- INVITED ARTICLE -

Background Subtraction for Time of Flight Imaging

Javier Giacomantone, Lucía Violini, and Luciano Lorenti

Institute of Research in Computer Science - School of Computer Science - University of La Plata, Argentina
{jog, lviolini, llorenti}@lidi.info.unlp.edu.ar

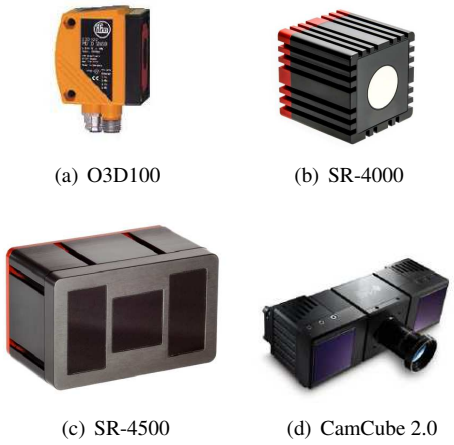
Abstract

A time of flight camera provides two types of images simultaneously, depth and intensity. In this paper a computational method for background subtraction, combining both images or fast sequences of images, is proposed. The background model is based on unbalanced or semi-supervised classifiers, in particular support vector machines. A brief review of one class support vector machines is first given. A method that combines the range and intensity data in two operational modes is then provided. Finally, experimental results are presented and discussed.

Keywords: Industrial TOF Cameras, Machine Vision, Pattern Recognition, Support Vector Machines

1 Introduction

Background subtraction methods are used in machine vision to detect foreground objects in image sequences [1]. The background representation is called background model and methods involving different models have been proposed over recent years [2][3][4]. A broad distinction can be considered between parametric and nonparametric models if an assumption about the intensity probabilistic distribution of background pixels is made or not [5]. Furthermore, a model can be divided into pixel based or block-based. Although the former are more common, the latter can be used to model distributions based on data of adjacent pixel blocks. Non-stationary background modelling due to overall lighting conditions and local image pattern fluctuations have been proposed for background subtraction. A common assumption in background modeling is that only initial background data is available to train a two class pattern recognition system [6]. Time of flight (TOF) cameras are an alternative to other sensing techniques like laser scanners or image based stereovision. Based on CMOS technology and time of flight principle, provide range distance information in real time and intensity images. By illuminating the scene with near infrared light and measuring the phase shift of the reflected light for each pixel they provide depth information, $21/2D$ images at fast rates [7][8]. Sensibility to external light sources, calibration and noise are important factors to consider in background modelling based on TOF images [9][10]. Figure 1 shows four different TOF phase based camera models.



(a) O3D100 (b) SR-4000

(c) SR-4500 (d) CamCube 2.0

Figure 1: a)Ifm electronic©, b)HEPTAGON™, c)HEPTAGON™, d)PMD[Vision] ®

Glazer et al. [11] proposed a nonparametric block based model for intensity images, unbalanced classes and support vector machine classification to perform background subtraction. Cho et. al [12] proposed a probabilistic background model for both intensity and range images. The obtained foreground models are combined and morphological filters are used to remove noise. This paper presents a computational background subtraction method for TOF images. The background model assumes unbalanced training data and is based on one class support vector machines (OCSVM). A preliminary version of this paper presented OCSVM to operate with both images simultaneously [13].

This paper is organized as follows: Section 2 briefly describes the main ideas behind one class support vector machines. Section 3 presents the proposed background subtraction method. In section 4 experimental results are presented. Finally, conclusions are given in section 5.

Citation: J. Giacomantone, L. Violini and L. Lorenti. *"Background Subtraction for Time of Flight Imaging"*. Journal of Computer Science & Technology, vol. 17, no. 2, pp. 135-139, 2017.

2 Support vector machines

Support Vector Machines (SVM), based on statistical learning theory [14], are very useful classification tools in machine vision. In this section, we briefly review the notation and the classification method. Consider a set of l training patterns from two classes $\{(x_i, y_i), i = 1, ..., l\}$, where $x_i \varepsilon \mathbb{R}^N$ is an N dimensional vector and $y_i \varepsilon \{1, -1\}$ is its class label. Consider a data set $\mathbf{x}_1, \ldots, \mathbf{x}_l \in \mathscr{X}$ where $l \in \mathbb{N}$ is the number of observations and $\mathscr{X}$ is a compact subset of $\mathbb{R}^N$. A SVM determines a linear classifier, a hyperplane, that maximizes the distance between the two classes. Mathematically, it is equivalent to maximizing the margin magnitude $\frac{2}{||\mathbf{w}||}$ subject to $y_i((x_i.w) + b) \geq 1$, where b and w determine the hyperplane $(w.x) + b = 0$. In order to classify nonseparable samples, training errors are accepted by introducing slack variables $\xi_i \geq 0$. The hyperplane is then obtained by minimizing:

$$\begin{aligned} \min \quad & \tfrac{1}{2}||\mathbf{w}||^2 + C\textstyle\sum_i \xi_i \\ \text{s.t.} \quad & y_i[(w - x_i) + b]) \geq 1 - \xi_i \end{aligned} \tag{1}$$

where C controls the tradeoff between the hyperplane complexity and training error, with a small C value corresponding to a large margin. The final decision function is obtained by solving the Lagrangian dual of eq. (1),

$$f(\mathbf{x}) = \text{sgn}[\sum_i \alpha_i y_i (x.x_i) + b]$$

where α_i are Lagrange multipliers. A SVM classifier can be extended to nonlinear classification by using kernel methods. Kernel methods project data into a higher dimensional feature space to obtain a linear decision function equivalent to the nonlinear in the original space. The kernel used in this paper is the Radial Basis Function (RBF) defined as $k(x, x_i) = e^{-\gamma \|x - x_i\|^2}$, where γ is the width parameter, a small γ value corresponds to a large kernel width.

There are two main one class classification algorithms based on SVM, support vector data description [15] and one class SVM (OCSVM) [16]. The one-class SVM estimates a function f that is positive for a subset of the sample space and negative for the complementary. The algorithm maps the data into a feature space corresponding to the kernel and separates them from the origin with maximum margin. Let $\Phi : \mathscr{X} \to \mathscr{F}$ be a feature map, that is, a map into an inner product space $\mathscr{F}$ such that the inner product in the image of Φ can be computed by evaluating a simple kernel.

$$k(\mathbf{x}, \mathbf{z}) = (\Phi(\mathbf{x}) \cdot \Phi(\mathbf{z}))$$

It can be formulated as an optimization problem.

$$\begin{aligned} \min_{\mathbf{w} \in \mathscr{F}, \xi \in \mathbb{R}^l, \rho \in \mathbb{R}} \quad & \tfrac{1}{2}||\mathbf{w}||^2 + \tfrac{1}{vl}\textstyle\sum_i \xi_i - \rho \\ \text{s.t.} \quad & (\mathbf{w} \cdot \Phi(\mathbf{x}_i)) \geq \rho - \xi_i \end{aligned} \tag{2}$$

where $\xi_i \geq 0$ and $\nu \in (0, 1]$ is a parameter controlling the penalized term and ξ_i are slack variables. By solving the optimization problem (2) we obtain $\mathbf{w}$ and ρ and the decision function is -1 for outliers in the data set and +1 for the rest of the samples in the data set.

$$f(\mathbf{x}) = \text{sgn}(\mathbf{w} \cdot \Phi(\mathbf{x}) - \rho) \tag{3}$$

Introducing Lagrangian multipliers $\alpha_i, \beta_i \geq 0$, we obtain

$$\begin{aligned} L(\mathbf{w}, \xi, \rho, \alpha, \beta) \quad & = \tfrac{1}{2}||\mathbf{w}||^2 + \tfrac{1}{vl}\textstyle\sum_i \xi_i - \sum_i \beta_i \xi_i \\ & \quad -\rho - \textstyle\sum_i \alpha_i (\mathbf{w} \cdot \Phi(\mathbf{x}) - \rho + \xi_i) \end{aligned}$$

Setting the derivatives with respect to the primal variables $\mathbf{w}, \xi, \rho$ equal to zero yields

$$\mathbf{w} = \textstyle\sum_i \alpha_i \Phi(\mathbf{x}_i),$$

$$\alpha_i = \tfrac{1}{vl} - \beta_i \leq \tfrac{1}{vl},$$

$$\textstyle\sum_i \alpha_i = 1.$$

The decision function can be written as

$$f(\mathbf{x}) = \text{sgn}(\sum_i \alpha_i k(\mathbf{x}_i, \mathbf{x}) - \rho)$$

The multipliers α_i can be solved from the dual problem:

$$\begin{aligned} \min_{\alpha} \quad & \tfrac{1}{2}\textstyle\sum_{ij} \alpha_i \alpha_j k(\mathbf{x}_i, \mathbf{x}_j) \\ \text{s.t.} \quad & 0 \leq \alpha_i \leq \tfrac{1}{vl},\ \textstyle\sum_i \alpha_i = 1. \end{aligned}$$

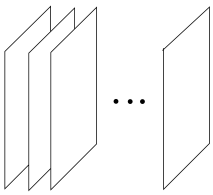
The parameter ρ can be recovered by exploiting that for any such α_i and the corresponding pattern $\mathbf{x}_i$ satisfies:

$$\rho = (\mathbf{w} \cdot \Phi(\mathbf{x}_i)) = \sum_j \alpha_i k(\mathbf{x}_i, \mathbf{x}_j). \tag{4}$$

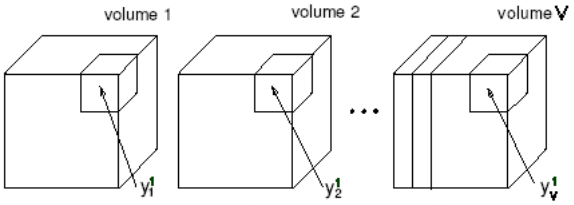
3 Background subtraction

The proposed method consists of two operational modes. Each one has two stages. The first stage trains a OCSVM classifier with background data. The second stage classifies each pixel or voxel as a member of the background or the foreground. Both operational modes are based on adequately combining depth and intensity TOF data. The operational mode 2 works with arrays of volumes of TOF images. Figure 2 shows the data structure for both operational modes. Let TOF_r and TOF_g be two arrays of depth images and intensity images respectively. Each one consists of l images

$21/2D$ and l images $2D$. The parameters of the background model are obtained by the procedure described in section 2. Let l be the number of training vectors, $k = 1...l$, for each pixel (i, j), where:



(a) Mode 1



(b) Mode 2

Figure 2: (a) Training images, (b) V volumes

$$\mathbf{x}_{i,j}^{k} = [x_1, x_2]^T$$

is the k-th training vector for pixel the pixel (i, j) where:

$$x_1 = TOF_r^k(i, j) \quad and \quad x_2 = TOF_g^k(i, j)$$

After the reference background model is defined depth and intensity images are obtained and each pixel is classified as belonging to the background or the foreground by the equation:

$$f(\mathbf{x}_{i,j}) = \text{sgn}(\mathbf{w} \cdot \Phi(\mathbf{x}_{i,j}) - \rho) \qquad (5)$$

The second operational mode operates with large data sets. l volumes, $v = 1...l$, of image arrays, TOF_r and TOF_g are the training data of the first stage. M images of depth an intensity, respectively formed each volume. The OCSVM parameters of the background model are defined from V vectors, $v = 1...l$, for each volumen element (i, j, k), where:

$$\mathbf{x}_{i,j,k}^{v} = [x_1, x_2]^T$$

is the v-th training vector for voxel (i, j, k) where:

$$x_1 = TOF_r^v(i, j, k) \quad and \quad x_2 = TOF_g^v(i, j, k)$$

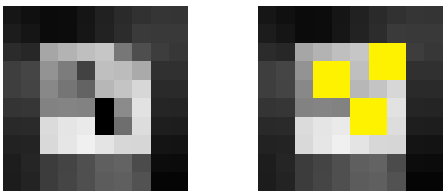
The second stage of mode 2 determines the class of each voxel by:

$$f(\mathbf{x}_{i,j,k}) = \text{sgn}(\mathbf{w} \cdot \Phi(\mathbf{x}_{i,j,k}) - \rho) \qquad (6)$$

A tradeoff between available memory, complexity and speed depends on the camera characteristics and the particular conditions of a problem.

4 Experimental Results

In order to test and compare results of the proposed method three main TOF data sets were evaluated. The first TOF set is a completely synthetic one with different levels and noise statistical distributions. These types of experiments provide controlled conditions and knowledge of the exact foreground object location, namely a gold standard. The second alternative is to generate artificial foci of activated pixels or voxels in real background TOF data sequences. The third possible data set involves working with real TOF background and foreground objects. Figure 3 depicts one synthetic image and foreground objects in a pseudo-color image. The synthetic TOF data are formed by 10x10x3 voxels per volume and 20 volumes. Voxels values were 16000 corrupted by zero-mean Gaussian noise with standard deviation 4000. Foreground voxels had their values increased between 600 and 1500 in 100 steps. Images TOF1 to TOF10 were generated with Gaussian noise and TOF11 to TOF20 with Rayleigh noise.

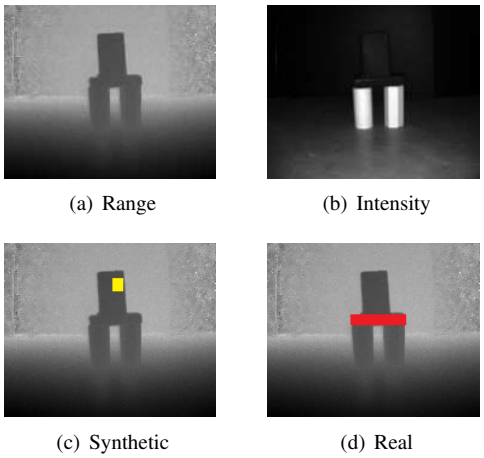


(a) Synthetic (b) Pseudo-color

Figure 3: TOF data

Images for real and semi-synthetic data sets were obtained using the MESA SwissRanger SR4000 time-of-flight camera [17][18]. The SR4000 is an active measurement device that obtains depth and intensity images of 144x176 pixels up to 50fps as shown in figures 4(a) and 4(b). The operational range is between 0,8 to 8m. Figure 4(c) and 4(d) are depth images with synthetic and real foreground objects shown in pseudo-color. TOF depth measurements at high rates can be noisy and display errors where scenes contain materials with low infrared reflectivity [19][20]. Another error known as flying pixel, could appear in depth image levels between the foreground and the background [21][22].

Table 1 presents comparative results. The values of the true positive fraction (TPF) and the false negative fraction (FPF) correspond to the optimal operating point of the receiving operating characteristics (ROC). The



(a) Range (b) Intensity

(c) Synthetic (d) Real

Figure 4: Time of flight images

optimal operating point distance and the area under the curve are indicated for each data set. C1 is the method proposed in [12] and C2 also includes morphological filters. $M2_{AF}$ refers to a test of the proposed method in mode 2, and anisotropic diffusion filtering [23]. Figure 5 presents ROC curves corresponding to TOF9, TOF12 and TOFS.

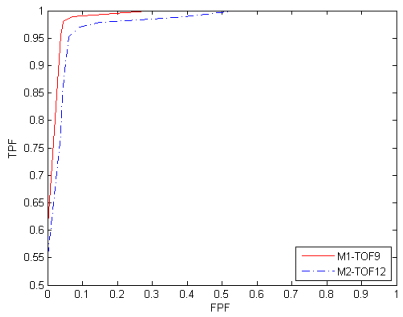
	Area	d_{po}	TPF_{po}	FPF_{po}
TOF9				
$M1_{v=0.37}$	0.989	0.661	0.979	0.046
TOF12				
$M2_{v=0.4}$	0.977	0.693	0.892	0.049
$M2_{AF}$	0.985	0.639	0.946	0.045
TOFS				
$M1_{v=0.38}$	0.986	0.645	0.955	0.043
C_1	0.954	0.573	0.931	0.121
C_2	0.975	0.618	0.935	0.060

Table 1: Performance metrics

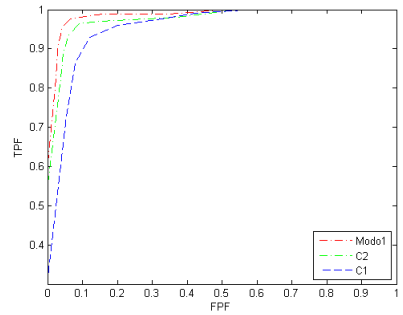
Figure 6 shows a real range scene and the result of the proposed method for background subtraction.

5 Conclusions

In this paper, a computational method for background subtraction of time of flight imaging, have been proposed. The background model is constructed by using depth and intensity data to train a one class support vector classifier. The second stage of the method is a pixel based background scheme that classifies each pixel. Experimental results show that the method can combine depth and intensity data to build an effective model suitable for real time applications. Future work should consider evaluating the specific performance of the method for more extensive experimental scenarios involving noise filtering and the overhead influence in real time.

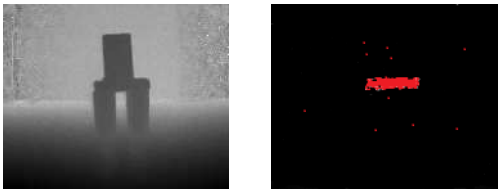


(a) TOF



(b) TOFS

Figure 5: ROC curves



(a) Depth image (b) Subtraction

Figure 6: TOFR

Competing interests

The authors have declared that no competing interests exist.

References

[1] P. Rajan and S. Prakash, "Moving foreground object detection and background subtraction using adaptive-K GMM: A survey," *International Journal of Advance Research in Computer Science and Management Studies*, vol. 2, pp. 300–308, 2014.

[2] M. Piccardi, "Background subtraction techniques: a review," in *IEEE International Confer-*

ence on System. Man and Cybernetics, pp. 3099–3104, 2004.

[3] T. Bouwmans, L. Davis, J. Gonzalez, M. Piccardi, and C. Shan, "Special issue on background modeling for foreground detection in real-world dynamic scenes," *Machine Vision and Applications*, vol. 25, pp. 1101–1103, 2014.

[4] A. Vacavant, L. Tougne, and T. Chateau, "Special section on background models comparison," *Computer Vision and Image Understanding*, vol. 122, pp. 1–202, 2014.

[5] C. Stauffer and W. L. Grimson, "Adaptive background mixture models for real-time tracking," *Computer Vision and Pattern Recognition*, pp. 2246–2252, 1999.

[6] R. D. Cajote *et al.*, "Framework of surveillance video analysis and transmission system using background modeling and MIMO-OFDM," in *IEEE International Conference on Digital Signal Processing*, pp. 1071–1075, 2015.

[7] F. Chiabrando, R. Chiabrando, D. Piatti, and F. Rinaudo, "Sensors for 3d imaging: Metric evaluation and calibration of a ccd/cmos time-of-flight camera," *Sensors*, vol. 9, pp. 10080–10096, 2009.

[8] A. Kolb, E. Barth, R. Koch, and R. Larsen, "Time-of-flight sensors on computer graphics," in *Eurographics*, 2009.

[9] S. Foix, G. Alenya, and C. Torras, "Lock-in time-of-flight (ToF) cameras: A survey," *Sensors*, vol. 11, no. 9, pp. 1917–1926, 2011.

[10] T. Oggier *et al.*, "Novel pixel architecture with inherent background suppression for 3d time-of-flight imaging," in *Videometrics VIII*, 2005.

[11] A. Glazer, M. Lindenbaum, and S. Markovitch, "One-class background model," *Lecture Notes in Computer Science*, vol. 7728, pp. 301–307, 2012.

[12] S. H. Cho *et al.*, "Background subtraction based object extraction for time-of-flight sensor," in *IEEE Global Conference on Consumer Electronics*, pp. 48–49, 2013.

[13] J. Giacomantone, L. Violini, L. Lorenti, and A. D. Giusti, "Supresión de segundo plano en imágenes de tiempo de vuelo," in *XXII Congreso Argentino de Ciencias de la Computación*, pp. 1064–1073, 2016.

[14] V. Vapnik, "An overview of statistical learning theory," *IEEE Transactions on Neural Networks*, vol. 10, no. 5, pp. 988–999, 1999.

[15] D. J. Tax and R. P. W. Duin, "Support vector data description," *Machine Learning*, vol. 54, pp. 45–66, 2004.

[16] B. Schlkopf, J. C. Platt, J. Shawe-Taylor, A. J. Samola, and R. C. Williamson, "Estimating the support of a high dimensional distribution," *Neural Computation*, vol. 13, no. 7, pp. 1443–1471, 2001.

[17] F. Chiabrando, D. Piatti, and F. Rinaudo, "Sr-4000 tof camera: Further experimental tests and first applications to metric surveys," in *V Symposium on Remote Sensing and Spatial Information Sciences*, pp. 149–154, 2010.

[18] M. Lindner, I. Schiller, A. Kolb, and R. Koch, "Time-of-flight sensor calibration for accurate range sensing," *Computer Vision and Image Understanding*, vol. 114, no. 12, pp. 1318–1328, 2010.

[19] M. Reynolds *et al.*, "Capturing time-of-flight data with confidence," in *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 945–952, 2011.

[20] F. Lenzen *et al.*, "Denoising strategies for time-of-flight data," *Lecture Notes in Computer Science*, vol. 8200, pp. 25–45, 2013.

[21] A. Sabov and J. Kruguer, "Identification and correction of flying pixels in range camera data," in *Computer Graphics*, pp. 135–145, 2010.

[22] H. Rapp, M. Frank, F. Hamprecht, and B. Jahne, "A theoretical and experimental investigation of the systematic errors and statistical uncertainties of time of flight cameras," *Intelligent Systems Technologies and Applications*, vol. 5, pp. 402–413, 2008.

[23] H. Schöner, "Denoising 3d images from time-of-flight cameras using extended anisotropic diffusion," *SPIE Newsroom*, 2012.