# A global probabilistic framework for the foreground, background and shadow classification task

**Conference Paper** *in* Proceedings / ICIP ... International Conference on Image Processing · December 2009

DOI: 10.1109/ICIP.2009.5414388 · Source: IEEE Xplore



**6 authors**, including:

David Marimon
Catchoom
**27** PUBLICATIONS **300** CITATIONS
SEE PROFILE

Oscar Divorra Escoda
Vonage
**42** PUBLICATIONS **491** CITATIONS
SEE PROFILE

# A GLOBAL PROBABILISTIC FRAMEWORK FOR THE FOREGROUND, BACKGROUND AND SHADOW CLASSIFICATION TASK

*Jose-Luis Landabaso, Jose-Carlos Pujol-Alcolado, Tomas Montserrat, David Marimon, Jaume Civit, Oscar Divorra Escoda*

Telefonica Research
Barcelona, Spain - http://www.tid.es

## ABSTRACT

Over the years, many works have been published on the two-dimensional foreground segmentation task, describing different methods that treat to extract that part of the scene containing active entities. In most of the cases, the stochastic background process for each pixel is modeled first, and then the foreground pixels are classified as an exception to the model or using maximum a posteriori (MAP) or maximum likelihood (ML). The shadow is usually removed in a later stage and salt and pepper noise is treated with connected component analysis or mathematical morphology. In this paper, we propose a global method that classifies each pixel by finding the best possible class (foreground, background, shadow) examining the image globally. A Markov Random Field is used to represent the dependencies between all the pixels and classes and the global optimal solution is approximated with the Belief Propagation algorithm. The method can extend most local methods and increase their accuracy. In addition, this approach brings a probabilistic justification of the classification problem and it avoids the use of additional post-processing techniques.

***Index Terms***— Foreground Segmentation, Belief Propagation, Shadow Removal, Global, Markov Random Fields

## 1. INTRODUCTION

The most popular methods for detecting foreground observations at each pixel correspond to those that classify observations into foreground as an exception to a background model [1, 2, 3, 4, 5, 6]. Of course, foreground models can also help. However, it is often difficult to model the color appearances of moving objects. Besides, the foreground models of each object have to be mapped to each pixel at each instant before performing a classification, which is also prone to errors. On the contrary, the background appearance of a pixel can be robustly learned using fixed cameras and, therefore, background models usually provide the most reliable source of information in the segmentation task.

Shadow removal algorithms are usually incorporated after the background subtraction step as follows. First, the expected background for each frame is estimated using the background modeling algorithm. Then, some cues are extracted from the expected background. These cues are used then to identify if a pixel is a cast shadow/highlight pixel or not. Prati et al. have presented an in-depth survey of those algorithms [7].

This work has been partially performed within the framework of EU FP7 Project 3DPresence and within the framework of the Spanish Agency CDTI project CENIT-VISION 2007-1007.

In this paper, we propose a global classification framework. This framework allows an easy incorporation of well-known background models. We also provide a simple solution for not having to model the foreground appearances and still be able to use the probabilistic global setting. Finally, the proposed framework incorporates the shadow process as an additional class in the classification instead of being a post-processing step as is it often found in the literature.

This paper is structured as follows. Section 2 describes the local methods applied to classify the background and shadow pixels. The proposed global foreground segmentation based on local classification is explained in section 3. Experiments and results are dealt with in section 4. The paper concludes with a summary of the improvements accomplished and the future paths of research.

## 2. LEARNING THE BACKGROUND WITH LOCAL METHODS

### 2.1. Local foreground segmentation

We adopt a single-class statistical model for modeling the background color of a pixel $\mathbf{x}$ (indicating its spatial coordinates), given observations of its color value $\mathbf{I}(\mathbf{x})$ across time. For this purpose, we use a Gaussian probability density function. Gaussians have been previously proposed in [8, 5, 6], among others, to ensure that the cameras thermal noise does not produce classification errors. Some of these works [8, 5], adopt multi-class models to model repetitive background, such as in weaving flags, or moving trees. However, a single-class model is enough in our approach, since our system is being developed to operate in a scene that consists of a relatively static situation:

$$G_{\mathbf{x}}(\mathbf{I}(\mathbf{x})) = \frac{1}{(2\pi)^{3/2}\sqrt{|\mathbf{\Sigma_x}|}} e^{-\frac{1}{2}(\mathbf{I}(\mathbf{x})-\mu_{\mathbf{x}})^T \mathbf{\Sigma_x^{-1}} (\mathbf{I}(\mathbf{x})-\mu_{\mathbf{x}})}, \quad (1)$$

corresponding to the Gaussian that models the color of the background process of pixel $\mathbf{x}$, and where pixel color values ($\mathbf{I}(\mathbf{x})$) are expressed as a vector of three dimensions in the $RGB$ color space. Often it is assumed that the covariance matrix is diagonal with $R$, $G$ and $B$ sharing the same variances: $\mathbf{\Sigma_x} = \sigma_{\mathbf{x}}^2 \cdot \mathbf{Id_{3\times 3}}$.

Similarly as in [5], model adaptation is implemented as a low pass filter procedure. Thus, once the pixel value has been classified into the background, the model is adapted as follows

$$\begin{aligned} \mu_{\mathbf{x}}[t] =& (1-\rho)\mu_{\mathbf{x}}[t-1] + \rho \mathbf{I}(\mathbf{x}) \\ \sigma_{\mathbf{x}}^2[t] =& (1-\rho)\sigma_{\mathbf{x}}^2[t-1] + \\ & + \rho(\mathbf{I}_{\mathbf{x}}(\mathbf{x}) - \mu_{\mathbf{x}}[t])^T (\mathbf{I}_{\mathbf{x}}(\mathbf{x}) - \mu_{\mathbf{x}}[t]), \end{aligned} \quad (2)$$

where $\rho$ is the adaptation learning rate: $\rho \propto G_{\mathbf{x}}(\mathbf{I}(\mathbf{x}))|_t$.

The foreground process can be modeled using histograms, Gaussians or any other *pdf*. However, we simply use a uniform *pdf* to model the foreground process in each pixel as in [9], which is in fact the probabilistic extension of classifying a foreground pixel as an exception to the model, as discussed in [10]. Since a pixel admits $256^3$ colors in the $RGB$ color space, we model its *pdf* as $p(\mathbf{I}(\mathbf{x})|\phi) = \frac{1}{256^3}$.

Then, the probability that a pixel $\mathbf{x}$ belongs to the foreground $\phi$, given an observation $\mathbf{I}(\mathbf{x})$, can be expressed in terms of the likelihoods of the foreground and background processes as follows

$$P(\phi|\mathbf{I}(\mathbf{x})) = \frac{P(\phi)p(\mathbf{I}(\mathbf{x})|\phi)}{p(\mathbf{I}(\mathbf{x}))},$$

where $P(\phi)$ is the prior probability of foreground. A pixel can be classified as foreground using maximum a posteriori if $P(\phi|\mathbf{I}(\mathbf{x})) > \frac{1}{2}$ is satisfied.

Finally, the Gaussian model is adapted using (2), when the pixel is classified into the background.

### 2.2. Learning shadows with color statistics

A shadow is normally an area that is not or only partially irradiated or illuminated because of the interception of radiation by an opaque object between the area and the source of radiation. Assuming that the irradiation consists only of white light, the chromaticity in a shadowed region should be the same as when it is directly illuminated. The same also applies to lightened areas in the image. Based on the same assumption, a normalized chromatic color space, *e.g.* $r = R/(R+G+B)$, $g = G/(R+G+B)$, is immune to shadows, but the lightness information is unfortunately lost. Keeping it is important in order to avoid some simple errors such as confusing a white car with a gray road.

It is relevant to underline the fact that we are only interested in detecting shadows that form part of the foreground objects. Shadows that form part of the background are not analyzed as they are assumed to be constant along time. Our method to detect shadows is based on the previous observation about chromaticity and brightness distorion over shadowed regions. More precisely, a shadow removal algorithm needs to analyze foreground pixels and detect those that have similar chromaticity but lower brightness to the corresponding region when it is directly illuminated. The adaptive background reference image provides the desired information.

In view of the fact that both brightness and chromaticity are very important, a good distortion measure between foreground and background pixels should account for the discrepancies in both their brightness and chromaticity components as proposed by Horpraset et al. in [2]. Following [2], Brightness Distortion ($BD$) is defined as a scalar value that brings expected background close to the observed chromaticity line. Similarly, color distortion ($CD$) is defined as the orthogonal distance between the expected color and the observed chromaticity line. Both measures can be graphically represented as in Fig. 1 and formulated as:

$$\begin{aligned} BD &= \underset{\alpha}{\operatorname{argmin}} ||\mathbf{I}(\mathbf{x}) - \alpha \cdot \mu_{\mathbf{x}}|| \\ CD &= ||\mathbf{I}(\mathbf{x}) - BD \cdot \mu_{\mathbf{x}}||, \end{aligned} \tag{3}$$

where $\mu_{\mathbf{x}}$ is the mean in the $RGB$ color-space of the background process defined in (1). Note that brightness distortion values over 1 correspond to lighter foreground and, on the other hand, the foreground is darker when BD is below 1.

Finally, a set of thresholds need to be defined to assist the classification into foreground, background and shadow [11].

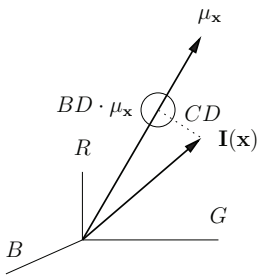



**Fig. 1**. Distortion measurements in the $RGB$ color space.

Note that this technique fulfills its objective not to remove self-shadows as they do not share similar brightness or chromaticity with their background reference image.

Although the algorithm gives good local classification of shadows, its application requires manual tunning of the thresholds. Our goal is to be able to globally classify shadows. Therefore, we need a method to characterize shadows locally. We propose to model the shadow statistics of a pixel as a bi-dimensional Gaussian distribution on the $BD$ and $CD$ space. Our purpose is to obtain the mean and standard deviation using the observed brightness and chromaticity for each pixel $\mathbf{x}$, incrementally. The samples of that observation are taken each time a pixel is classified as shadow. Then, assuming statistical independence between the brightness and chromaticity stochastic processes, this *pdf* can be expressed as:

$$p(BD, CD) = \frac{1}{2\pi\sigma_{BD}\sigma_{CD}} e^{-\frac{(BD-\mu_{BD})^2}{2\sigma_{BD}^2} - \frac{(CD-\mu_{CD})^2}{2\sigma_{CD}^2}}. \tag{4}$$

And, similarly as in (2), means $\mu_{BD}$, $\mu_{CD}$ and variances $\sigma_{BD}^2$, $\sigma_{CD}^2$ can be updated as a low pass-filter process with each incoming observation that is classified as shadow.

This learned local shadow model is further used in the global foreground segmentation method described in section 3.

## 3. EXTENSION TO A GLOBAL CLASSIFICATION

In the previous section, we have expressed the foreground, background and shadow stochastic processes in terms of their probabilistic density functions. For the particular case of the shadow, it is possible to compute $BD$ and $CD$ from the $RGB$ observation and evaluate it with (4) to obtain the likelihood of a particular observation of belonging to the shadow class. In fact, with the aforementioned mathematical expressions, it is possible to use a MAP setting to probabilistically classify each new observation.

However, we prefer to solve the problem globally, not assuming that pixel observations are independent among their neighbors. The problem we want to solve is:

$$\Gamma_j(\mathbf{x}) = \underset{\Gamma_k}{\operatorname{argmax}} P(\Gamma_k)P(\Gamma_k(\mathbf{x})|\mathbf{I}), \tag{5}$$

where $\Gamma_k$ is any of the 3 classes and $\mathbf{I}$ is the whole image, *i.e.*, not just one observation in a pixel $\mathbf{I}(\mathbf{x})$.

### 3.1. Belief propagation (BP)

In order to solve (5), we state the problem as a Markov Random Field. The Markov property asserts that the conditional probability of a site in the field depends only on its neighboring sites. And in

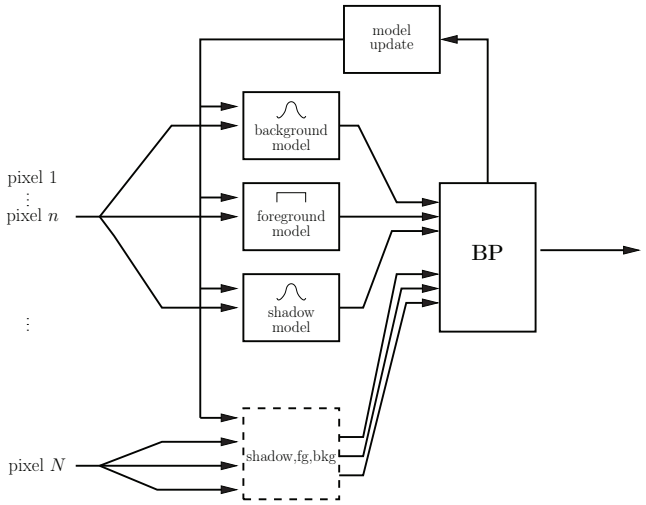



**Fig. 2**. The schematic diagram of the BP-based foreground, background and shadow classification method.

order to enforce spatial interaction between pixels, it is necessary to define which are the costs of assigning different classes to neighboring pixels. In our implementation, we assign two different types of transference costs. The cost we assign to transitions *not* going to/from the foreground class is $c = -ln(0.2)$, which corresponds to the 20% of the probability of staying in the same class. On the other hand, we double the costs $(2 \cdot c)$ to any transition that goes to/from the foreground class. The reason for doubling these costs is that the shadow class is semantically closer to background than foreground. Note that a shadowed region could also be considered a background region where there has been a change of illumination.

Note that even if the costs are given, it is intractable to solve the MAP setting of such a problem. In this paper, we propose to use the Loopy Belief Propagation [13] inference algorithm as an approximation to the solution.

### 3.2. Global foreground segmentation with shadow removal

The outline of the algorithm, which has been depicted in Fig. 2 is as follows. First, the likelihoods of each observation and class are obtained locally. From such likelihoods we obtain the local data costs which are the input to the BP algorithm. Then, the Markov Random Field is solved with the BP algorithm using the transition costs aforementioned. Once the image has been classified globally, then the background, $BD$ and $CD$ Gaussian models are updated following the low pass-filter procedure described in (2). The process is performed if the observation is classified with their corresponding class. The foreground model is not updated since it corresponds to a fixed uniform *pdf* function.

## 4. RESULTS

In this section, three different methods are compared. A local only method, a local method with some post-processing, consisting in a morphological aperture followed by a hole filling and the global described in the previous section. The ground truth has been obtained after manual segmentation of 10 consecutive frames and on a dataset captured in the context of the 3DPresence EU project by the Fraunhofer Institute for Telecommunications/Heinrich-Hertz-Institut in Berlin.

In order to assess the performance of the mentioned algorithms, two types of measures are calculated. The first set is intended to evaluate the ability of the system to recognize the true solution at every frame, independently from the consistency of the solution through time. The second set of measures is designed to assess the temporal homogeneity of the found solution; in other words, the *flickering* effect.

We include the *recall*, *precision* and *f-measure* in the first set. The *recall* is defined as $\frac{\#\text{ Correct detections}}{\#\text{ Correct detections}+\#\text{ misses}}$, the *precision* is calculated as $\frac{\#\text{ Correct detections}}{\#\text{ Correct detections}+\#\text{ false alarms}}$. Finally, the *f-measure* is intended to combine the two quantities: a good system will be balanced both in *recall* and *precision*: $\textit{f-measure} = \frac{2\times \textit{recall}\times \textit{precision}}{\textit{recall}+\textit{precision}}$.

The results for the different algorithms are shown in Table 1 and a particular frame has been presented in Fig. 3.

**Table 1**. Instant System Evaluation

| | G. truth | Local | Loc. & Post P. | Global |
|---|---|---|---|---|
| # Correct det. | 201642 | 158401 | 181902 | **183565** |
| # False al. | 0 | 43241 | 19740 | **18077** |
| # Misses | 0 | 12864 | **8348** | 8521 |
| F.A. rate | 0 | 0.005 | 0.003 | **0.002** |
| Miss rate | 0 | 0.002 | **0.001** | **0.001** |
| Error rate | 0 | 0.007 | 0.004 | **0.003** |
| *recall* | 1 | 0.950 | 0.934 | **0.956** |
| *precision* | 1 | 0.786 | 0,902 | **0.910** |
| *f-measure* | 1 | 0.860 | 0.918 | **0.932** |

Note that the global method gives more accurate results than the rest of methods in terms of *precision*, *recall* and *f-measure*. The second best method is the local method with post-processing and the worst one corresponds to the local only method. In order to be as fair as possible with the comparisons, we have employed exactly the same Gaussian classes in the local and global methods. However, the classification is done using MAP at the pixel level in the local methods and BP in the global method.

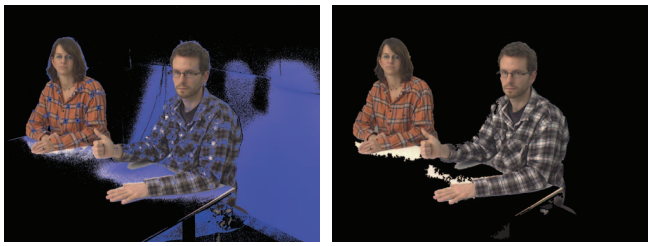
In Fig. 3, results of the three different methods are also presented for visual inspection. The shadows are highlighted in blue and the background is set to black. Note the noise in the classification of the local method due to not having any kind of spatial regularity. In addition, it should be pointed out that that post-processing technique has to be tunned for each new dataset while such tunning is not needed in the global method proposed here.

Temporal consistency is measured by comparing the *false alarms* and *misses* for every frame $t$. This new error frame is denoted as $E_t$, and is calculated with an XOR operation between the ground truth and estimated solution. We compute the accumulated temporal error image, $T$, as $T = \sum_{t=0,N-1}(XOR(E_t, E_{t+1}))$. The average temporal error energy per pixel (ATEE) corresponds to ATEE $= \frac{\sum_{x,y} T}{N-1}$. The lower the ATEE value, the less flickering effect. ATEE $= 1$ would mean that every single pixel varies from the ground truth solution to its inverse during the whole sequence. The obtained values of ATEE are shown in Table 2.

**Table 2**. Temporal System Evaluation

| | G. truth | Local | Loc. & Post P. | Global |
|---|---|---|---|---|
| ATEE | 0 | 0.019545 | 0.010435 | 0.006676 |

Note that the global method produces twice less flickering than the other methods. This effect can be clearly observed in the videos we made available: http://www.3dpresence.eu/icip09.



(Local method) (Post-processed foreground with local method)

(Global method) (Foreground with global method)

**Fig. 3**. Images corresponding to foreground segmentation with the discussed methods. Left column, classification results. Right column, foreground segmentation after removing shadows.

## 5. CONCLUSIONS

In this paper we have proposed a unified framework for the foreground, background and shadow classification task. The proposed framework operates in a global manner by transferring probabilistic information between neighboring pixels which are set as a Markov Random Field. The algorithm can be summarized as follows. First, the likelihoods of background, foreground and shadow are obtained per each pixel locally. Then, these costs, together with the transference cost are employed by a loopy BP technique to obtain an approximate solution of the global MAP classification problem. After all the pixels are classified, the background and shadow models are updated using a low-pass filter approach.

The proposed method is better than local methods in terms of *precision*, *recall* and *f-measure*. In addition, the global segmentation method presents lower flickering, which is an important factor to consider depending on the field of application of the segmentation.

These results are obtained at the cost of a higher computational cost. The high CPU requirements of the algorithm makes it unfeasible for real-time operation with current PC hardware. However, the algorithm presented here shares some principles with other algorithms focused on obtaining depth maps from stereo camera pairs with BP. Some of these works have been ported to GPU hardware bringing them closer to real-time operation. This fact opens the door to a future GPU-based implementation of our approach that would also make it run in real-time.

## 6. REFERENCES


[1] Ahmed Elgammal, Ramani Duraiswami, David Harwood, and Larry S. Davis, "Non-parametric model for background subtraction," in *Proceedings of International Conference on Computer Vision*. Sept 1999, IEEE Computer Society. 1

[2] T. Horpraset, D. Harwood, and L. Davis, "A statistical approach for real-time robust background subtraction and shadow detection," in *Proceedings of International Conference on Computer Vision*. 1999, IEEE Computer Society. 1, 2

[3] Haritaoglu, D. Harwood, and L. Davis, "W4: Real time surveillance of people and their activities," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, August 2000. 1

[4] Stephen J. McKenna, Sumer Jabri, Zoran Duric, Azriel Rosenfeld, and Harry Wechsler, "Tracking groups of people," *Computer Vision and Image Understanding*, vol. 80, no. 1, pp. 42–56, 2000. 1

[5] Chris Stauffer and W. Eric L. Grimson, "Learning patterns of activity using real-time tracking," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 747–757, 2000. 1

[6] Christopher Richard Wren, Ali Azarbayejani, Trevor Darrell, and Alex Paul Pentland, "Pfinder: Real-time tracking of the human body," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 780–785, 1997. 1

[7] Andrea Prati, Ivana Mikic, Mohan M. Trivedi, and Rita Cucchiara, "Detecting moving shadows: Algorithms and evaluation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, pp. 918–923, 2003. 1

[8] Nir Friedman and Stuart J. Russell, "Image segmentation in video sequences: A probabilistic approach," in *Proceedings of Conference on Uncertainty in Artificial Intelligence*, 1997, pp. 175–181. 1

[9] J. L. Landabaso and M. Pardàs, "Cooperative background modelling using multiple cameras towards human detection in smart-rooms (invited paper)," in *Proceedings of European Signal Processing Conference*, 2006. 2

[10] Wayne P. Power and Johann A. Schoonees, "Understanding background mixture models for foreground segmentation," in *Proceedings of Image and Vision Computing New Zealand*, 2002. 2

[11] J. L. Landabaso, M. Pardàs, and L.-Q. Xu, "Shadow removal with blob-based morphological reconstruction for error correction," in *Proceedings of International Conference on Acoustics, Speech and Signal Processing*, Philadelphia, PA, USA, March 2005, IEEE Computer Society. 2

[12] Ahmed Elgammal, Ramani Duraiswami, David Harwood, and Larry S. Davis, "Background and foreground modeling using nonparametric kernel density for visual surveillance," in *Proceedings of the IEEE*, July 2002, vol. 90, pp. 1151–1163.

[13] P. Felzenszwalb and D. Huttenlocher, "Efficient belief propagation for early vision," *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, vol. 1, pp. I–261–I–268 Vol.1, 2004. 3