas Javier Ruiz-Hidalgo

Signal Theory and Communications Department, Technical University of Catalonia
UPC-Campus Nord, C/ Jordi Girona, 1-3, 08034, Barcelona
E-Mail: {xsuau@gps.tsc.upc.edu, josep.ramon.casas@upc.edu, j.ruiz@upc.edu}

ABSTRACT

Illumination changes may lead to false foreground (FG) segmentation and tracking results. Most of the existing FG extraction algorithms obtain a background (BG) estimation from temporal statistical parameters. Such algorithms consider a quasi-static BG which does not change but slowly. Therefore, fast illumination changes are not taken into account by the BG estimator and they are considered as FG. The aim of the proposed algorithm is to reduce illumination effects in video sequences in order to improve foreground segmentation performances.

Index Terms— Illumination, Foreground Extraction, Video Surveillance

1. INTRODUCTION

Illumination correction in video sequences is of special interest when dealing with outdoors scenes. Such sequences are affected by moving clouds, sunset light variation and other meteorological phenomena which may strongly modify pixel levels. Modifications due to this kind of situation may occur fast enough to mislead the FG extraction algorithm. The abruptly modified pixels (i.e. by clouds passing by) may be taken into account as FG pixels since the algorithm does not have enough time to assimilate them as BG. Therefore, FG objects are detected in still scenes without moving objects, and the resulting BG model is polluted by the new pixel levels.

Compensating illumination has been subject of study during recent years. Multi-Scale Retinex [1] and its variational version [2], both based on illumination low-pass filtering, help to reduce the effects of illumination on color. A normalized color space [3] may also be used to compensate color alteration due to lighting. Intrinsic images have been used by [4], [5] and [6] combined with an illumination sub-space. Even if performances are rewarding, real-time is achieved with difficulty and knowledge of future illumination situations is required to construct the illumination sub-space.

Lighting changes in outdoors scenes are random and of great variety, which implies a random modification of the video sequence too. However, lighting changes present some common characteristics that may be exploited in order to compensate the effects of illumination. Actually, illumination changes usually result in a *zone-wise modification* of the image, that is, pixel's levels will be modified the same way than in neighbour pixels. Thus, we can say that illumination changes affect neighbourhoods (or zones) of pixels by modifying their mean and contrast (as shown in Figure 1). The size of these zones is not defined and may even cover the whole image.

This work has been partly supported by the projects CENIT-HESPERIA 2005 and TEC2007-66858/TCM PROVEC of the Spanish Government



Fig. 1. Illumination effects on a static outdoors scene

In order to be able to evaluate the performances of the proposed algorithm our results are compared with those of a typical FG extraction algorithm. In this paper we have used a Mixture of Gaussians (MoG) [7] consisting of 3 gaussians with a 25-frame training period.

The remainder of this paper is organized as follows. In the next section, we introduce the mean and variance estimation algorithms and their multi-resolution versions. In Section 3, the Multi-Resolution Illumination Compensation (MIC) algorithm is presented. In Section 4, the performance obtained by our system is discussed. Finally, in the last section, conclusions are drawn and the direction of our future work in this subject is presented.

2. MULTI-RESOLUTION MEAN AND VARIANCE ESTIMATION

As mentioned above, illumination changes affect video sequences by modifying the mean value and variance of certain zones in each frame, the size of these zones being unknown.

We consider that illumination changes are mainly contained in the luminance channel Y . If we want to restore the non-lightened values of an illumination-affected frame, it seems reasonable to estimate the mean and variance of the illuminated zones in the Y channel. In order to do so, each frame Y of size $(W \times H)$ is divided into $(\mathcal{D}_k \times \mathcal{D}_k)$ rectangular sectors.

We subsequently calculate the mean and variance of each zone $Z_{i,j}$ $i, j \in [1 \dots \mathcal{D}_k]$ as follows :

$$\mu_{i,j} = \frac{1}{H_z W_z} \left[\sum_{x=1}^{H_z} \sum_{y=1}^{W_z} Y_{i,j}(x,y) \right] \quad (1)$$

$$\sigma_{i,j}^2 = \frac{1}{H_z W_z} \left[\sum_{x=1}^{H_z} \sum_{y=1}^{W_z} (Y_{i,j}(x,y) - \mu_{i,j})^2 \right] \quad (2)$$

$(W_z, H_z) = \left(\frac{W}{\mathcal{D}_k}, \frac{H}{\mathcal{D}_k} \right)$ is the size of a zone and $x, y \in Zone_{i,j}$

For each zone $Z_{i,j}$, we obtain two matrices (\tilde{M}_k and \tilde{V}_k of size $(\mathcal{D}_k \times \mathcal{D}_k)$), containing the calculated means $\mu_{i,j}$ and variances $\sigma_{i,j}$ respectively. These matrices are resized by a bilinear interpolation to the size of the original frame. The resulting *images* are called :

$$\mathbf{M}_k = \text{bilinear} \left(\tilde{M}_k \right)_{(\mathcal{D}_k \times \mathcal{D}_k) \rightarrow (W \times H)} \quad (3)$$

$$\mathbf{V}_k = \text{bilinear} \left(\tilde{V}_k \right)_{(\mathcal{D}_k \times \mathcal{D}_k) \rightarrow (W \times H)} \quad (4)$$

Since the size of the illuminated zones may vary, we consider a resolutions set $\mathcal{R} = \{\mathcal{D}_1, \mathcal{D}_2, \dots, \mathcal{D}_L\}$, each \mathcal{D}_k being the number of divisions per image side. *i.e.* $\mathcal{D}_k = 4 \Rightarrow 4 \times 4 = 16$ zones. For each element $\mathcal{D}_k \in \mathcal{R}$ we calculate a mean and variance estimation, obtaining two sets $\mathcal{R}_M = \{\mathbf{M}_1, \mathbf{M}_2, \dots, \mathbf{M}_L\}$ and $\mathcal{R}_V = \{\mathbf{V}_1, \mathbf{V}_2, \dots, \mathbf{V}_L\}$ of estimations, which are called Multi-Resolution Estimation.

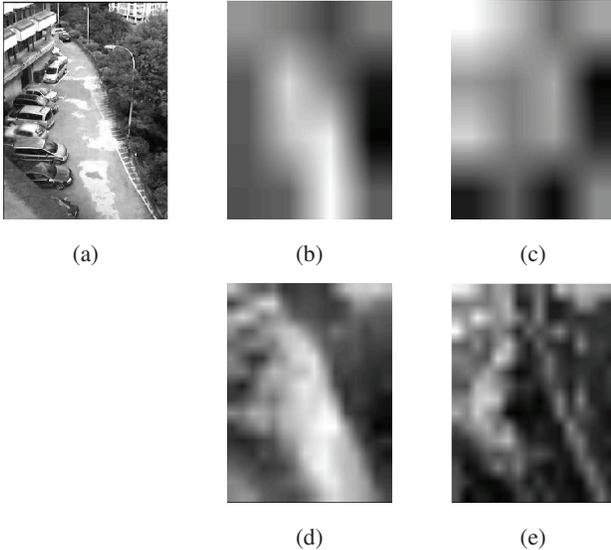


Fig. 2. Multi-Resolution Estimation for $\mathcal{R} = \{4, 16\}$. (a) Original luminance. (b) \mathbf{M}_1 for $\mathcal{D}_1 = 4$. (c) \mathbf{V}_1 for $\mathcal{D}_1 = 4$. (d) \mathbf{M}_2 for $\mathcal{D}_2 = 16$. (e) \mathbf{V}_2 for $\mathcal{D}_2 = 16$. (Remark that (b,d) constitute \mathcal{R}_M and (c,e) constitute \mathcal{R}_V)

3. MULTI-RESOLUTION ILLUMINATION COMPENSATION (MIC)

The main purpose of the research carried out in this paper is to compensate the effects of illumination changes in a video sequence in order to improve FG extraction results. Our idea is to exploit multi-resolution mean and variance masks calculated as explained in section 2 to accomplish this objective.

We assume that the resolutions set \mathcal{R} verifies our size constraint presented in Equation (5). Actually, every image division must be bigger than the objects to be tracked. Otherwise, it would be impossible to determine if changes are due to illumination variations or to objects passing by.

$$Size(Zone_k) > Size(Objects) \quad \forall k \quad (5)$$

3.1. Mean-Variance Normalization

An image I with mean μ_I and variance σ_I^2 can be normalized to a given mean μ and variance σ^2 by means of the following expression:

$$\bar{I} = (I - \mu_I) \cdot \frac{\sigma}{\sigma_I} + \mu \quad (6)$$

Ideally, every frame of our sequence should be normalized to a constant illumination no matter the situation to be faced. However, constant illumination is impossible to achieve since the real illumination field is unknown (it is precisely what we are looking for). Instead of using the *real* illumination and trying to remove it, our approach is to bring every frame to a constant mean level and contrast, called *nominal illumination* or Y_0 . Mean level and variance of Y_0 should be chosen in such a way that they cover the whole dynamic range. Since we work with 256 intensity levels, the mean value of Y_0 is then set to $\mu_0 = 128$. As for the variance, we set a nominal standard deviation of $\sigma_0 = 40$ in order to adjust $\mu_0 \pm 3\sigma_0$ to the 256 levels of the image range.

$$Y_0 : \left\{ \begin{array}{l} \mu_0 = 128 \\ \sigma_0 = 40 \end{array} \right\} \quad (7)$$

3.2. Illumination compensation

First of all, a resolution set $\mathcal{R} = \{\mathcal{D}_1, \mathcal{D}_2, \dots, \mathcal{D}_L\}$ of dimension L must be defined. Subsequently, we calculate the multi-resolution estimations \mathcal{R}_M and \mathcal{R}_V as seen in Section 2. The luminance channel Y is then normalized according to equation (6)¹ with every element of the estimation sets. Each resolution layer is weighted by $p_k = \frac{1}{L}$ and added-up to the final illumination compensated luminance \bar{Y} .

$$\bar{Y} = p_k \cdot \sum_{k=1}^L \left[(Y - \mathbf{M}_k) \cdot \frac{\sigma_0}{1 + \mathbf{V}_k} + \mu_0 \right] \quad (8)$$

\bar{Y} in equation (8) is illumination compensated for a great range of illumination resolutions². Actually, we compensate illumination changes which light zones of size $\mathcal{D}_k \in \mathcal{R}$. Furthermore, the elements and the dimension of \mathcal{R} are to be adapted to each application depending on the size of the expected illuminated areas and the size of the target objects.

4. RESULTS

Tests are effectuated on two types of scenes:

- *Static sequences* with no moving objects and fast illumination changes due to clouds passing by (Scenes 1, 2 and 3).
- *Dynamic sequences* with moving objects such as pedestrians and vehicles. These scenes will help to confirm tracking performances after MIC processing.

¹division by $(1 + \mathbf{V}_k)$ is done to avoid division by zero

²equation (8) corresponds to our MIC algorithm

Results are focused on static scenes for which we ideally expect no FG detection. However, fast illumination transitions mislead MoG algorithm, which detects large areas as new FG objects. Actually, in order to be able to overcome illumination transitions, we may increase MoG's *learning factor* α . Unfortunately, this solution will make MoG algorithm *learn* slow-moving objects as BG, preventing a correct FG detection. Two main strategies are compared:

- (MIC+MoG) : Our MIC pre-processing algorithm followed by MoG with a typical value of $\alpha = 0.005$. MIC algorithm is adjusted with a resolutions set of $\mathcal{R} = \{1, 4, 8, 16\}$
- (MoG $_{\alpha}$) : No frame pre-processing followed by MoG with $\alpha = 0.005, 0.2$

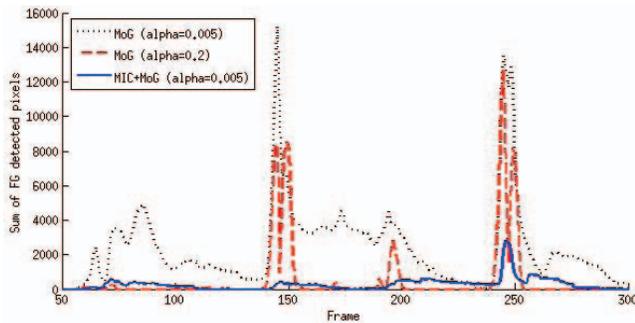


Fig. 3. Sum of FG detected pixels per frame for a static sequence (Scene 1) with 3 main illumination episodes starting at frames 60, 140 and 240 approx. (ideal detection should be 0 pixels)

Figure 3 shows how increasing α reduces the number of detected FG pixels. Nevertheless, it has been tested in other dynamic sequences that objects are rapidly lost when they stop or move slowly. Our MIC+MoG solution improves MoG $_{\alpha}$ in terms of false FG detections, the peaks due to lighting changes being strongly smoothed. Throughout the whole sequence, 595754 FG pixels have been detected with MoG $_{0.005}$, 144049 with MoG $_{0.2}$ and 66354 with the MIC+MoG solution, which represents around a 90% and 50% reduction respectively. Furthermore, we have confirmed that object detection and tracking is reliable in MIC-processed sequences. A frame pre-processing with *our MIC algorithm followed by a state-of-the-art FG extraction algorithm such as MoG is an interesting strategy to overcome heavy illumination variations.*

Figure 4 contains the resulting MIC-processed images for two different frames in terms of illumination. It can be appreciated how MIC brings both images to the Y_0 reference luminance levels.

As an example, the frame in Figure 5.(a,b,c) corresponds to an episode of static Scene 1 with heavy illumination transitions. It can be clearly seen how MIC+MoG strategy drastically reduces both the number of detected FG pixels and the number of tracked objects (ideally, no FG pixel or object should be detected). During the whole Scene 1 sequence, we have detected 59 false objects with MoG $_{0.005}$ and only 9 false objects with MIC+MoG strategy. Results for static Scenes 1, 2 and 3 are summarized in Table 1.

As for tracking, objects detected and tracked with MoG $_{0.005}$ and MIC+MoG strategies are compared. MoG $_{0.02}$ strategy is abandoned because of its tracking problems with slow objects, as shown in Figure 5.e. On the other hand, MIC+MoG strategy does not reduce persons, or other objects, tracking capacity (Figure 5.(d,f)).

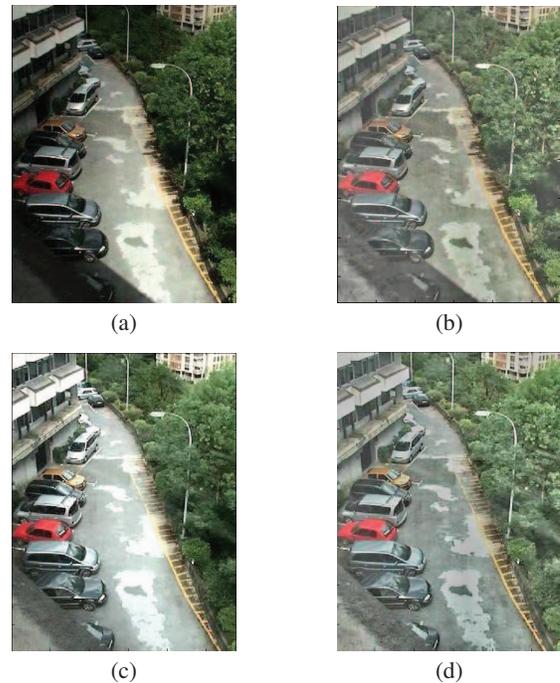


Fig. 4. Multi-Resolution Illumination Correction results for Scene 1.(a) Original frame 143. (b) MIC-processed frame 143. (c) Original frame 200. (d) MIC-processed frame 200.

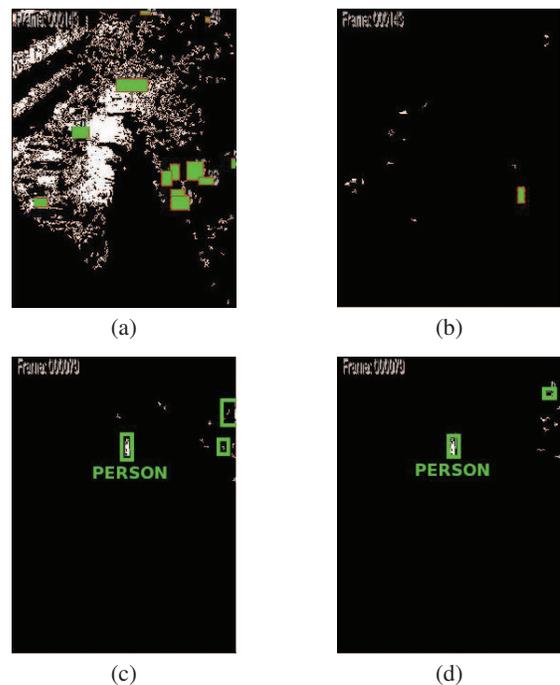


Fig. 5. Detected FG pixels and tracked objects for static Scene 1 (a,b) and for dynamic scene with one person (c,d). (a,c) MoG $_{0.005}$ output. (b,d) MIC+MoG $_{0.005}$ output

	Strategy	FG pixels	Tracked objects
Scene 1:	MoG _{0.005}	595754	59
	MoG _{0.2}	144049	-
	MIC+MoG	66354	9
	Strategy	FG pixels	Tracked objects
Scene 2:	MoG _{0.005}	843658	25
	MoG _{0.2}	176354	-
	MIC+MoG	35057	1
	Strategy	FG pixels	Tracked objects
Scene 3:	MoG _{0.005}	86464	29
	MoG _{0.2}	5863	-
	MIC+MoG	2697	11

Table 1. Obtained results for 3 static scenes polluted by heavy illumination changes

The temporal evolution of MoG’s BG model for a given pixel is presented in Figure 6, which contains the mean \pm standard deviation of each MoG _{α} gaussian (up to 3 gaussians). We compare BG models obtained from the original sequence and from the MIC pre-processed sequence. The stand-alone MoG_{0.005} algorithm uses the 3 available gaussians throughout almost the whole sequence, with a great amount of shifting and re-ordering³ phases. This jumping phenomena appear since MoG tries to adapt its BG model to new pixel intensities. On the other hand, MIC+MoG strategy uses one single gaussian, illumination effects being barely visible. MIC+MoG solution strongly smooths the evolution of MoG gaussians and reduces the number of used gaussians in the BG model, which *gives more degrees of freedom to be able to account for additional actual BG changes*.

5. CONCLUSIONS AND FUTURE WORK

The proposed MIC algorithm smoothes illumination effects drastically, which improves FG detection with state-of-the-art algorithms. Even if image contrast is modified, both the number of wrongly detected pixels and false objects are reduced. What is more, our frame processing does not affect real objects tracking.

Furthermore, MoG’s background model is modified by our MIC algorithm, becoming more stable against illumination modifications. In addition, the BG is represented with a lower number of gaussians. Reducing the number of used gaussians gives the system more degrees of freedom to incorporate additional real BG variations into the model.

Spatial adaption is supplied by the resolutions set \mathcal{R} . The elements in \mathcal{R} give an idea of the illumination compensation resolution, which may be adapted to different applications. The more elements in \mathcal{R} the better the illumination will be corrected, a larger range of zones being processed. On the other hand, calculations volume will considerably increase and real-time may be more difficult to attain. Moreover, size condition (5) must be verified by every element in \mathcal{R} in order to preserve tracking properties.

Future research will be focused on studying the temporal dimension of the algorithm, instead of limiting it to a per-frame processing. The weights p_k in Equation (8) may be adapted too, with a view to attach importance to particular resolution levels. Eventually, feed-backing MIC with MoG’s parameters, and vice versa, is envisaged.

³MoG gaussians are re-ordered according to their associated weights

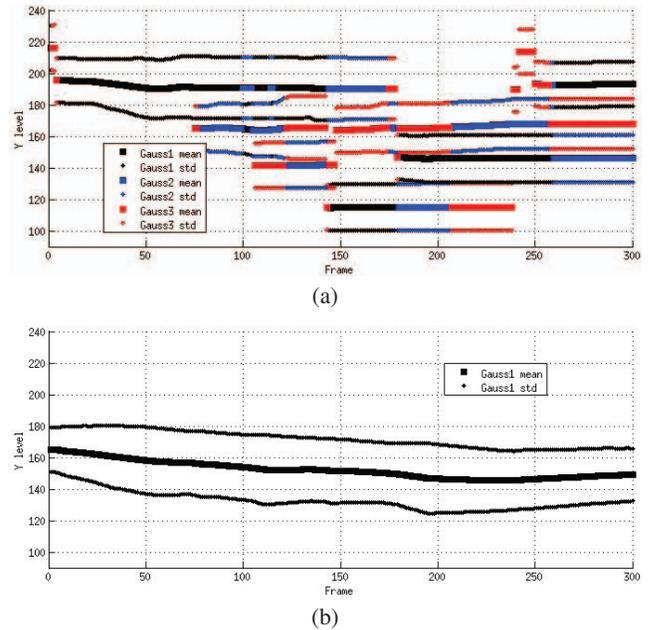


Fig. 6. Evolution of $(\mu \pm \sigma)$ of MoG gaussians for a given pixel in Scene 1. The chosen pixel (131,94) is placed on the pavement of the street in Figure 6.a. (a) MoG on original sequence. (b) MoG on MIC-processed sequence.

6. REFERENCES

- [1] Vivek Agarwal, “Achieving color constancy and illumination compensation,” *ECE 574*, pp. 1–29, 2003.
- [2] R. Kimmel, M. Elad, D. Shaked, R. Keshet, and I. Sobel, “A variational framework for retinex,” in *International Journal of Computer Vision*, 2003, vol. 52, pp. 7–23.
- [3] M. Xu and T. Ellis, “Illumination-invariant motion detection using colour mixture models,” 2001.
- [4] Y. Weiss, “Deriving intrinsic images from image sequences,” in *International Conference on Computer Vision, ICCV*. IEEE, 2001, vol. 2, pp. 68–75.
- [5] Y. Matsushita, K. Nishino, K. Ikeuchi, and M. Sakauchi, “Illumination normalization with time-dependent intrinsic images for video surveillance,” in *Computer Society Conference on Computer Vision and Pattern Recognition*. IEEE, 2005, vol. 26, pp. 1336–1347.
- [6] F. Porikli, “Multiplicative background-foreground estimation under uncontrolled illumination using intrinsic images,” in *Workshop on Motion and Video Computing (WACV/MOTION’05)*. IEEE, 2005, vol. 2, pp. 20 – 27.
- [7] C. Stauffer and W.E.L. Grimson, “Adaptive background mixture models for real-time tracking,” in *Computer Society Conference on Computer Vision and Pattern Recognition*. IEEE, 1999, vol. 2, pp. 246–252.