# ROBUST 3D SFS RECONSTRUCTION BASED ON RELIABILITY MAPS

*Jaime Gallego, Montse Pardàs*

Technical University of Catalonia (UPC)
{jgallego,montse}@gps.tsc.upc.edu

**ABSTRACT**

*This paper deals with Shape from Silhouette (SfS) volumetric reconstruction in the context of multi-view smart room scenarios. The method that we propose first computes a 2D foreground object segmentation in each one of the views, by using region-based models to model the foreground, and shadow classes, and a pixel-wise model to model the background class. Next, we calculate the reliability maps between foreground and background/shadow classes in each view, by computing the hellinger distance among models. These 2D reliability maps are taken into account finally, in the 3D SfS reconstruction algorithm, to obtain an enhanced final volumetric reconstruction. The advantages of our system rely on the possibility to obtain a volumetric representation which automatically defines the optimal tolerance to errors for each one of the voxels of the volume, with a low rate of false positive and false negative errors. The results obtained by using our proposal improve the traditional SfS reconstruction computed with a fixed tolerance for the overall volume.*

***Index Terms—*** 3D reconstruction, Shape from Silhouette, Multi-view foreground segmentation, reliability maps, SCGMM, region models.

## 1. INTRODUCTION

3D reconstruction from multiple calibrated planar images is a major challenge in the image processing area in order to obtain a realistic volumetric representation of the objects and people under study. In this field, Shape from Silhouette (SfS) gather all the techniques to reconstruct the 3D structure from a set of segmentation masks obtained from multi-view smart-room scenarios. Many of the SfS proposals are based on the Visual Hull concept presented by [11] and based on the 3D geometric modeling, first introduced by [2].

Since the Visual Hull is based on the intersection of the rays that 2D foreground points in each view define in 3D space, these methods are highly dependent on the quality and consistency of the silhouettes obtained in each one of the views since a miss in a view propagates this error into the 3D volume reconstruction.

The most common errors appear due to the presence of shadows and camouflage situations between foreground and background regions. Therefore, there is a clear dependency of the 3D reconstruction with respect to the foreground segmentation, which makes foreground segmentation central to the problem of obtaining a correct volumetric reconstruction.

In this paper we focus on multi-view smart-room sequences recorded by means of $C \equiv \{C_1, \dots C_j, \dots C_M\}$ static color cameras used for a posterior 3D reconstruction. Our objective is to establish a more complete communication between the foreground segmentation process and the 3D reconstruction in order to obtain an enhanced object volume.

### 1.1. Previous work

For several years, many authors have been working in 3D reconstruction techniques that deal with the inconsistency of the silhouettes proposing SfS techniques with enhanced robustness. In these proposals, consistency tests between views and further processing is applied in order to overcome the limitations in the silhouette extraction. [1] uses techniques based on minimization of energy functions based on the local neighborhood structures of 3D elements and smoothing factors. Algorithms based on graph cuts allow to obtain a global minimum of the defined energy function ([9]) with computational efficiency. [6] proposed the Space occupancy grids where each pixel is considered as an occupancy sensor, and the visual hull computation is formulated as a problem of fusion of sensors with Bayesian networks. [10] worked with the Shape from Inconsistent Silhouette by combining the probabilities of each one of the pixels, while [5] proposed a *SfS* using Dempster-Shafer theory, which takes into account the positional relationships between camera pairs and voxels to determine the degree in which a voxel belongs to a foreground object.

Although these techniques increase the computational cost, the results obtained overcome the simple systems that consider the foreground segmentation and the 3D reconstruction as separated steps.

### 1.2. Proposed method

In this paper, we propose a *SfS* system that improves the robustness of the final object volume by including the information obtained from the reliability maps of each one of the views, in the volumetric reconstruction. The proposed system is based on three main processes that are applied for each one of the frames of the sequence under analysis:

*Foreground segmentation:* We use a Bayesian region-based foreground segmentation method (based on [7]) for each one of the views, which combines pixel-wise background model with region-based foreground and shadow models. The advantages of the probabilistic modeling is two-fold: it improves the foreground detection of the objects, thanks to a precise modeling of each class, and, moreover, it allows us to compute the reliability map of each view by comparing the probabilistic models.

*Reliability maps:* For each pixel of the 2D views, we compute the Hellinger distance [3] between the foreground model and the background and shadow models. This distance will give us a $[0, 1]$ bounded value which will be used as an indicator of how reliable the foreground segmentation has been in each one of the views. If we assume a correct modeling for each one of the classes, the similarity between class models gives us information about the reliability of the segmentation obtained: pixels under camouflage or shadow situation between classes will result in small distances.

*3D volume reconstruction:* After computing the foreground segmentation and its reliability for each one of the views, we compute

the Visual Hull reconstruction, but using only the pixels of each view that present enough reliability to be taken into account in the process. *i.e*, working only with those pixels where foreground model is separated in the color domain from the background and shadow models, thus dealing with inconsistent silhouettes obtained under foreground-background camouflage situations or shadow effects.

Our system allows us a reconstruction which automatically determines the optimal tolerance to errors for each one of the voxels of the volume in order to obtain a robust 3D volume of the object, improving the traditional *SfS* reconstruction obtained by defining a fixed tolerance for the overall volume.

The remainder of the paper is organized as follows: Section 2 explains the multi-view foreground segmentation method utilized in each view. Section 3 is devoted to the reliability maps obtained from computing the Hellinger distance between classes, while Section 4 describes the reliable 3D reconstruction. Finally, Section 5 and Section 6 focus on the results and conclusions respectively.

## 2. MULTI-VIEW FOREGROUND SEGMENTATION

Specific probabilistic models are used to represent the foreground and background classes for each one of the camera views $C_j$. Analogously to [7], we use one pixel-wise Gaussian model in the color domain for the background, and two region based models for the foreground and shadow classes: Spatial-Color Gaussian Mixture Model (SCGMM) and Spatial-Color Gaussian Model (SCGM) respectively. The color domain used in the formulation is denoted as $c = (r, g, b)$, while the spatial domain is denoted as $s = (x, y)$. The combination of both color and space domains are defined as the joint domain-range representation $z = (r, g, b, x, y)$.

### 2.1. Background model

The background (bg) model consists of one Gaussian per pixel in the RGB domain. The likelihood of the model is:

$$P(z_i|\text{bg}) = \sum_{k=1}^{N} \frac{1}{N} G_{\text{bg}}(z_i, \mu_{z,k}, \Sigma_{z,k}), \tag{1}$$

where:
$$G_{\text{bg}}(z_i, \mu_{z,k}, \Sigma_{z,k}) = \delta(s_i - \mu_{s,k}) P(c_i|\text{bg}), \tag{2}$$

$z_i \in \mathbb{R}^5$ is the $i$-th pixel value ($i = 1, ..., N$), $\mu_{z,k} \in \mathbb{R}^5$ is the Gaussian mean, $\Sigma_{z,k} \in \mathbb{R}^{5\times5}$ is the covariance matrix, $s_i \in \mathbb{R}^2$ is the spatial pixel's coordinate, $c_i \in \mathbb{R}^3$ is the pixel's color value, $\mu_{s,k}$ is the spatial mean of the $k$-th Gaussian and $P(c_i|\text{bg})$ is the likelihood of each color pixel-wise Gaussian [13]:

$$\begin{aligned} P(c_i|\text{bg}) &= G_{\text{bg}}(c_i, \mu_{c,k}, \Sigma_{c,k}) \\ &= \frac{1}{(2\pi)^{3/2}|\Sigma_{c,k}|^{1/2}} e^{-\frac{1}{2}(c_i-\mu_{c,k})^T \Sigma_{c,k}^{-1}(c_i-\mu_{c,k})}, \end{aligned} \tag{3}$$

where $\mu_{c,k} \in \mathbb{R}^3$ is the color mean of the Gaussian $k$ and $\Sigma_{c,k} \in \mathbb{R}^{3\times3}$ its covariance matrix.

Therefore, we use $N$ color Gaussians, each one centered (in space) at each pixel position ($\mu_{k,s}$) with a zero spatial variance. The initialization of the model is done using training frames, and the updating according to low pass equations for the mean and the variance [13].

### 2.2. Shadow model

We use the Spatial Color Gaussian Model (SCGM) to model the shadow (sh) regions that the object under analysis generates in the scene. The likelihood of the model is:

$$P(z_i|\text{sh}) = G_{\text{sh}}(zi, \mu_z, \Sigma_z) = \frac{e^{-\frac{1}{2}\left[(z_i-\mu_z)^T \Sigma_z^{-1}(z_i-\mu_z)\right]}}{(2\pi)^{5/2}|\Sigma_z|^{1/2}} \tag{4}$$

The initialization of the model is done by analyzing the pixels that accomplish the Color Distortion and Brightness Distortion conditions given in [14]. In each frame, spatial and color mean and variance are updated with the detected shadow pixels.

### 2.3. Foreground model

In order to achieve a correct and precise foreground (fg) modeling in each one of the views, we use the Spatial Color Gaussian Mixture Models (SCGMM). Therefore, for each 2D view, the likelihood of pixel $i$ will be:

$$P(z_i|\text{fg}) = \sum_{k=1}^{K_{\text{fg}}} \omega_k G_{\text{fg}}(z_i, \mu_{z,k}, \Sigma_{z,k}), \tag{5}$$

where $K_{\text{fg}}$ is the number of Gaussians that forms the foreground model and $\omega_k$ is the weighting coefficient.

### 2.4. Bayesian Foreground/Background classification

A pixel $i$ is assigned to the class $l \in \{\text{fg, bg, sh}\}$ that maximizes $P(l_i|z_i) \propto P(z_i|l_i)P(l_i)$.

Analogously to [7, 15, 12], we consider a MAP-MRF framework in order to take into account neighborhood information that can be solved using standard graph-cut algorithm [4].

## 3. RELIABILITY MAPS

In order to get the reliability maps of each camera view: $\gamma^{C_j}$, we propose to analyze the foreground similarity with background and shadow classes for each one of the camera sensors, assuming that:

-High similarity implies that both classes are modeling the same space in a camouflage situation, and thus, the decision is not reliable.

-Low similarity implies classes separated enough to achieve a correct decision.

Hence, for each one of the image pixels $z_i \in I_t$, in each view $C_j$, we propose to compute the Hellinger distance ([3]), in the color $c = (r, g, b)$ domain to detect the degree of similarity between foreground and $l' \in \{\text{bg, sh}\}$ models that each one of the cameras presents:

$$H_i^{C_j}(q_{\text{fg},i}^{C_j}, q_{l',i}^{C_j}) = \sqrt{1 - \text{BC}_i^{C_j}}, \tag{6}$$

where $0 \leqslant H(q_{\text{fg},i}^{C_j}, q_{l',i}^{C_j}) \leqslant 1$, $q_{\text{fg},i}^{C_j}$ and $q_{l',i}^{C_j}$ are the p.d.f.'s that model the $i$-th pixel for the foreground and $l'$ classes respectively in the camera view $C_j$. BC is the Bhattacharyya Coefficient, which is formulated, for a multivariate Gaussian distribution, as follows:

$$\text{BC}_i^{C_j} = \frac{1}{\left(\frac{|\Sigma_i^{C_j}|}{\sqrt{|\Sigma_{\text{fg},i}^{C_j}||\Sigma_{l',i}^{C_j}|}}\right)^{\frac{1}{2}}} e^{-\frac{(\mu_{\text{fg},i}^{C_j} - \mu_{l',i}^{C_j})^T (\Sigma_i^{C_j})^{-1} (\mu_{\text{fg},i}^{C_j} - \mu_{l',i}^{C_j})}{8}}, \tag{7}$$

where $0 \leqslant \text{BC} \leqslant 1$, $\Sigma_{\text{fg},i}^{C_j}$ and $\Sigma_{l',i}^{C_j}$ are the covariance matrices of the models associated to the $i$-th pixel, for the $C_j$-view of the foreground and $l' \in \{\text{bg,sh}\}$ classes respectively. $\mu_{\text{fg},i}^{C_j}$ and $\mu_{l',i}^{C_j}$ are the mean vectors of each class, and $\Sigma_i^{C_j} = \frac{\Sigma_{\text{fg},i}^{C_j} + \Sigma_{l',i}^{C_j}}{2}$.

Note that $H(q_{\text{fg}}, q_{l'}) = 0$ means that foreground and $l'$ models are equal, and thus, strong camouflage situation is present in this pixel, and otherwise $H(q_{\text{fg}}, q_{l'}) = 1$ implies that both models are completely different and there is not similarity between them.

Since the foreground classes are modeled by means of SCGMMs, $q_{\text{fg},i}^{C_j}$ will be chosen according to the Gaussian $k$ that maximizes the probability of the $i$-th pixel under analysis for each view:

$$P(k|z_i, \text{fg}) = \frac{\omega_k G_{\text{fg}}(z_i, \mu_k, \sigma_k)}{\sum_k \omega_k G_{\text{fg}}(z_i, \mu_k, \sigma_k)} \tag{8}$$

In the case of the background, since we have defined a pixel-wise model, $q_{\text{bg},i}^{C_j}$ will be directly obtained from the background Gaussians associated to this pixel. For the shadow class, $q_{\text{sh},i}^{C_j}$ is the SCGM used to model the shadow projected by the person. The foreground-shadow reliability will be utilized only over the spatial region modeled by the shadow Gaussian, since it is the only region affected by the shadow effects.

Therefore, for each one of the pixels of $C_j$, we will obtain the final reliability value $\gamma_i^{C_j}$, according to the comparison between fg-bg and fg-sh models:

$$\gamma_i^{C_j} = \begin{cases} \min\left[H_i^{C_j}(q_{\text{fg},i}^{C_j}, q_{\text{bg},i}^{C_j}),\ H_i^{C_j}(q_{\text{fg},i}^{C_j}, q_{\text{sh},i}^{C_j})\right] & \rightarrow \quad \text{sh region} \\ H_i^{C_j}(q_{\text{fg},i}^{C_j}, q_{\text{bg},i}^{C_j}) & \rightarrow \quad \text{otherwise} \end{cases} \tag{9}$$

where the most restrictive distance between fg-bg and fg-sh is chosen in the regions belonging to the spatial shadow model, and the distance between fg-bg in the rest of the image.

## 4. ROBUST 3D RECONSTRUCTION

The concept of Visual Hull (VH) is strongly linked to the one of silhouettes consistency. Total consistency hardly ever happens in realistic scenarios due to inaccurate calibration or wrong silhouettes caused by errors during the 2D foreground detection process. Because of that, some *SfS* methods have been designed in the past assuming that the silhouettes can not be consistent, thus adding a tolerance to error ($\tau$) in the number of views necessary to consider a voxel as occupied. Hence, adding error tolerance to the 3D reconstruction, the estimate of the visual hull is conservative in the sense of assuming that $\tau$ foreground under-segmentation errors can occur. This approach will lead to reduce the number of false negative errors although losing precision in the final reconstructed volume.

We propose a *SfS* reconstruction method based on the silhouette reliability principle. Our system validates the regions in the silhouettes which are reliable and uses only these regions of each view to compute the robust Visual Hull of the object, thus dealing with 2D errors.

The robust shape from silhouette algorithm that we propose is shown in Algorithm 1, where the projection test (PT) consists in testing the central pixel within the splat of the voxel in camera $Cj$. Once the projection Test has been carried out, we can use the voxel-pixels correspondence to check the reliability that each one of the pixels present. The Reliability Test (RT) checks for the pixel that appear in the voxel's projection in each view $C_j$, the reliability value $\gamma_i^{C_j}$.

**Algorithm 1** Reliable Shape from Silhouette algorithm

**Require:** : Silhouettes: S(c), Reliability Test: RT(voxel, camera), Projection Test: PT (voxel, silhouette)

```
for all voxel do
    voxel ← Foreground
    for all cameras do
        if PT (voxel, S(c)) is false and RT(voxel, camera) > R_th then
            voxel ← Background
        end if
    end for
end for
```

We define the Reliability threshold $R_{\text{th}}$ as a value $0 < R_{\text{th}} < 1$ which will determine the minimum reliability value to consider the pixels in the final reconstruction process. In our experiments, we have tested that a reliability factor $R_{th} = 0.7$ yields correct results in the final reconstruction process.

This 3D reconstruction is equivalent to define an optimal error tolerance value $\tau$ for each one of the voxels of the image, improving the precision of the volume in those regions where no tolerance is necessary, while reducing the false negative errors.

## 5. RESULTS

We have evaluated our proposal by analyzing four multi-view sequences, of the database presented in [8], which present strong difficulties to achieve a correct 3D reconstruction due to the similarity between some foreground regions and the background. These sequences have been recorded with different acquisition setups in order to better analyze the effect of the errors tolerance in the volumetric reconstruction:

Figure 1 displays the results obtained in these four sequences recorded with 18 cameras (first row), 16 cameras (second and fourth rows) and 8 cameras (third row). One representative view of the overall multi-view sequence has been selected in each case.

The qualitative evaluation is done comparing the volumetric reconstruction results with the ones obtained by using the Visual Hull reconstruction with different tolerance to errors ($\tau$). The segmentation masks used in all these reconstructions are the ones obtained with the 2D segmentation exposed in Section 2. These segmentation results are displayed in the second column. As we can observe, the resultant foreground segmentation presents a low ratio of false positive and false negative detections although some false negative errors are present in some of the views, due to the foreground-background camouflage problem and the presence of shadows.

In the third column we can see the spatial representation of the projected foreground model. Each ellipse represents one Gaussian of the foreground model, and are colored with the mean color that each distribution is modeling.

From fourth to sixth column, we can observe the different 3D volumes that we can obtain using the Visual Hull reconstruction with different tolerance to errors. When we do not use any tolerance to errors ($\tau = 0$) (Fourth column), any false negative error that appears in the 2D segmentation is propagated to the final 3D volume, thus generating critical false negative errors in the resultant reconstruction. When using tolerance to errors in fifth and sixth column, we reduce significantly the propagation of the false negative errors to

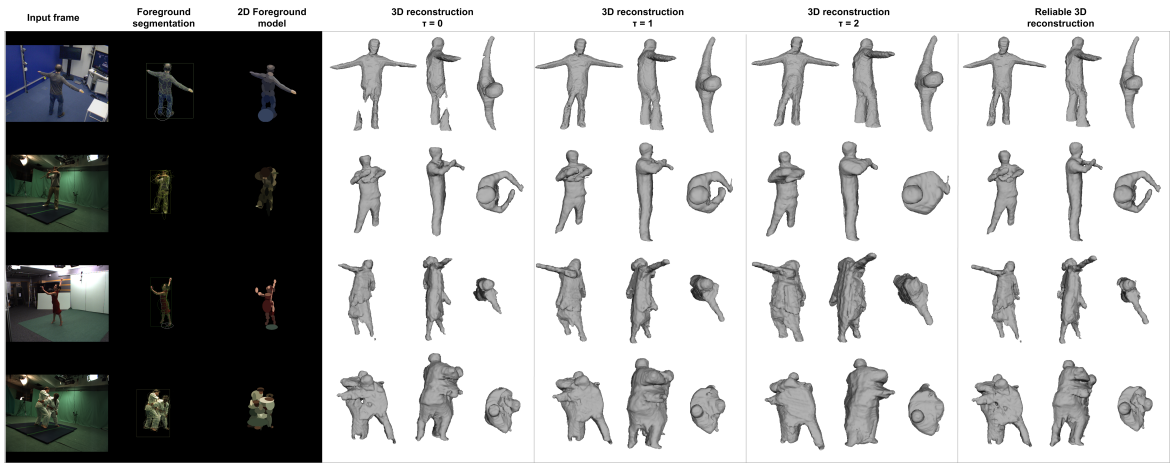


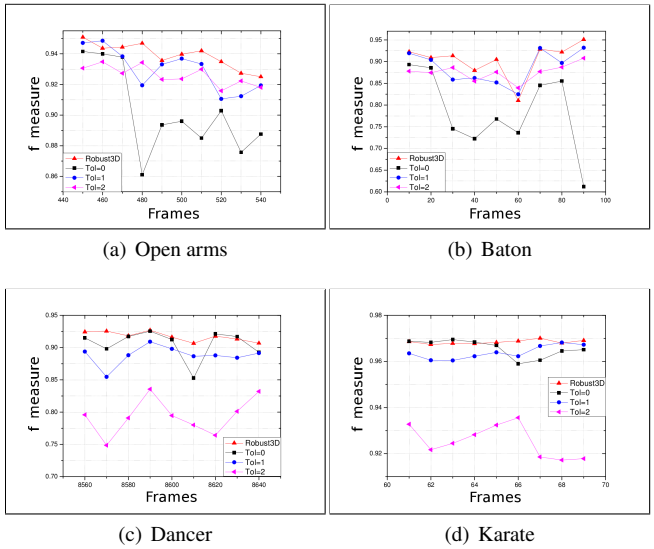

**Fig. 1**. Foreground segmentation and 3D volume reconstruction results. From left to right: original view; Bayesian foreground segmentation proposed in the paper: Color ellipses correspond to the Gaussians of the foreground model, white ellipse corresponds to the spatial representation of the shadow model; Foreground model colorized with the mean color that are modeling; the projected volume computed with tolerance $\tau = 0$; volume with $\tau = 1$; volume with $\tau = 2$; Robust 3D reconstruction using our method.

the 3D space, although losing precision in the volumetric reconstruction, thus obtaining a coarse representation of the object. Our system (seventh column) achieves a 3D reconstruction that only applies the tolerance to errors in those pixels where the reliability between foreground and background and shadow classes is low, thus reducing the propagation of those errors to the 3D space. As we can see, our system achieves an object reconstruction that presents similar precision than the Visual Hull reconstruction without tolerance ($\tau = 0$), but solving the false negative errors.

Finally, quantitative results of these sequences are displayed in Figure 2, where we use the projection of the volume to a 2D view in order to compute the $f_{\text{measure}}$. We compute this metric over equally distributed frames on parts of the sequences that present special difficulty due to the foreground-background similarity. As we can see, our proposal (in red color), achieves a volumetric reconstruction that adapts better to the circumstances of the sequence under analysis than the reconstructions with fixed tolerance. Our method maintains a high $f_{\text{measure}}$ value for the sequences under study, maintaining the precision of the volumetric reconstruction while reducing the false negative detections. More qualitative and quantitative results will be accessible in our web page [1]. The computational cost of our system is 0.1 frames/second analyzing a standard sequence and using an Intel Core2 Duo 3GHz processor and 20 GB RAM.

## 6. CONCLUSIONS

We have introduced in this paper a novel multi-view segmentation and 3D reconstruction system. To this end, we have proposed a robust Visual Hull reconstruction that uses the reliability of the pixels to avoid those views where the pixels detected as background, present high similarity between foreground, background and shadows models. Although the system is highly dependent on the foreground segmentation model and how it represents the foreground object in each one of the views, our approach achieves better accuracy of the reconstructed volume while reducing the critical misses that appear in a direct 3D reconstruction with $\tau = 0$, and reducing the false positive regions that appear if we decide to use a direct $\tau = 1$ or $\tau = 2$ reconstruction.




(a) Open arms (b) Baton

(c) Dancer (d) Karate

**Fig. 2**. Quantitative evaluation. $f_{\text{measure}}$ for each sequence.

## 7. REFERENCES


[1] M. Alcoverro and M. Pardas. Voxel occupancy with viewing line inconsistency analysis and spatial regularization. *Int. Conf. on Computer Vision Theory and Applications*, pages 464–469, 2009.

[2] B. G. Baumgart. Geometric modeling for computer vision. Technical report, Ph.D. thesis, CS Stanford University Document, 1974.

[3] R. Beran. Minimum hellinger distance estimates for parametric models. *The Annals of Statistics*, 5(3):445–463, 1977.

[4] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 23(11):1222–1239, 2001.


[1] http://www.jaimegallego.com.es/icip2014_hellinger_3d

[5] L. Díaz-Más, R. Muñoz-Salinas, F. Madrid-Cuevas, and R. Medina-Carnicer. Shape from silhouette using dempster–shafer theory. *Pattern Recognition*, 43(6):2119–2131, 2010.

[6] J.-S. Franco and E. Boyer. Fusion of multiview silhouette cues using a space occupancy grid. In *Int. Conf. on Computer Vision*, volume 2, pages 1747–1753. IEEE, 2005.

[7] J. Gallego, M. Pardàs, and G. Haro. Enhanced foreground segmentation and tracking combining bayesian background, shadow and foreground modeling. *Pattern Recognition Letters*, 2012.

[8] INRIA. 4D Repository. http://4drepository.inrialpes.fr/.

[9] V. Kolmogorov and R. Zabih. Multi-camera scene reconstruction via graph cuts. *European Conf. on Computer Vision*, pages 8–40, 2002.

[10] J. Landabaso and M. Pardàs. Cooperative background modelling using multiple cameras towards human detection in smart-rooms. In *In Proc. of European Signal Processing Conference*, 2006.

[11] A. Laurentini. The visual hull: A new tool for contour-based image understanding. *Proc. 7th. Scandinavian Conf. on Image Analysis*, pages 993–1002, 1991.

[12] Y. Sheikh and M. Shah. Bayesian modeling of dynamic scenes for object detection. *IEEE trans. on Pattern Analysis and Machine Intelligence*, 27(11):1778–1792, 2005.

[13] C. Wren, A. Azarbayejani, T. Darrell, and A. Pentland. Pfinder: Real-time tracking of the human body. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 19(7):780–785, 1997.

[14] L. Xu, J. Landabaso, and M. Pardas. Shadow removal with blob-based morphological reconstruction for error correction. In *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing.*, volume 2, 2005.

[15] T. Yu, C. Zhang, M. Cohen, Y. Rui, and Y. Wu. Monocular video foreground/background segmentation by tracking spatial-color Gaussian mixture models. In *IEEE Workshop on Motion and Video Computing*, pages 5–5, 2007.