

**CARACTERIZACIÓN PARA LA UBICACIÓN EN LA CAPTURA DE VIDEO
APLICADO A TÉCNICAS DE VISIÓN ARTIFICIAL EN LA DETECCIÓN DE
PERSONAS**

**CHARACTERIZATION FOR LOCATION IN THE CAPTURE OF VIDEO
APPLIED TO ARTIFICIAL VISION TECHNIQUES IN THE PEOPLE
DETECTION**

**Ing. Carlos V. Niño Rondón, MSc. Sergio A. Castro Casadiego,
PhD. Byron Medina Delgado**

Universidad Francisco de Paula Santander, Facultad de Ingeniería.
Grupo de investigación y desarrollo en Electrónica y Telecomunicaciones (GIDET).
Av. Gran Colombia N° 12E-96, San José de Cúcuta, Norte de Santander, Colombia.
Tel.: 57-7-5776655, Ext. 202
E-mail: {carlosvicentenr, sergio.castroc, byronmedina}@ufps.edu.co

Resumen: Este documento muestra los resultados de la caracterización para la ubicación del dispositivo de captura de video para implementar en Python las técnicas de visión artificial de sustracción de fondo, detectores en cascada y flujo óptico, utilizadas en la detección de personas. Se confrontaron aspectos como altura y ángulo de inclinación del dispositivo, obteniendo una exactitud de 98.34% para la técnica de sustracción de fondo, con la cámara ubicada a 4 metros de altura con inclinación de 50°, siendo la de mejor rendimiento en el número de detecciones. En promedio, las pruebas realizadas a 4 metros presentan rendimiento de 87.96%, a 8 metros, 88.19%, a 50°, 87.94% y a 90° el rendimiento es de 88.19%.

Palabras clave: Caracterización, dispositivo de captura de video, sustracción de fondo, detectores en cascada, flujo óptico.

Abstract: This document shows the results obtained during the characterization for the location of video capture device to implement in Python the techniques of artificial vision of background subtraction, cascade detectors and optical flow, used in the people detection. Aspects such as height and angle of inclination of the device were compared, obtaining an accuracy of 98.34% for the technique of subtraction of bottom, with the camera located at 4 meters high with an inclination of 50°, being the best performance in the number of detections. On average, tests carried out at 4 meters show a yield of 87.96%, at 8 meters 88.19%, at 50°, 87.94% and at 90° the yield is 88.19%.

Keywords: Characterization, video capture device, background subtraction, cascade detectors, optical flow.

1. INTRODUCCIÓN

La visión artificial hace referencia al conjunto de técnicas que permiten obtener cierta información y simular las capacidades de las personas para capturar una imagen o un vídeo, analizarlos y tomar decisiones en consecuencia (García S & Caranqui S, 2015).

Igualmente, la visión artificial se presenta como una alternativa en los procesos de localización de objetos y personas, así como en la extracción de características correspondientes a las actividades realizadas por los humanos. (Leo, Medioni, Trivedi, Kanade, & Farinella, 2017).

Actualmente, la detección de personas es uno de los temas de gran relevancia en los procesos de investigación (Franco, Ospina, Cuevas, & Capacho, 2017). El propósito de estos procesos es identificar y monitorear las personas presentes en el cuadro de video analizado con aplicabilidad en video vigilancia y conteo automático de personas (Kajabad & Ivanov, 2019).

El rendimiento de las técnicas de visión por computador depende de la calidad de la imagen de video suministrada al sistema de procesamiento, por lo que es común que aparezcan fallas al implementar este tipo de tecnología en espacios abiertos, ya que se presentan factores como cambios en la luminosidad, sombras y en algunos casos ausencia de luz. (Jeon et al., 2015).

En este artículo se propone determinar las consideraciones de ubicación óptimas para mejorar el resultado en el número de aciertos en la detección de personas para las técnicas de sustracción de fondo, detectores de cascada y flujo óptico. (Pérez, J, Castro, J 2018).

2. METODOLOGÍA

La metodología propuesta se basa en tres etapas: la toma de muestras de imágenes en diferentes condiciones de ubicación, la aplicación de las técnicas de detección de personas y la comparación con respecto al número de aciertos en las detecciones (Moctezuma-Ochoa, 2016).

2.1. Consideraciones de ubicación

La precisión de los procesos de detección humana depende tanto del rendimiento del dispositivo de captura de vídeo como de las características espaciales del lugar (Thacker et al., 2008).

Las características de ubicación que deben considerarse son: la altura de la ubicación y el

ángulo de inclinación de la cámara, así como el nivel de luminosidad presente en la zona de captura de vídeo (Vidalía et al., 2019).

Los videos de prueba fueron tomados a alturas de 4 y 8 metros y los ángulos de inclinación del dispositivo de captura de video fueron 50° y 90° (Díaz, 2012). Asimismo, el nivel de luminosidad presente en el momento de la captura de dichos videos fue medido con el uso de un luxómetro digital. (Hernández Tolosa C, Contreras Eugenio B, Torres Sánchez C. 2016).

2.2. Aplicación de las técnicas de detección

La detección de personas tiene como objetivo distinguir entre los objetos en movimiento en una imagen de vídeo para su posterior clasificación (Paul, Haque, & Chakraborty, 2013). En esta sección se estudian las técnicas de sustracción de fondo, detectores de cascada y flujo óptico y su codificación en el lenguaje de programación Python.

2.2.1 Sustracción de fondo

Detecta los cambios de posición de los objetos presentes en una imagen de vídeo. En términos generales, la técnica consiste en hacer una sustracción entre el fotograma de la imagen que se va a procesar y el fondo de la misma, obteniendo así el objeto presente en el fotograma de vídeo (Sobral & Vacavant, 2014). En la tabla 1, se muestran los procesos generales necesarios para la aplicación de la técnica de sustracción de fondo en Python.

Tabla 1. Algoritmo de sustracción de fondo

- 1: **definir** valor en pixeles de ancho y alto
- 2: **ubicar** una línea de prueba para conteo
- 3: **crear** vector para las detecciones
- 4: **inicializar** imagen de video
- 5: **inicializar** el algoritmo de sustracción de fondo de mezclas gaussianas
- 6: **crear** función para la posición de las detecciones
- 7: **mientras corre el video:**
 devolver el cuadro de video en curso
- 8: **binarizar** imagen de video
- 9: **aplicar** filtros morfológicos
- 10: **definir** umbral
- 11: **buscar** contornos en la imagen
- 12: **analizar** contornos
- 13: **validar** contornos respecto a los valores de ancho y alto
- 14: **dibujar** rectángulo a las detecciones
- 15: **si** los valores de la función están en el vector de detecciones

16: **+1** en el número de detecciones
17: **pausa**
18: **fin**

La figura 1, muestra la aplicación de la técnica de sustracción de fondo en los videos de prueba tomados.

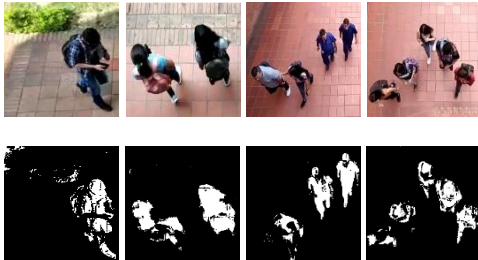


Fig. 1. Aplicación de la técnica de sustracción de fondo

2.2.2 Detectores en cascada

Acopla varios clasificadores con el fin de crear una estructura de varios niveles, acelerando así el proceso de detección. (Setjo, Achmad, & Faridah, 2017).

La clasificación es el proceso más importante en un detector de cascada. Por eso se requiere un conjunto de datos entrenados para que el objeto en cuestión sea filtrado paso a paso por el clasificador, mejorando así los resultados de las detecciones (Mahkonen, Virtanen, & Kämäräinen, 2018). En la tabla 2, se muestran los procesos generales necesarios para la aplicación de la técnica de detectores de cascada de cuerpo completo en Python.

Tabla 2. Algoritmo para un detector en cascada de cuerpo completo

1: **crear** dataset
2: **entrenar** dataset
3: **crear** vector para las detecciones
4: **inicializar** imagen de video
5: **inicializar** clasificador en cascada
6: **mientras corre el video:**
 devolver el cuadro de video en curso
7: **binarizar** imagen de video
8: **buscar** cuerpos completos
 definir parámetros
9: **dibujar** rectángulo a las detecciones
10: **si** los valores de la función están en el vector de detecciones
11: **+1** en el número de detecciones
 pausa
12: **fin**

En la figura 2, se observa la aplicación del algoritmo de detector en cascada de cuerpo completo para los videos de prueba capturados.



Fig. 2. Aplicación de un detector en cascada de cuerpo completo

2.2.3 Flujo óptico

Obtiene las estructuras correspondientes a la variación de las velocidades de los patrones de luminosidad en una imagen (Min & Huang, 2016). Aunque permite obtener información relevante de los objetos, es susceptible variaciones en la luminosidad de la escena de vídeo (Sharmin & Brad, 2012). En la tabla 3, se muestran los procesos generales requeridos para la aplicación de la técnica de flujo óptico en Python.

Tabla 3. Algoritmo de flujo óptico

1: **definir** parámetros para buscar esquinas
2: **definir** parámetros para el flujo
3: **crear** vector para las detecciones
4: **crear** colores aleatorios
5: **inicializar** imagen de video
6: **buscar** esquinas en el primer cuadro
7: **crear** máscara
8: **mientras corre el video:**
 devolver el cuadro de video en curso
9: **binarizar** imagen de video
10: **calcular** el flujo óptico
11: **dibujar** línea para las detecciones
12: **si** los valores de la función están en el vector de detecciones
13: **+1** en el número de detecciones
 pausa
14: **fin**

Así mismo, la figura 3 muestra la aplicación de la técnica de flujo óptico en los videos de prueba tomados.

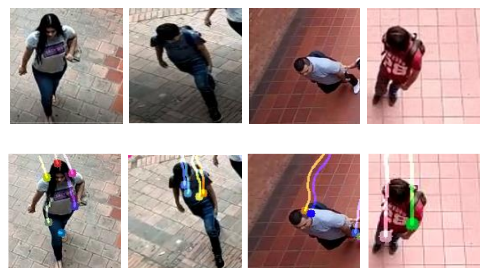


Fig. 3. Flujo óptico

vídeos utilizados en la aplicación de las técnicas de detección de personas.

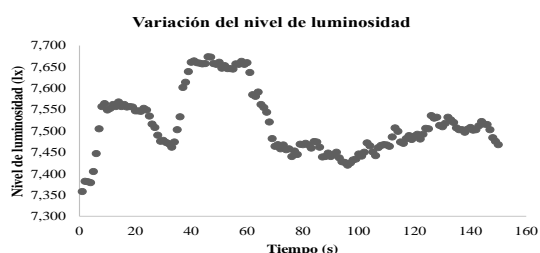


Fig. 4. Variación del nivel de luminosidad

El nivel de luminosidad obtenido fluctúa entre 7358 y 7674 luxes. Esta diferencia de 316 luxes no se considera relevante si se tiene en cuenta que el nivel de luminosidad toma como referencia una superficie en la que el flujo luminoso se distribuye uniformemente.

2.3. Comparativa

El método propuesto para comparar los resultados de los algoritmos de las técnicas de

Además, en la figura 4, se muestra la variación del nivel de luminosidad obtenida al grabar los videos de detección de personas se basa en la determinación del número de aciertos en las detecciones, así como el número de falsos positivos y falsos negativos (Guerrero Balaguera, 2016). Estos valores se determinan para cada una de las tres técnicas en estudio, tanto para las dos alturas de localización como para los dos ángulos de inclinación del dispositivo de captura de video, así como el nivel de luminosidad medio en el momento de la captura de cada video, a fin de determinar el porcentaje de precisión de cada una de las técnicas.

3. RESULTADOS

En la tabla 4, se muestra la comparación de las técnicas de sustracción de fondo, detectores en cascada y flujo óptico, respecto al número de detecciones realizadas. En ella, se muestra la precisión de cada una de las técnicas teniendo en cuenta el número de falsos positivos y falsos negativos, en relación al número de personas que intervienen en el cuadro de video en análisis.

Tabla 4. Comparativa de las técnicas de detección de personas

Técnica	Altura	Ángulo de inclinación	Número real de personas	Detecciones realizadas	Falsos positivos	Falsos negativos	Exactitud
Sustracción de fondo	4m	50°	60	59	0	1	98.34%
Sustracción de fondo	4m	90°	60	56	1	3	94.73%
Sustracción de fondo	8m	50°	60	55	2	3	94%
Sustracción de fondo	8m	90°	60	54	1	5	91.5%
Detectores en cascada	4m	50°	60	57	1	4	93.12%
Detectores en cascada	4m	90°	60	54	0	6	90%
Detectores en cascada	8m	50°	60	55	2	7	87.86%
Detectores en cascada	8m	90°	60	54	1	7	88.21%
Flujo óptico	4m	50°	60	45	1	16	73.28%
Flujo óptico	4m	90°	60	48	1	13	78.26%
Flujo óptico	8m	50°	60	52	3	11	81%
Flujo óptico	8m	90°	60	53	1	8	86.56%

La técnica de sustracción de fondo aplicada con la cámara de video ubicada a 4 metros de altura con ángulo de inclinación de 50° presenta mejores resultados en comparación con las demás configuraciones de ubicación, puesto que, de las 60 personas presentes en el video analizado, logró la detección de 59 de ellas. Para este caso, no se registra ningún caso de falso positivo, garantizando que cada una de las detecciones efectuadas correspondan a personas. Por otra parte, la técnica de flujo óptico probada a una altura de 4 metros con ángulo de inclinación de 50° presenta los resultados con el mayor número de desaciertos, ya que, de las 60 personas a detectar, se realizó la detección correcta solo de 44 de ellas, sumado a que una de las detecciones correspondió a un falso positivo.

No obstante, en términos generales, las 3 técnicas de detección de personas presentan mejorías en el número de aciertos en las detecciones al ser implementadas en alturas de 8 metros con ángulo de inclinación del dispositivo de captura de video de 90°, ya que presentan una exactitud medida de 88.76%, seguida con 88.25% de la implementación a una altura de 4 metros con 50° de inclinación, por su parte, la implementación a 4 metros de altura con inclinación de 90° y a 8 metros de altura con 50° de inclinación, presentan rendimiento promedios de 87.66% y 87.62%.

4. CONCLUSIONES

Al caracterizar el espacio en el cual se ubicará el dispositivo de captura, se aumenta el porcentaje de aciertos en las detecciones. Para la técnica de sustracción de fondo que fue la de mejor rendimiento, permitió pasar de 91,5% de exactitud al estar ubicada a 8 metros de altura y con inclinación de 90°, a 98.34% si se ubica a 4 metros de altura con inclinación de 50°.

Las técnicas de sustracción de fondo y detectores en cascada al tener en promedio el mayor porcentaje de aciertos en las detecciones con 94.64% y 89.79% respectivamente, pueden ser aplicadas en procesos de video vigilancia y conteo automático de personas. Por su parte, la técnica de flujo óptico al no tener un porcentaje promedio de acierto superior a 80%, no es recomendable su uso en procesos de detección de personas en tiempo real.

REFERENCIAS

- Díaz, A. (2012). Cámaras de medición. *Revista Universidad EAFIT*, 26, 7–16. Retrieved from <http://publicaciones.eafit.edu.co/index.php/revista-universidad-eafit/article/view/1528>
- Franco, C. E., Ospina, C. T., Cuevas, E. S., & Capacho, D. V. (2017). RECONOCIMIENTO FACIAL BASADO EN EIGENFACES, LBHP Y FISHERFACES EN LA BEAGLEBOARD-xM. *Revista Colombiana De Tecnologias De Avanzada (Rcta)*, 2(26). <https://doi.org/10.24054/16927257.v26.n26.2015.2387>
- García S, I., & Caranqui S, V. (2015). La visión artificial y los campos de aplicación. *Tierra Infinita*, 1, 94–103. <https://doi.org/2631-2921>
- Guerrero Balaguera, J. D. (2016). Algoritmos De Procesamiento De Imágenes Y Redes Neuronales Artificiales Para El Reconocimiento De La Lengua De Señas Colombiana (Lsc) Image Processing Algorithms and Artificial Neural Networks To Recognition of Colombian Sign Language (Lsc). *Revista Colombia de Tecnologías de Avanzada RCTA*, 1–8.
- Hernández Tolosa C, Contreras Eugenio B, Torres Sánchez C. (2016). Desarrollo de libros electrónicos: “taller pedagógico”. *Revista Tecnologías de Avanzada*, ISSN: 1692-7257
- Jeon, E. S., Choi, J. S., Lee, J. H., Shin, K. Y., Kim, Y. G., Le, T. T., & Park, K. R. (2015). Human detection based on the generation of a background image by using a far-infrared light camera. *Sensors (Switzerland)*, 15(3), 6763–6788. <https://doi.org/10.3390/s150306763>
- Kajabad, E. N., & Ivanov, S. V. (2019). People Detection and Finding Attractive Areas by the use of Movement Detection Analysis and Deep Learning Approach. *Procedia Computer Science*, 156, 327–337. <https://doi.org/10.1016/j.procs.2019.08.209>
- Leo, M., Medioni, G., Trivedi, M., Kanade, T., & Farinella, G. M. (2017). Computer vision for assistive technologies. *Computer Vision and Image Understanding*, 154, 1–15. <https://doi.org/10.1016/j.cviu.2016.09.001>
- Mahkonen, K., Virtanen, T., & Kämäräinen, J. (2018). Cascade of Boolean detector combinations. *Eurasip Journal on Image and Video Processing*, 2018(1). <https://doi.org/10.1186/s13640-018-0303-9>

- Min, Q., & Huang, Y. (2016). Motion detection using binocular image flow in dynamic scenes. *Eurasip Journal on Advances in Signal Processing*, 2016(1). <https://doi.org/10.1186/s13634-016-0349-8>
- Moctezuma-Ochoa, D. A. (2016). Re-identificación de personas a través de sus características soft-biométricas en un entorno multi-cámara de video-vigilancia. *Ingeniería, Investigación y Tecnología*, 17(2), 257–271. <https://doi.org/10.1016/j.riit.2016.06.010>
- Paul, M., Haque, S. M. E., & Chakraborty, S. (2013). Human detection in surveillance videos and its applications - a review. *Eurasip Journal on Advances in Signal Processing*, 2013(1), 1–16. <https://doi.org/10.1186/1687-6180-2013-176>
- Pérez, J, Castro, J (2018). LRS1: un robot social de bajo costo para la asignatura “Programación 1”. *Revista Tecnologías de Avanzada*, ISSN: 1692-7257
- Setjo, C. H., Achmad, B., & Faridah. (2017). Thermal image human detection using Haar-cascade classifier. *Proceedings - 2017 7th International Annual Engineering Seminar, InAES 2017*, (January 2020). <https://doi.org/10.1109/INAES.2017.8068554>
- Sharmin, N., & Brad, R. (2012). Optimal filter estimation for Lucas-Kanade optical flow. *Sensors (Switzerland)*, 12(9), 12694–12709. <https://doi.org/10.3390/s120912694>
- Sobral, A., & Vacavant, A. (2014). A comprehensive review of background subtraction algorithms evaluated with synthetic and real videos. *Computer Vision and Image Understanding*, 122, 4–21. <https://doi.org/10.1016/j.cviu.2013.12.005>
- Thacker, N. A., Clark, A. F., Barron, J. L., Ross Beveridge, J., Courtney, P., Crum, W. R., ... Clark, C. (2008). Performance characterization in computer vision: A guide to best practices. *Computer Vision and Image Understanding*, 109(3), 305–334. <https://doi.org/10.1016/j.cviu.2007.04.006>
- Vidalía, H., Sandoval, G., Rogelio, L., Gill, V., Curiel, A. A., Alberto, F., ... L, A. M. (2019). Una estrategia didáctica interactiva e interdisciplinar para el proceso de aprendizaje. *Técnica Estenoipeica; Reproducción Visual de La Realidad.*, 1(2), 9–27. <https://doi.org/10.22517/25393812.21971>