

LiDAR de estado sólido multimodal para aplicaciones de percepción avanzada

Multimodal solid-state LiDAR for advanced perception applications

Pablo GARCÍA-GÓMEZ ⁽¹⁾, Noel RODRIGO⁽¹⁾, Jordi RIU⁽²⁾, Josep R.CASAS⁽³⁾,
Santiago ROYO ⁽¹⁾

1. Centre de Desenvolupament de Sensors, Instrumentació i Sistemes, Universitat Politècnica de Catalunya (UPC-CD6), Rambla Sant Nebridi 10, E08222, Terrassa, Spain.
2. Beamagine S.L., Carrer de Bellesguard 16, 08755 Castellbisbal, Spain.
3. Grup de Processat d'Imatge, Departament TSC, Universitat Politècnica de Catalunya (UPC), Carrer de Jordi Girona 1-3, 08034 Barcelona, Spain.

Persona de contacto: Pablo García-Gómez (pablo.garcia.gomez@upc.edu).

RESUMEN:

La percepción exacta del entorno es un requisito indispensable en aplicaciones de percepción avanzada como sistemas de navegación de vehículos autónomos, visión en seguridad y la robótica. Por ello, los sistemas de imagen LiDAR se han convertido en sensores clave debido a su capacidad de medida de la geometría de la escena en 3D. Sin embargo, se requiere mucha información para tomar decisiones fiables y por ello se suelen combinar diferentes sensores para obtener la mayor cantidad posible. Dentro de este marco, se presenta un sistema multimodal basado en un LiDAR de estado sólido combinado con otros tres sensores de imagen capaz de dar información multimodal con un reducido error de paralaje en la fusión.

Palabras clave: fusión de sensores, LiDAR de estado sólido, visión por computador, inteligencia artificial, percepción mejorada, navegación autónoma, robótica

ABSTRACT:

Perception of the environment is an essential requirement for the fields of autonomous vehicles and robotics. Consequently, LiDAR imaging sensors have become crucial sensors for such applications due to their 3D geometry sensing capability. However, autonomous systems claim for high amounts of data to make reliable decisions so many different sensors are often combined. In this context, we present a multimodal imaging system based on a solid-state LiDAR combined with three other imaging sensors that provides multimodal information with low parallax fusion error.

Key words: sensor fusion, solid-state LiDAR, computer vision, artificial intelligence, enhanced perception, autonomous navigation, robotics

1.- Introducción

Actualmente se dispone de una gran variedad de sensores capaces de aportar información de nuestro entorno, basados en diferentes principios de medida. Gracias a ellos y a través de la actividad de investigación en Inteligencia Artificial (AI) y Visión por Computador

(CV), podemos dotar de percepción avanzada a sistemas artificiales, imitando la percepción humana.

Los algoritmos de percepción [1] requieren la mayor cantidad posible de información sobre el entorno, incluso redundante, para mejorar la toma de decisiones de los sistemas autóno-

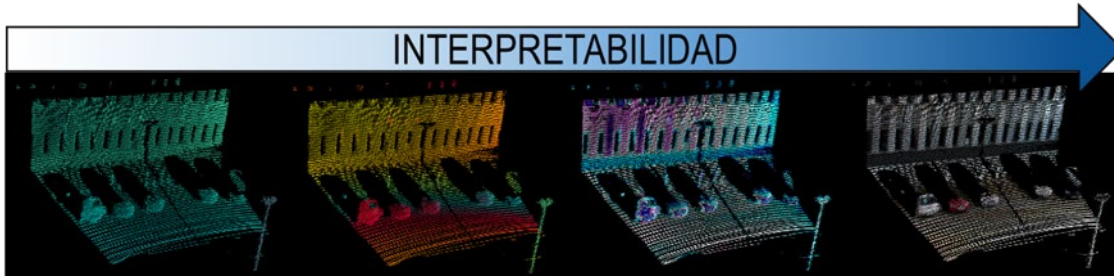


Fig. 1: Mejora de la percepción de una escena gracias a la fusión de sensores. En este caso, se muestra una nube de puntos de un LiDAR fusionado con el color de una cámara convencional.

mos en aplicaciones críticas, esencialmente porque cada sensor tiene modos de fallo y prestaciones diferentes. En este marco, la combinación de sensores, o más bien la fusión de sus datos, da respuesta a esta necesidad para aplicaciones ambiciosas como la navegación autónoma y la robótica [2] [3].

Para dichas aplicaciones, se requiere información de profundidad (con información detallada de la geometría del entorno) combinada con distintos modos de imagen (proyecciones de la geometría en un plano con información no disponible en la imagen LiDAR) para ofrecer mayor fiabilidad y robustez [4].

Los sensores LiDAR (Light Detection And Ranging) miden la distancia a un objetivo a través de la medida física del tiempo entre eventos sobre una magnitud medible de la luz [5]. Esta medida temporal se conoce como el tiempo de vuelo (TOF). Generalmente, en aplicaciones de exterior se utiliza un haz láser pulsado para emitir pulsos de luz y se cuenta el tiempo que tarda en ir al objeto y volver. A través de un escaneo del entorno, se generan imágenes en 3D llamadas nubes de puntos.

En la Fig. 1 se puede observar cómo se mejora la percepción de una nube de puntos gracias a la fusión de información en color adicional. De izquierda a derecha, primero se muestra la nube de puntos con un color uniforme de visualización. Luego, se le da un falso color a cada punto en función de su distancia al dispositivo. En la tercera escena, se muestra la información de la intensidad de los puntos a la longitud de onda del láser, relacionada con la reflectancia de los objetos. Finalmente, se fusiona con una cámara de color convencional. Con esta figura se muestra la mejora en la interpretación del entorno, objetivo de la fusión.

En la Sección 2 introduciremos la base matemática de la fusión y qué sensores se han utilizado en el caso que se presenta, así como sus modelos de funcionamiento. Luego, en la Sección 3 se mostrarán los resultados del sistema de fusión y en la Sección 4 se comentarán brevemente sus posibles aplicaciones. Para acabar, la Sección 5 presentará las conclusiones.

2.- Método

2.1.- Base matemática

Para combinar la información de cada sensor de un conjunto, se debe entender la percepción individual de cada uno de ellos y conocer cuál es su posición y la orientación relativa entre ellos. De esta forma, hablamos de intrínsecos cuando nos referimos a la relación entre su medida y el mundo, y de extrínsecos cuando lo hacemos sobre su posición relativa.

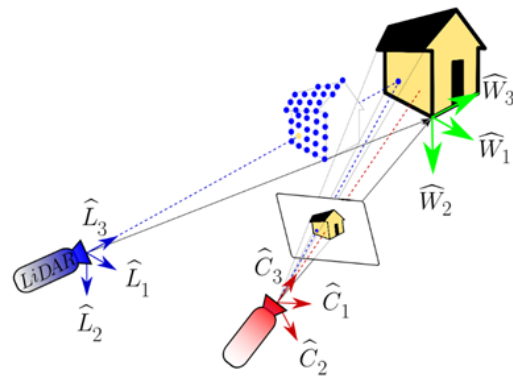


Fig. 2: Esquema de fusión entre un LiDAR (azul) y una cámara convencional (rojo).

Tal como muestra la Fig. 2, cada sensor consta de un sistema de referencia propio definido por una posición y tres ejes de coordenadas. Como consecuencia, los extrínsecos se definen matemáticamente como un vector de traslación y una matriz de rotación [6]. Esto se conoce en álgebra como una transformación

rígida. Cada sensor realiza las medidas del entorno desde su punto de vista, ergo su propio sistema de referencia, a través de su principio de funcionamiento particular.

La siguiente ecuación muestra la transformación entre el sistema de referencia del entorno W y el del LiDAR, L , para vectores fila:

$${}^L\mathbf{p} = {}^W\mathbf{p} \cdot \mathbf{R}_{WL} + \mathbf{t}_{WL} \quad (1)$$

Donde ${}^W\mathbf{p}$ y ${}^L\mathbf{p}$ son el mismo punto 3D expresado en el sistema de referencia del mundo y del LiDAR, respectivamente; $\mathbf{R}_{WL}$ es la matriz de rotación del mundo al LiDAR y $\mathbf{t}_{WL}$ su vector de traslación.

Por un lado, la matriz de rotación define la orientación de un sistema respecto al otro ya que expresa sus vectores de coordenadas en función de los del otro. En el caso anterior, $\mathbf{R}_{WL}$ expresa la orientación del mundo W en L . En concreto, cada fila de la matriz es uno de los vectores de coordenadas expresados en el otro sistema. Gracias a que las matrices de rotación son ortonormales, la inversa es directamente su traspuesta, por lo que se puede definir la orientación del sensor LiDAR en el mundo de la siguiente forma:

$$\mathbf{R}_{LW} = (\mathbf{R}_{WL})^T = \begin{bmatrix} {}^W\hat{L}_1 \\ {}^W\hat{L}_2 \\ {}^W\hat{L}_3 \end{bmatrix} \stackrel{\text{def}}{=} \mathbf{ORI}_{L|W} \quad (2)$$

Por otro lado, el vector de traslación define la posición de los sistemas relacionando sus orígenes. Entonces, se puede hallar el origen del LiDAR en el mundo ${}^W\mathbf{o}_L$ utilizando la inversa de (1) y (2), y aplicando que ${}^L\mathbf{o}_L = [0,0,0]$:

$$\begin{aligned} {}^W\mathbf{o}_L &= {}^L\mathbf{o}_L \cdot \mathbf{ORI}_{L|W} + \mathbf{t}_{LW} = \mathbf{t}_{LW} \\ {}^W\mathbf{o}_L &= \mathbf{t}_{LW} \stackrel{\text{def}}{=} \mathbf{loc}_{L|W} \end{aligned} \quad (3)$$

La relación inversa se obtiene fácilmente.

$$\mathbf{loc}_{L|W} = -\mathbf{loc}_{W|L} \cdot (\mathbf{ORI}_{W|L})^T \quad (4)$$

Estas expresiones son válidas para cada sensor, permitiendo el traspaso de información entre sensores y la fusión de sus datos.

2.2.- Sensores del sistema

Una vez hemos visto como se relacionan los sensores con el entorno y entre sí, vamos a presentar cuáles son los sensores utilizados

para este trabajo. Típicamente, se fusionan tanto sensores capaces de medir características geométricas como el LiDAR, el radar y los ultrasonidos, con sensores de imagen que proporcionan información 2D con gran resolución [7]–[9] en distintos dominios del espectro electromagnético. De esta forma, se obtiene información 3D multimodal

2.2.1.- LiDAR

En nuestro sistema, consideramos el LiDAR como el sensor principal del sistema, aportando la información 3D. El LiDAR utilizado es un sistema pre-comercial propiedad de Beamagine: un sistema pulsado de estado sólido basado en la tecnología MEMS [10] que proporciona una gran densidad de puntos (300x150 puntos en la escena) y un campo de visión (FOV) similar al de una cámara. Su rango supera los 100 metros gracias a su resistencia a la contaminación solar de fondo y su haz láser tiene una divergencia menor a 0.1°. Ofrece datos a 10 FPS con la información de intensidad a la longitud de onda de 1064 nm.

El modelo de intrínsecos utilizado relaciona cada píxel (i, j) de adquisición con una dirección de escaneo definida por dos ángulos (θ_H, θ_V) con una función parametrizable [11]:

$$\mathbf{f}: (i, j) \xleftrightarrow{x} (\theta_H, \theta_V) \quad (5)$$

Tal que una medida del TOF t_{TOF} se convierte en un punto 3D en el sistema del LiDAR así:

$$\begin{aligned} {}^L\mathbf{p} &= \left(\frac{c}{2} t_{TOF}\right) \begin{bmatrix} \sin(\theta) \cos(\phi) \\ \sin(\theta) \sin(\phi) \\ \cos(\theta) \end{bmatrix}^T \\ \theta &= \arctan\left(\sqrt{\tan(\theta_H)^2 + \tan(\theta_V)^2}\right) \end{aligned} \quad (6)$$

$$\phi = \arctan\left(\frac{\tan(\theta_V)}{\tan(\theta_H)}\right)$$

Donde c es la velocidad de la luz en el vacío ya que el pulso viaja a través del aire.

Gracias a (5) y al método de calibración propuesto en [11], las variaciones en la resolución angular a lo largo del FOV se ven corregidas, proporcionando nubes de puntos de gran exactitud y precisión – a 100 m, el error lateral es de menos de 5 cm.

2.2.2.- Sistemas de imagen

Los sensores de imagen complementarios al LiDAR presentado tienen como objetivo abarcar distintos rangos del espectro con tal de aportar diferentes características a los algoritmos de percepción. Por ello, hemos escogido una cámara en color convencional (RGB: 400 – 700 nm), una polarimétrica monocromática (POL: polarización lineal entre 320 – 1000 nm) y otra térmica (LWIR: 8 – 14 μ m).

Las cámaras han sido objeto de estudio desde los inicios de la CV [12]. El modelo de funcionamiento comúnmente utilizado es el de la cámara estenopeica. Es un modelo de proyección basado en la aproximación óptica de la lente delgada y se combina con modelos de distorsión [13]. Con este modelo, se relaciona un punto 3D de la cámara ${}^c\mathbf{p}$ con su proyección en píxeles $[u, v]$ en la imagen a través de su matriz de intrínsecos $\mathbf{K}_c$.

$$\lambda \begin{bmatrix} u \\ v \\ 1 \end{bmatrix}^T = {}^c\mathbf{p} \cdot \begin{bmatrix} f_x & 0 & 0 \\ s & f_y & 0 \\ c_x & c_y & 1 \end{bmatrix} = {}^c\mathbf{p} \cdot \mathbf{K}_c \quad (7)$$

En esta expresión, λ es una constante arbitraria, f_x y f_y son las distancias focales en píxeles horizontal y vertical, s es un parámetro relacionado con la perpendicularidad de ambos ejes y c_x y c_y son los píxeles del centro óptico. Estos parámetros se determinan utilizando el método de Bouguet [14], igual que la distorsión óptica en la imagen.

2.2.3.- Ensamblaje final

A continuación, se detalla el sistema de fusión LiDAR multimodal (Fig.3), con la información y tamaño otorgados por cada sensor:

Sensor		Medida y Tamaño
LiDAR (L)	L3CAM	3D, 1064 nm 300 x 150
RGB (R)	PHX050S-CC	Color 1224 x 1024
POL (P)	PHX050S-PC	Pol. Lineal 1224 x 1024
LWIR (T)	OPTRIS Xi 400	Temperatura 382 x 288

Tabla^o1: Listado de los sensores fusionados.

Para determinar la relación de extrínsecos, se ha seguido el método propuesto en [6] utilizando 14 capturas.



Fig.º3: Sistema de fusión LiDAR multimodal propiedad de Beamagine.

3.- Resultados

En esta sección, presentaremos algunos ejemplos de fusión obtenidos con nuestro sistema multimodal. Gracias a las reducidas dimensiones del sistema de fusión, el error de paralaje obtenido promediado para todas las combinaciones de sensores es de 3.6 cm (Tabla 2), dentro del rango expuesto en [6], y obtenido con la mitad de capturas propuestas en la referencia. La fusión entre el LiDAR y la cámara térmica es la que mayor error presenta debido a sus bajas resoluciones espaciales.

L-R	L-P	L-T	R-P	R-T	P-T
3.0	3.6	6.4	0.9	3.9	3.7

Tabla^o2: Error de paralaje en la fusión en cm para cada pareja de sensores.

A continuación, se muestran nubes de puntos fusionadas con las tres cámaras.

3.1.- Fusión en color

Las siguientes figuras muestran escenas exteriores interurbanas fusionadas en color.

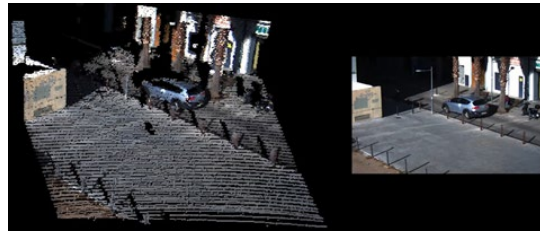


Fig.º4: Fusión en color a 40 m.

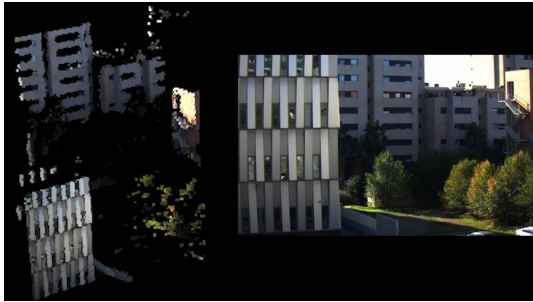


Fig. 5: Fusión en color a 120 metros.

Se puede apreciar que el color aporta información relevante para la interpretación de la escena, tal como se puede determinar en la siguiente Fig.6 en la que el FOV se ha dividido. En la parte izquierda se muestra el color mientras que, en la derecha, la intensidad a la longitud de onda del láser del LiDAR, siendo más difícil percibir que es un parque con coches de distintos colores.

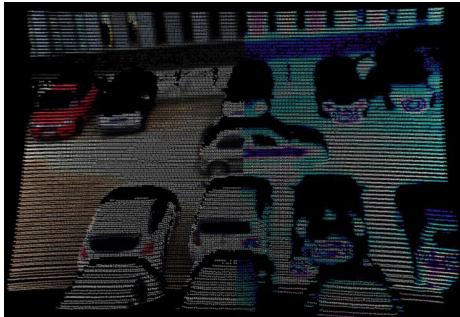


Fig. 6: Fusión en color a 30 metros junto con la información de intensidad a 1064 nm.

3.2.- Fusión polarimétrica

En este apartado, se muestran ejemplos de nubes de puntos fusionadas con la cámara polarimétrica junto con su correspondiente imagen en color. La imagen polarimétrica muestra el grado de polarización lineal, indicando que materiales polarizan la luz en estados lineales y cuáles la despolarizan.

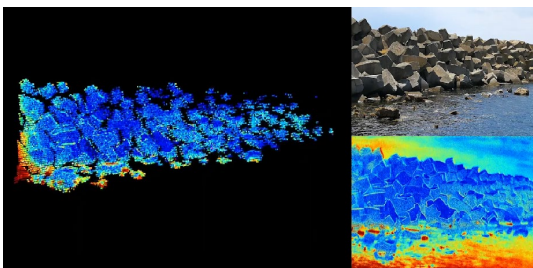


Fig. 7: Fusión polarimétrica de un espigón portuario para detección sobre agua.

En la Fig.7, se puede observar como el LiDAR sólo presenta puntos por encima del agua debido a las propiedades ópticas de ésta. De igual forma, la imagen polarimétrica revela que el agua polariza la luz mientras que las rocas que sobresalen no. Por otro lado, la Fig.8 a continuación muestra una escena de conducción urbana en un día lluvioso y se puede apreciar como la pintura y el asfalto ofrecen un gran contraste en la imagen polarimétrica cuando están mojados.

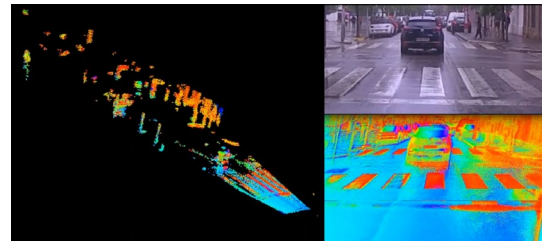


Fig. 8: Fusión polarimétrica de una escena de conducción urbana sobre pavimento mojado.

3.3.- Fusión térmica

Finalmente, se presentan los resultados de fusionar con la cámara térmica (Fig.9) y también la fusión conjunta de las cámaras térmica y de color (Fig.10).

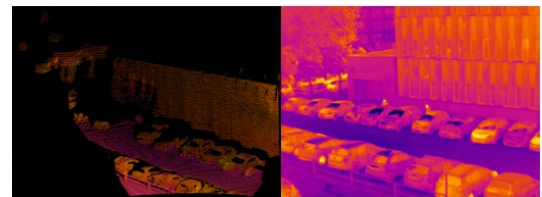


Fig. 9: Fusión térmica a 40 metros.

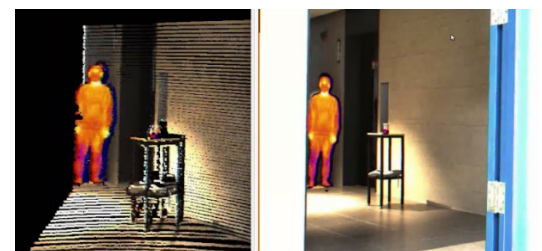


Fig. 10: Fusión en 3D, color y térmica para detección de personas y objetos calientes.

En el segundo caso, sólo se muestra la información térmica de aquellos objetos con temperatura superior a un valor de corte. De lo contrario, se muestra su color. Con esto, se demuestra que la información es complementaria y ayuda a discernir diferentes propiedades del entorno; además de que gracias al LiDAR, se dispone de información geométrica detallada del entorno.

4.- Conclusiones

Se ha presentado un sistema de fusión con información fusionada 3D, visible, polarimétrica, NIR (1064 nm) y LWIR (térmica) con una fiabilidad elevada por su bajo error de paralaje, en parte debido a su tamaño reducido. Se han mostrado diferentes ejemplos de superposición de información 3D con los demás tipos de imágenes, e incluso entre imágenes, demostrando su complementariedad.

Las nubes de puntos fusionadas con distintos modos de imagen son la base de los algoritmos de percepción que se están desarrollando hoy en día para numerosas aplicaciones, en particular en campos de gran demanda como la seguridad y el vehículo autónomo. La existencia de múltiples modos de imagen proporciona robustez y redundancia al procesamiento de los datos, permitiendo un mayor nivel de automatización. Este sistema, además, es capaz de ofrecer tanto los datos por separado para procesarlos primero y fusionar *a posteriori* de la percepción (*early data fusion*), como fusionarlos *a priori* para procesarlos conjuntamente (*late data fusion*). No hay, ahora mismo, un acuerdo unánime en cuanto a qué sistema será preferible finalmente [15].

Agradecimientos: Este trabajo ha sido realizado con el soporte del Ministerio de Ciencia, Innovación y Universidades (MICINN) del Gobierno de España con la beca con núm. DI-17-09181, y el proyecto FIS2017-89850-R.

Referencias

- [1] Richard SZELISKI, “*Computer Vision: Algorithms and Applications*” (Springer Science & Business Media, 2010).
- [2] Babak SHAHIAN JAHROMI, Theja TULABANDHULA, Sabri CETIN, “*Real-Time Hybrid Multi-Sensor Fusion Framework for Perception in Autonomous Vehicles*”, *Sensors*, 19, 20, 4357, 2019.
- [3] Zhangjing WANG, Yu WU, Qingqing UN, “*Multi-Sensor Fusion in Automated Driving: A Survey*”, *IEEE Access*, 8, 2847-2868, 2019.
- [4] Pablo GARCÍA-GÓMEZ, Jordi RIU, Santiago ROYO, “*Lidar and Data Fusion Increase AI Performance*”, *Photonics.com*, https://www.photonics.com/Articles/Lidar_and_Data_Fusion_Increase_AI_Performance/a66490. (2021)
- [5] Santiago ROYO, Maria BALLESTA-GARCIA, “*An Overview of Lidar Imaging Systems for Autonomous Vehicles*”, *Applied Sciences*, 9, 19, 4093, 2019.
- [6] Joris DOMHOF, Julian F. P. KOIJ, Dariu M. GAVRILA, “*An Extrinsic Calibration Tool for Radar, Camera and Lidar*”, en *Proceedings de 2019 International Conference on Robotics and Automation (ICRA)*, 8107-8113, 2019.
- [7] SAE J3016 automated-driving graphic, <https://www.sae.org/site/news/2019/01/sae-updates-j3016-automated-driving-graphic>.
- [8] Gustavo VELASCO-HERNANDEZ, De Jong YEONG, John BARRY, Joseph WALSH, “*Autonomous Driving Architectures, Perception and Data Fusion: A Review*”, en *Proceedings de 2020 IEEE 16th International Conference on Intelligent Computer Communication and Processing (ICCP)*, 315-321, 2020.
- [9] De Jong YEONG, Gustavo VELASCO-HERNANDEZ, John BARRY, Joseph WALSH, “*Sensor and Sensor Fusion Technology in Autonomous Vehicles: A Review*”, *Sensors*, 21, 6, 2140, 2021.
- [10] Sven T. S. HOLMSTRÖM, Utku BARAN, John BARRY, Hakan UREY, “*MEMS Laser Scanners: A Review*”, *Journal of Microelectromechanical Systems*, 23, 2, 17, 2014.
- [11] Pablo GARCÍA-GÓMEZ, Santiago ROYO, Noel RODRIGO, Josep R. CASAS, “*Geometric Model and Calibration Method for a Solid-State LiDAR*”, *Sensors*, 20, 10, 2898, 2020.
- [12] Thomas B. MOESLUND, Erik GRANUM, “*A Survey of Computer Vision-Based Human Motion Capture*”, *Computer Vision and Image Understanding*, 81, 3, 231-268, 2001.
- [13] Duane C. BROWN, “*Decentering Distortion of Lenses*”, *Photogrammetric Engineering*, 32, 444 - 462, 1966.
- [14] Jean-Yves BOUGUET, Pietro PERONA, “*Camera Calibration from Points and Lines in Dual-Space Geometry*”, p.18, 1998.
- [15] Konrad GADZICKI, Razieh KHAMSEHASHARI, Christoph ZETZSCHE, “*Early vs Late Fusion in Multimodal Convolutional Neural Networks*”, en *Proceedings de 2020 IEEE 23rd International Conference on Information Fusion (FUSION)*, 1-6, 2020.