

Supresión de segundo plano en imágenes de tiempo de vuelo

Javier Giacomantone, Lucía Violini, Luciano Lorenti, Armando De Giusti

Instituto de Investigación en Informática (III-LIDI),
Facultad de Informática - Universidad Nacional de La Plata - Argentina.
La Plata, Buenos Aires, Argentina.
{jog, lviolini, llorenti, degiusti}@lidi.info.unlp.edu.ar

Resumen. En este artículo se presenta un método computacional para detectar y extraer el plano de fondo, segundo plano, a partir de datos obtenidos por cámaras de tiempo de vuelo. Se utiliza una variante de un método de clasificación basado en máquinas de soporte vectorial. Considerando las características particulares del tipo de cámaras utilizadas, se incorpora adecuadamente la información de rango e intensidad, y se utiliza la capacidad para obtener secuencias rápidas de datos en una modalidad particular. El artículo revisa las técnicas específicas de reconocimiento de patrones utilizadas, presenta la solución propuesta y muestra resultados experimentales preliminares del método propuesto.

Palabras clave: Visión por computador, cámaras TOF industriales, Reconocimiento de patrones, Máquinas de soporte vectorial.

1. Introducción

Los métodos de sustracción de plano de fondo, *background*, son utilizados en visión por computador normalmente para detectar objetos de interés, *foreground objects*, en una secuencia de imágenes [1]. La representación del plano de fondo se denomina modelo. En los últimos años han sido propuestos distintos métodos [2][3][4]. Estos varían en cuanto a sus beneficios y limitaciones dependiendo del tipo de imágenes, el ruido, los artefactos y las hipótesis básicas sobre las que el correspondiente método computacional se sustenta. Los principales métodos se pueden clasificar en paramétricos o no paramétricos según establezcan o no suposiciones sobre la distribución de probabilidades de la intensidad del plano de fondo [5]. Otro criterio de clasificación posible es entre los métodos basados en analizar cada *pixel* (*picture element*), o los basados en bloques que dividen la imagen en un número de bloques de pixels sobre los cuales operan. Un método normalmente parte del conocimiento previo de imágenes del plano de fondo, pero no de los objetos de interés. Por lo tanto, desde el punto de vista de reconocimiento de patrones, es un problema de clasificación semisupervisado [6]. El plano de fondo puede ser estacionario o no estacionario. Las cámaras de tiempo de vuelo, *time of flight* (*TOF*), utilizan un arreglo de sensores de tecnología

CMOS y su propia fuente de iluminación en el infrarrojo cercano [7][8]. A partir de la diferencia de fase entre la señal emitida y la reflejada por la escena para cada posición obtienen imágenes de rango, $2\frac{1}{2}D$, e intensidad $2D$. A pesar de ser cámaras de baja resolución, permiten obtener secuencias rápidas de imágenes [9]. La sensibilidad a la iluminación de fondo requiere no sólo consideraciones de diseño particulares sino también la adecuada calibración del tiempo de integración [10][11]. Las características generales anteriores requieren consideraciones y métodos específicos para su posprocesamiento en general, y en particular para la sustracción o también denominada supresión del *background*. La figura 1 muestra cuatro modelos distintos de cámara TOF.

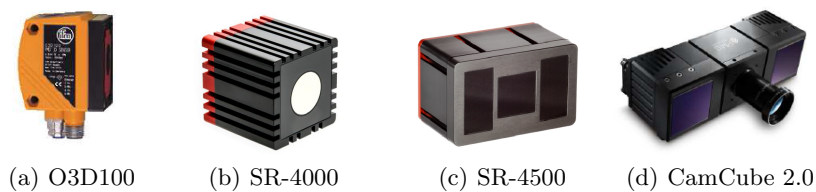


Figura 1. (a) Ifm electronic[©], (b-c) HEPTAGONTM, (d) PMD[Vision][®]

Glazer et. al [12] proponen un método orientado a bloques semisupervisado, utilizando máquinas de soporte vectorial, para sustraer el plano de fondo en imágenes de intensidad pero no en imágenes de rango. Cho et. al [13] proponen, utilizando imágenes de tiempo de vuelo, generar un modelo estadístico del fondo tanto para la imagen de intensidad como para la de rango. Una vez obtenido el modelo, cada pixel de una nueva imagen es comparado con el modelo y se lo clasifica como perteneciente o no al fondo. El método propuesto en este artículo aborda el mismo tipo de problema, es orientado a pixels y asume clases desbalanceadas. Considera los datos obtenidos en una estructura particular y utiliza clasificación basada en máquinas de soporte vectorial de una clase, *one class support vector machines (OCSVM)*. El objetivo de la alternativa propuesta es sustraer el plano de fondo combinando la información de rango e intensidad que permita determinar los objetos de la escena que no pertenecen al *background*.

El artículo ha sido organizado del siguiente modo. En la sección 2 se presenta una breve revisión de los conceptos fundamentales utilizados para construir el modelo. La sección 3 presenta el método propuesto. La sección 4 expone resultados experimentales. Finalmente, en la sección 5 se presentan conclusiones.

2. Máquinas de soporte vectorial

Las máquinas de soporte vectorial tienen como fundamento la teoría de aprendizaje estadístico [14]. En esta sección se presenta en forma breve el método de clasificación y la notación utilizada. Dados l patrones de entrenamiento para dos

clases $\{(x_i, y_i), i = 1, \dots, l\}$, donde $x_i \in \mathbb{R}^N$ es un patrón, un vector de características de dimensión N e $y_i \in \{1, -1\}$ es la etiqueta de la clase a la que pertenece x_i . Decimos que $\mathbf{x}_1, \dots, \mathbf{x}_l \in \mathcal{X}$ donde $l \in \mathbb{N}$ es el número de observaciones y \mathcal{X} es un subconjunto compacto de \mathbb{R}^N . Una SVM determina el hiperplano que maximiza la separación entre ambas clases y es matemáticamente equivalente a maximizar el margen dado por $\frac{2}{\|\mathbf{w}\|}$ sujeto a $y_i(x_i \cdot \mathbf{w} + b) \geq 1$. El hiperplano $(\mathbf{w} \cdot \mathbf{x}) + b = 0$ queda determinado por b y \mathbf{w} . Se puede relajar la condición de muestras de entrenamiento separables incluyendo variables artificiales que permitan mejorar también la generalización. Para obtener el hiperplano minimizamos

$$\begin{aligned} \text{mín} \quad & \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_i \xi_i \\ \text{sujeto a} \quad & y_i[(\mathbf{w} \cdot \mathbf{x}_i) + b] \geq 1 - \xi_i \end{aligned} \tag{1}$$

donde C controla el compromiso entre complejidad del hiperplano y el error sobre el conjunto de entrenamiento. Resolviendo el Lagrangiano en la forma dual obtenemos la función de decisión como:

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

donde α_i son los multiplicadores de Lagrange. Si las clases no son linealmente separables es posible extender el método proyectando los patrones en un espacio de mayor dimensión donde si sea posible clasificarlos linealmente, para luego obtener la función de decisión no lineal en el espacio original. Pueden ser utilizados distintos tipos de *kernels* que corresponden a estimadores no lineales en el espacio de entrada. En particular en este trabajo es utilizado un kernel de base radial $k(x, x_i) = e^{-\gamma \|x - x_i\|^2}$, γ es el parámetro que determina el ancho del kernel.

Han sido propuestos dos algoritmos basados en SVM para operar con una clase, *support vector data description* [15] y *one-class SVM (OCSVM)* [16]. El algoritmo OCSVM estima una función de decisión f que es positiva para un subconjunto del espacio de características y negativa para el complemento. El mapeo de los datos al espacio de características corresponde al kernel y separa los datos del origen con máximo margen. Sea $\Phi : \mathcal{X} \rightarrow \mathcal{F}$ un mapeo de características en \mathcal{F} de tal modo que el producto interno en la imagen de Φ puede ser calculado evaluando un kernel simple.

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

Se puede reescribir como un problema de optimización.

$$\begin{aligned} \text{mín}_{\mathbf{w} \in \mathcal{F}, \xi \in \mathbb{R}^l, \rho \in \mathbb{R}} \quad & \frac{1}{2} \|\mathbf{w}\|^2 + \frac{1}{\nu l} \sum_i \xi_i - \rho \\ \text{sujeto a} \quad & (\mathbf{w} \cdot \Phi(\mathbf{x}_i)) \geq \rho - \xi_i, \xi_i \geq 0, \end{aligned} \tag{2}$$

donde $\nu \in (0, 1]$ es un parámetro que controla el término de penalización y ξ_i son variables de holgura. Resolviendo el problema de optimización (1) obtenemos \mathbf{w}

y ρ y la función de decisión es -1 para valores atípicos en el conjunto de datos y +1 para el resto de los patrones en el conjunto de datos.

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

Introduciendo multiplicadores de Lagrange $\alpha_i, \beta_i \geq 0$, obtenemos

$$L(\mathbf{w}, \xi, \rho, \boldsymbol{\alpha}, \boldsymbol{\beta}) = \frac{1}{2} \|\mathbf{w}\|^2 + \frac{1}{\nu l} \sum_i \xi_i - \sum_i \beta_i \xi_i - \rho - \sum_i \alpha_i (\mathbf{w} \cdot \Phi(\mathbf{x}_i) - \rho + \xi_i)$$

Fijando las derivadas con respecto a las variables primarias \mathbf{w}, ξ, ρ igual a cero obtenemos

$$\begin{aligned} \mathbf{w} &= \sum_i \alpha_i \Phi(\mathbf{x}_i), \\ \alpha_i &= \frac{1}{\nu l} - \beta_i \leq \frac{1}{\nu l}, \\ \sum_i \alpha_i &= 1. \end{aligned}$$

La función de decisión puede ser escrita como

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

Los multiplicadores α_i se pueden resolver en el problema dual:

$$\begin{aligned} \underset{\boldsymbol{\alpha}}{\text{mín}} \quad & \frac{1}{2} \sum_{ij} \alpha_i \alpha_j k(\mathbf{x}_i, \mathbf{x}_j) \\ \text{sujeto a} \quad & 0 \leq \alpha_i \leq \frac{1}{\nu l}, \quad \sum_i \alpha_i = 1. \end{aligned}$$

El parámetro ρ puede ser obtenido notando que para cualquier α_i y el correspondiente patron \mathbf{x}_i se satisface que:

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

3. Supresión de plano de fondo por medio de OCSVM

El método propuesto consta de dos modos de operación posible y dos etapas cada uno. La primera etapa determina el modelo del fondo que en una segunda etapa permite suprimir el plano de fondo e identificar los *pixels* del *foreground*. El modelo de *background* se determina para cada pixel mediante el cálculo de los parámetros necesarios para definir la máquina de soporte vectorial de una clase como en la sección 2 a partir de un arreglo de imágenes de entrenamiento. El segundo modo de operación es una extensión del primer modo que explota la capacidad de las cámaras TOF de adquirir secuencias rápidas de datos. La figura (2) muestra la estructura de datos adoptada para ambas modalidades.

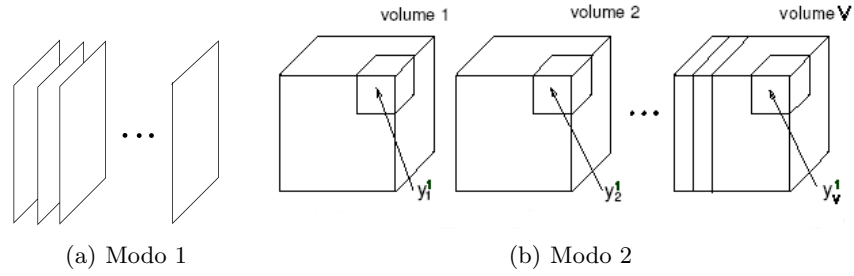


Figura 2. (a) Imágenes de entrenamiento, (b) Voxel correspondiente en V volúmenes

El primer modo de operación propuesto para la etapa de entrenamiento utiliza dos arreglos de imágenes, TOF_r y TOF_g . Cada uno consta de l imágenes $2\frac{1}{2}D$ y $2D$ representando el valor de rango y de intensidad respectivamente. Se calculan los parámetros que definen el modelo aplicando el procedimiento descrito en la sección 2 considerando l vectores de entrenamiento, $k = 1...l$, para cada pixel (i, j) , siendo:

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

el k -ésimo vector de entrenamiento para el pixel (i, j) donde:

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

En la segunda etapa para cada nueva imagen de tiempo de vuelo provista por la cámara TOF se determina para cada pixel su pertenencia al fondo o al objeto de interés mediante la siguiente ecuación:

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

El segundo modo de operación propuesto para la etapa de entrenamiento utiliza dos arreglos de l volúmenes ($v = 1...l$) conformados por arreglos de imágenes, TOF_r y TOF_g . Cada uno consta de M imágenes $2\frac{1}{2}D$ y $2D$ representando el valor de rango y de intensidad respectivamente. Se calculan los parámetros correspondiente para definir un clasificador OCSVM a partir de V vectores $v = 1...l$ para cada *voxel (volumen element)* (i, j, k) , siendo:

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

el v -ésimo vector de entrenamiento para el voxel (i, j, k) donde:

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

En la segunda etapa para cada nueva imagen de tiempo de vuelo provista por la cámara TOF se determina para cada voxel su pertenencia al fondo o al objeto de interés mediante la siguiente ecuación:

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

La resolución temporal tiene una cota máxima determinada por la cámara TOF utilizada, pero las características del problema particular a resolver, son las que determinarán el diseño adecuado y el compromiso entre velocidad, memoria disponible y complejidad.

4. Resultados Experimentales

La evaluación del método propuesto fue realizada utilizando secuencias de datos de tiempo de vuelo de rango e intensidad. Se generaron imágenes TOF artificiales simples con diferentes niveles y tipo de ruido. Se utilizaron también imágenes con un objeto sintetizado y con un objeto de *foreground* real. La figura 3 muestra una imagen sintética y una imagen real de rango e intensidad. Las imágenes sintéticas utilizadas fueron de 10x10x3 voxels por volumen y 20 volúmenes. Todos los voxels tienen sumado ruido gaussiano de media 16000 y desvío estándar 4000, figura 3(a). El valor de los voxels simulando *foreground* fue modificado entre 600 y 1500 con incrementos de 100 generando así 10 imágenes sintéticas denominadas TOF1 a TOF10 con ruido gaussiano. Otras 10 imágenes artificiales fueron generadas con ruido Rayleigh denominadas TOF11 a TOF20.

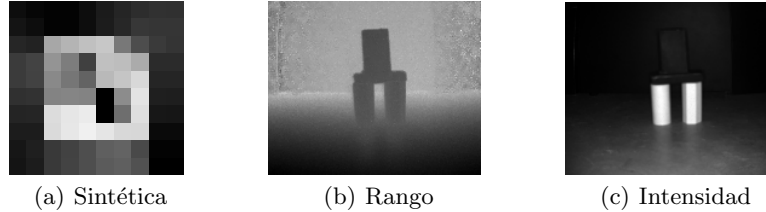


Figura 3. Imágenes de tiempo de vuelo

Las imágenes reales fueron adquiridas con una cámara TOF MESA SwissRanger SR4000 [17][18] que permite obtener imágenes de rango e intensidad de 144x176 pixels, hasta 50fps y rango de trabajo de 0,8 a 8m. Se construyó una escena simple y se adquirieron secuencias de imágenes, conformando un arreglo de imágenes TOF para construir el modelo de *background*, como muestran las figuras 3(b) y 3(c). Sobre una de las imágenes de fondo se sintetizó un objeto y la denominamos imagen de evaluación TOFS. También se adquirieron imágenes para la etapa de evaluación de un objeto real y denominamos al conjunto o a la imagen TOFR, utilizándose la notación de la sección 3, para diferenciar cada imagen o pixel en el arreglo. En la figura 4 podemos observar en color, simplemente para facilitar su identificación en la imagen, los objetos pertenecientes al *foreground* en las imágenes de evaluación. En las figuras 4(a) y 4(b) se muestran las imágenes de

los objetos sintetizados sobre la imagen artificial y sobre la real respectivamente. La imagen de la figura 4(c) se obtiene mediante el desplazamiento de un objeto del *background* y la captura de una nueva imagen TOF. En imágenes TOF reales no sólo existe ruido en la adquisición de las series temporales de cada pixel [19][20] sino también errores en la frontera entre objetos, efecto conocido como *flying pixels* [21][22].

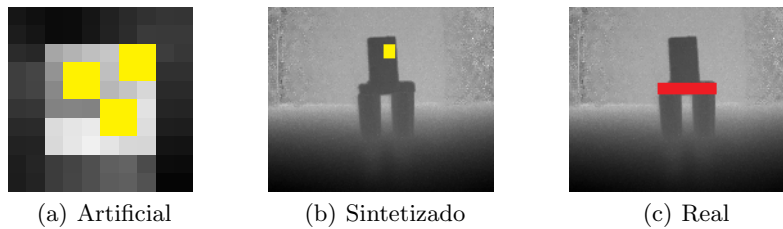


Figura 4. *Foreground* superpuesto en imágenes de rango

La tabla 1 presenta resultados comparativos expresados como la fracción de verdaderos positivos (TPF) y la fracción de falsos negativos (FPF). Estos valores corresponden al punto óptimo de operación en una curva ROC (*Receiving Operating Characteristics*). Se indican también otros datos de interés como la distancia al punto óptimo de operación y el área bajo la curva. Denominamos métodos C1 y C2 al método propuesto en [13] sin filtros morfológicos y utilizando procesamiento morfológico respectivamente. La figura 5 presenta las curvas ROC correspondientes a la evaluación con TOF9, TOF12 y TOFS.

Imagen TOF	Método	Area	d_{po}	TPF_{po}	FPF_{po}
TOF9	$Modo1_{\nu=0,37}$	0.989	0.661	0.979	0.046
TOF12	$Modo2_{\nu=0,4}$	0.977	0.693	0.892	0.049
TOFS	$Modo1_{\nu=0,38}$	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

Tabla 1. Métricas de rendimiento

La figura 6 muestra una imagen de rango con *foreground* real y el resultado de aplicar el método propuesto y suprimir el *background* en una imagen real.

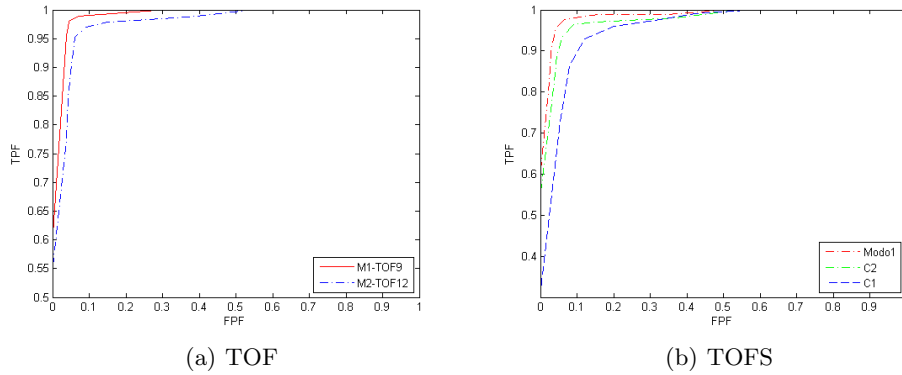


Figura 5. Curvas ROC

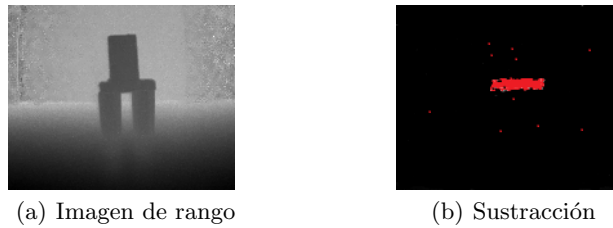


Figura 6. TOFR

5. Conclusiones

En este artículo se ha presentado un método que permite sustraer el plano de fondo en secuencias de imágenes de cámaras de tiempo de vuelo. Los resultados experimentales preliminares muestran que es viable, como resultado de combinar la información de rango e intensidad en un clasificador para clases desbalanceadas y basado en secuencias de pixels correspondientes. El método prevé una tercera etapa, que utilice la propiedad de correlación espacial de pixels o voxels y combine la información de rango e intensidad para filtrar anisotrópicamente o morfológicamente las imágenes y así permitir reducir ruido y *flying objects*. Una extensión futura de la evaluación experimental tiene previsto comparar los resultados obtenidos con métodos basados en bloque y considerar la significancia estadística de los resultados comparativos.

Referencias

1. P. Rajan, 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, (2):300-308, 2014.

2. Maasimo Piccardi. Background subtraction techniques: a review. *IEEE International Conference on System. Man and Cybernetics*, 3099-3104, 2004.
3. T. Bouwmans, L. Davis, J. Gonzalez, M. Piccardi, C. Shan. Special Issue on Background Modeling for Foreground Detection in Real-World Dynamic Scenes. *Machine Vision and Applications*, (25):1101-1103, 2014.
4. A. Vacavant, L. Tougne, T. Chateau. Special section on background models comparison. *Computer Vision and Image Understanding*,(122), 2014.
5. Chris Stauffer. W.E.L Grimson. Adaptive background mixture models for real-time tracking. *Computer Vision and Pattern Recognition*, (2), 1999.
6. Rhandley D. Cajote et. al. Framework of Surveillance Video Analysis and Transmission System using Background Modeling and MIMO-OFDM. *IEEE International Conference on Digital Signal Processing*, 1071-1075, 2015.
7. F. Chiabrando, R. Chiabrando, D. Piatti, F. Rinaudo. Sensors for 3D imaging: Metric evaluation and calibration of a CCD/CMOS time-of-flight camera. *Sensors*, (9):10080–10096, 2009.
8. A. Kolb, E. Barth, R. Koch, R. Larsen. Time-of-Flight Sensors on Computer Graphics. *Proceedings of the Eurographics, Munich, Germany*, 2009.
9. S. Foix, G. Alenya, C. Torras. Lock-in Time-of-Flight (ToF) Cameras: A Survey. *IEEE Sensors Journal*, 11(9):1917-1926, 2011.
10. Thierry Oggier et. al. Novel Pixel Architecture with Inherent Background Suppression for 3D Time-of-Flight Imaging. *Proc. SPIE 5665, Videometrics VIII*, 2005.
11. P. Gil et. al . Calibración de cámaras de tiempo de vuelo: Ajuste adaptativo del tiempo de integración y análisis de la frecuencia de modulación. *Revista Iberoamericana de Automática e Informática industrial*, (10):453–464, 2013.
12. Assaf Glazer, Michael Lindenbaum, Shaul Markovitch. One-Class Background Model. *Computer Vision Workshop. Lecture Notes in Computer Science*, (7728):301-307, 2012.
13. Shung Han Cho, Kwanghyuk Bae, Kyu-Min Kyung, Seongyeong Jeong, Tae-Chan Kim. Background Subtraction Based Object Extraction for Time-of-Flight Sensor. *IEEE Global Conference on Consumer Electronics*, 48-49, 2013.
14. V. Vapnik. An overview of statistical learning theory. *IEEE Transactions on Neural Networks*, 10(5):988-999, 1999.
15. D. M. J. Tax, R. P. W. Duin. Support vector data description. *Machine Learning*, 54:45-66, 2004.
16. B. Schölkopf, J. C. Platt, J. Shawe-Taylor, A. J. Samola, R. C. Williamson. Estimating the support of a high dimensional distribution. *Neural Computation*, 13(7):1443-1471, 2001.
17. F. Chiabrando, D. Piatti, F. Rinaudo. SR-4000 TOF Camera: Further Experimental Tests and First Applications to Metric Surveys. *Proceedings of V Symposium on Remote Sensing and Spatial Information Sciences*, 149-154, 2010.
18. M. Lindner, I. Schiller, A. Kolb, R. Koch. Time-of-flight sensor calibration for accurate range sensing. *Comput. Vis. Image Underst.*, (114):1318–18, 2010.
19. M. Reynolds et. al. Capturing Time-of-Flight data with confidence. *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 945-952, 2011.
20. Frank Lenzen et. al. Denoising Strategies for Time-of-Flight Data, 2013. *Lecture Notes in Computer Science (8200):25-45, Springer*, 2013.
21. A. Sabov, J. Kruguer. Identification and Correction of Flying Pixels in Range Camera Data. *Proc. Computer Graphics. Slovakia*, 135-142, 2010.
22. H. Rapp, M. Frank, F.A. Hamprecht, B. Jahne. A theoretical and experimental investigation of the systematic errors and statistical uncertainties of time of flight cameras. *Int. J. Intell. Syst. Technol. Appl.*, (5):402–413, 2008.