A physically motivated pixel-based model for background subtraction in 3D images

M. Braham, A. Lejeune and M. Van Droogenbroeck

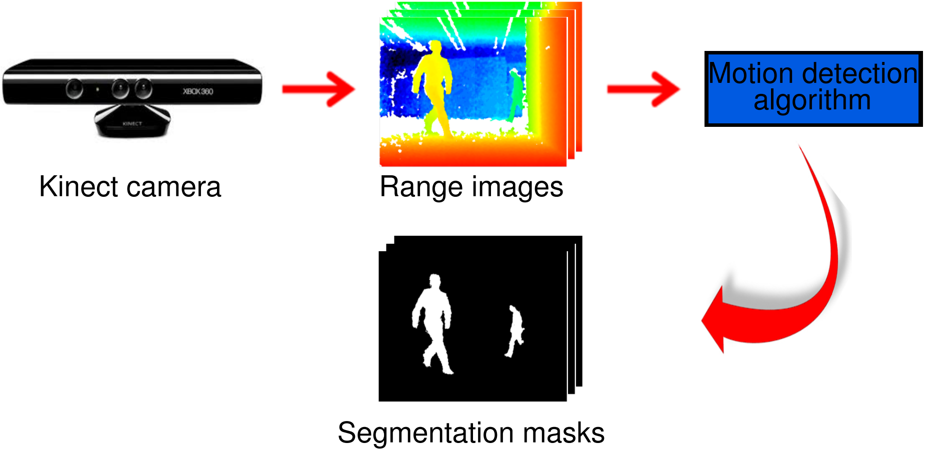
INTELSIG, Montefiore Institute, University of Liège, Belgium

IC3D - December 10, 2014

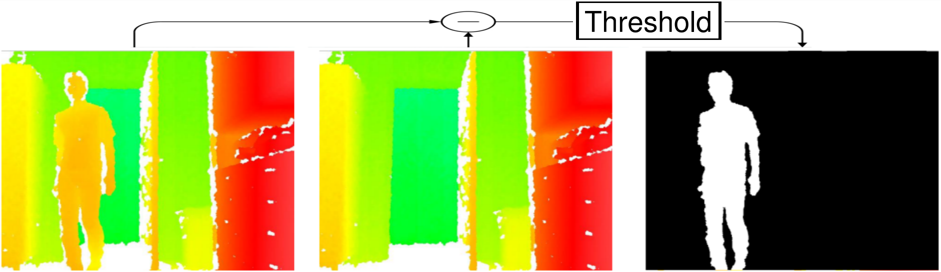
Outline

1. Introduction
   - Topic of this work
   - Background subtraction: principle
2. Background subtraction in range images
   - Advantages, opportunities and challenges
   - Related work
3. Proposed technique
   - Towards a hybrid background model
   - Considering holes in one model
   - Depth-based background model
   - Post-processing
4. Experimental results
   - Benchmarking: dataset and algorithms
   - Qualitative results
   - Comparison of methods in the ROC space
5. Conclusion

Topic of this work: real-time motion detection in a sequence of range images



Motion detection through background subtraction



Current frame Background model Output binary mask

Main questions

- How to model the background ?
- How to initialize and update the background model ?
- How to classify pixels?

Background subtraction in range images
Advantages, opportunities and challenges

Advantages of range images (when compared to color images)

- Insensitive to lighting changes (in a first approximation)
- Insensitive to the true colors of objects

Opportunity

The physical meaning of the depth signal can be leveraged to improve the foreground segmentation.

Challenges

- Holes
- Non-uniform spatial distribution of noise

Background subtraction in range images
Related work

- Most of the work for motion detection is dedicated to color imaging.
- RGB-D background subtraction techniques focus on the combination of depth and color, not on the depth signal.
- Researchers apply almost exclusively basic methods (static background, exponential filter, ...) or well-known color-based methods (GMM, ViBe, ...) to range images.
- To the best of our knowledge, only one motion detection algorithm is tailored for depth imaging:

  del-Blanco *et al.*, "Foreground segmentation in depth imagery using depth and spatial dynamic models for video surveillance applications", January 2014.

Characteristics of our background model

Our background model is:

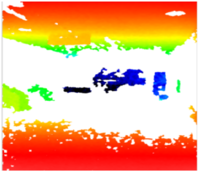
- Pixel-based
- Physically motivated
- Hybrid:
    - Model of *constant holes*
    - Depth-based background model

Definition

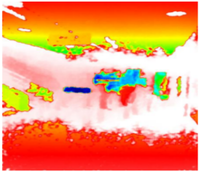
A *constant hole* is a pixel for which the Kinect camera is unable to measure depth when the background is not occluded by a foreground object.

Relevance of a hybrid background model

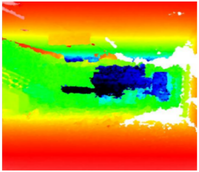
Color image[1]



Depth map

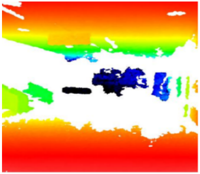


Model of Pfinder[2]



Depth-based model

Constant holes



Hybrid model

[1] Taken from an existing database: Spinello *et al.*, "People detection in RGB-D data", 2011
[2] Wren *et al.*, "Pfinder: Real-time tracking of the human body", 1997

Analysis of the dynamics of holes

- Use of N counters C_i (N = number of pixels) and two global heuristic parameters N_H and T_W with $N_H \ll T_W$.

Definition

$C_i = k$ indicates that the last depth value in pixel i was observed at frame $t - k$.

Identification of a constant hole

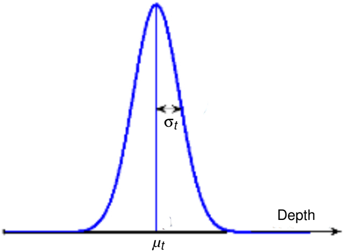
$C_i \geq N_H \Rightarrow$ pixel i is labeled as a constant hole.

Reset of a constant hole

$C_i < N_H$ during at least T_W frames $\Leftrightarrow$ pixel i switches from the state *constant hole* to the state *standard pixel*.

Unimodal Gaussian depth-based model

- Parametric model
- Only two parameters memorized for each pixel: μ_t and σ_t.



Depth-based background model: gaussian pdf

- μ_t updated with a physical interpretation of the depth signal.
- σ_t updated according to a law defined by the sensor noise.

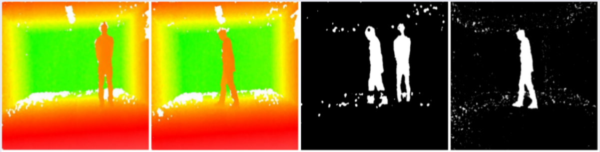
Physical interpretation of the depth signal

Background is always located behind foreground !

Physically motivated updating strategy of the mean μ_t.

$\mu_t \approx MAX\left(D_k\right)$ for $k \in [0, t]$, where D_k denotes the measured depth at time k.

Ghosts challenge solved !



Initial background | Lateral moving | Mask of PBAS[3] | Our mask

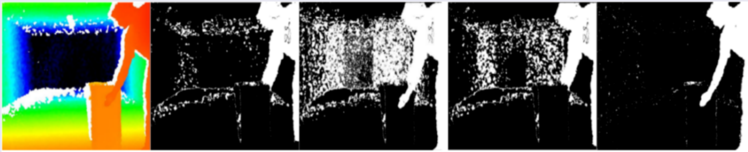
[3] Hofmann *et al.*, "Background segmentation with feedback: The pixel-based adaptive segmenter", 2012

Depth-dependent BG/FG decision threshold

The noise of the Kinect depth sensor is depth-dependent. The spatial distribution of noise in range images is thus non-uniform.

- We use Khoshelham's relationship[3] to update the standard deviation: $\sigma_t = K_{kinect}\mu_t^2$
- Our BG/FG decision threshold τ_t is thus depth-dependent: $\tau_t = K\sigma_t = KK_{kinect}\mu_t^2$

Consequence: reliable segmentation for all depth values



Range image | Sakbot[4] (high τ) | Sakbot (low τ) | Sakbot (medium τ) | Our algorithm

[3] Khoshelham, "Accuracy analysis of Kinect depth data", 2011

[4] Cucchiara *et al.*, "Detecting moving objects, ghosts, and shadows in video streams", 2003

Kinematic constraint on foreground objects

The updating equation $\mu_t \approx MAX(D_k)$ for $k \in [0, t]$ removes ghosts after one frame.
$\rightarrow$ How can we eliminate ghosts instantaneously?

Kinematic constraint

The maximum depth jump of the foreground between two consecutive frames is upper bounded by:

$$\triangle P_{max} = \frac{V_{max}}{Fr}$$

where V_{max} is the maximum speed of foreground objects and Fr the frame rate of the camera.

Improved BG/FG classification process

- $\mu_t + K\sigma_t + \triangle P_{max} < D_t \Rightarrow BG$
- $\mu_t + K\sigma_t < D_t \leq \mu_t + K\sigma_t + \triangle P_{max} \Rightarrow FG$

$\rightarrow$ Ghosts are generally removed instantaneously.

Summary of the depth-based background model

Definitions

L_t and H_t are respectively defined by $\mu_t - K\sigma_t$ and $\mu_t + K\sigma_t$.

Updating equations and classification process					
	0	L_t	$K\sigma_t$ μ_t H_t	$H_t + \Delta P_{max}$	$Depth$
Condition	$D_t = 0$ (hole)	$0 < D_t < L_t$	$L_t \leq D_t \leq H_t$	$H_t < D_t \leq H_t + \Delta P_{max}$	$H_t + \Delta P_{max} < D_t$
μ_{t+1}	μ_t	μ_t	$(1-\alpha)\mu_t + \alpha D_t$	D_t	D_t
σ_{t+1}	σ_t	σ_t	$K_{kinect}\mu_{t+1}^2$	$K_{kinect}\mu_{t+1}^2$	$K_{kinect}\mu_{t+1}^2$
Class	BG	FG	BG	FG	BG
Initialization process					
$\mu_0 = D_0$			$\sigma_0 = K_{kinect}\mu_0^2$		

- Recursive filter on μ_t to enhance the estimation of the real background depth
- Sleeping foreground is not absorbed in the background
- Semi-conservative updating strategy

Post-processing filters

1. Background model controller
2. Morphological opening with a 3x3 cross as structuring element.
3. 7x7 median filter

Benchmarking: dataset and algorithms

To evaluate the performances of the proposed technique, we have built a new dataset:

- 8 depth maps sequences acquired with a Kinect camera: 3 sequences taken from an existing depth-based database + 5 sequences representing various challenges.
- 220 ground-truths have been labeled manually at the rate of one ground-truth image per 25 frames for each sequence.

We compare our results with those of 4 algorithms:

- 2 very popular Gaussian mixtures: GMM-STAUFFER[1] and GMM-ZIVKOVIC[2]
- 2 state-of-the-art algorithms for color videos: SOBS[3] and PBAS[4]

---

[1] Stauffer *et al.*, "Adaptive background mixture models for real-time tracking", 1999

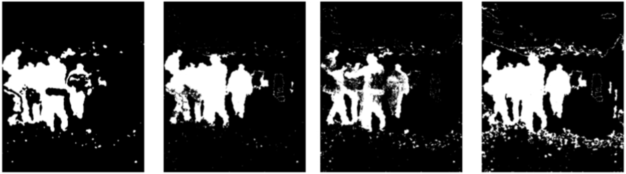
[2] Zivkovic *et al.*, "Efficient adaptive density estimation per image pixel for the task of background subtraction", 2006

[3] Maddalena *et al.*, "A self-organizing approach to background subtraction for visual surveillance applications", 2008

[4] Hofmann *et al.*, "Background segmentation with feedback: The pixel-based adaptive segmenter", 2012

Qualitative results

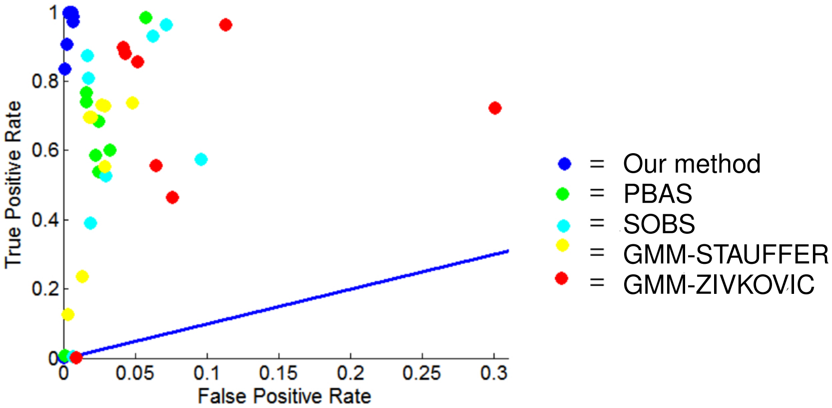
Range image[1] Ground-Truth Our algorithm



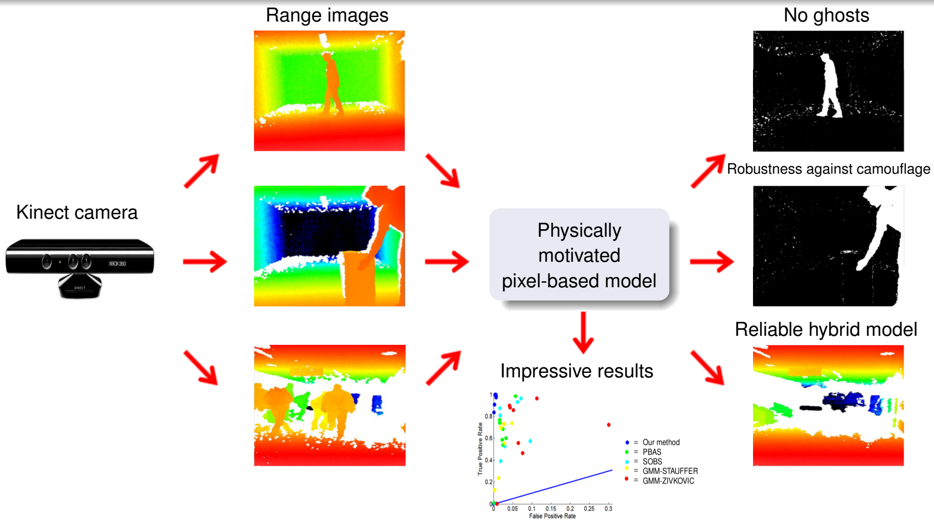
PBAS SOBS GMM-STAUFFER GMM-ZIVKOVIC

[1] Taken from an existing database: Spinello *et al.*, "People detection in RGB-D data", 2011

Comparison of methods in the ROC space



Conclusion


Range images
No ghosts
Robustness against camouflage
Kinect camera
Physically motivated pixel-based model
Reliable hybrid model
Impressive results
True Positive Rate
False Positive Rate
Our method
PBAS
SOBS
GMM-STAUFFER
GMM-ZIVKOVIC