

# REVISIÓN DE MÉTODOS DE RECUPERACIÓN DE VIDEO CON BASE EN CONTENIDO

## REVIEW OF CONTENT-BASED VIDEO RETRIEVAL METHODS

Ing. Gustavo Steve Luis Peña Riquer<sup>1</sup>,

Dra. Lisbeth Rodríguez Mazahua<sup>2</sup>

Dr. Mario Leoncio Arrijoja Rodríguez<sup>3</sup>

Dr. Ulises Juárez Martínez<sup>4</sup>

M.C. María Antonieta Abud Figueroa<sup>5</sup>

### RESUMEN

El crecimiento explosivo de datos multimedia, específicamente video, genera la necesidad de desarrollar métodos de recuperación basada en contenido. Este artículo presenta los resultados de una revisión de 32 artículos relacionados con recuperación de video basada en contenido (CBVR). Además, se propone una arquitectura de un módulo CBVR que se integrará a un sistema para la gestión de datos históricos del Instituto Tecnológico de Orizaba. Las tecnologías consideradas para el desarrollo del módulo son el lenguaje de programación Java con el marco de trabajo JavaServer Faces, el sistema gestor de bases de datos Cottontail DB, la biblioteca de visión artificial OpenCV y el descriptor SURF (*Speeded-Up Robust Features*, Funciones robustas aceleradas).

### ABSTRACT

The explosive growth of multimedia data, specifically video, generates the need to develop content-based retrieval methods. This paper presents the results of a review of 32 papers related to content-based video retrieval (CBVR). In addition, an architecture of a CBVR module is proposed that will be integrated to a system for the management of historical data of the Instituto Tecnológico de Orizaba. The technologies considered for the development of this module are the Java programming language with the JavaServer Faces framework, the database management system Cottontail DB, the open computer vision library (OpenCV), and the descriptor SURF (*Speeded-Up Robust Features*).

*Palabras clave:* CBVR, HITO, SURF, Base de datos multimedia

*Key words:* CBVR, HITO, SURF, Multimedia database

### INTRODUCCIÓN

Una base de datos multimedia es una colección controlada de elementos de datos multimedia, como texto, imágenes, objetos gráficos, bocetos, video y audio [1]. Es recomendable emplear una base de datos multimedia por sus características que facilitan la recuperación basada en contenido. La recuperación de video con base en contenido (CBVR) es el método donde los videos más relevantes para la consulta se obtienen desde conjuntos de datos con base en características (contenido del video) extraídas del video y la consulta [2].

En el Instituto Tecnológico de Orizaba, se recuperaron una gran cantidad de datos multimedia, los cuales se almacenaron en una base de datos llamada HITO (Historia del Instituto Tecnológico de

<sup>1</sup> Ing. Sistemas Computacionales Gustavo Steve Luis Peña Riquer, Estudiante de la maestría en Sistemas Computacionales del Instituto Tecnológico de Orizaba [msc.gsteveluis@ito-depi.edu.mx](mailto:msc.gsteveluis@ito-depi.edu.mx)

<sup>2</sup> Dra. Lisbeth Rodríguez Mazahua, Profesor investigador de la maestría en Sistemas Computacionales del Instituto Tecnológico de Orizaba, [lrodriguez@ito-depi.edu.mx](mailto:lrodriguez@ito-depi.edu.mx)

<sup>3</sup> Dr. Mario Leoncio Arrijoja Rodríguez, Profesor investigador de la maestría en Ing. Industrial del Instituto Tecnológico de Orizaba, [marriojar@ito-depi.edu.mx](mailto:marriojar@ito-depi.edu.mx)

<sup>4</sup> Dr. Ulises Juárez Martínez, Profesor investigador de la maestría en Sistemas Computacionales del Instituto Tecnológico de Orizaba, [ulises.jm@orizaba.tecnm.mx](mailto:ulises.jm@orizaba.tecnm.mx)

<sup>5</sup> M.C. María Antonieta Abud Figueroa, Profesor investigador de la maestría en Sistemas Computacionales del Instituto Tecnológico de Orizaba, [maria.af@orizaba.tecnm.mx](mailto:maria.af@orizaba.tecnm.mx)

Orizaba) y se desarrolló un sistema para la gestión de estos datos [3] el cual trabaja en base al contenido. Actualmente, se agregaron videos de eventos relevantes del Instituto a HITO, por lo que es necesario extender el sistema para que facilite la gestión de los videos en la base de datos. Adicionalmente, se desarrollará un módulo para la recuperación de videos con base en contenido que tendrá ventajas con respecto a las consultas basadas en texto. En este artículo, se presentan los resultados de una revisión de 32 artículos relacionados con CBVR y se propone la arquitectura del módulo CBVR.

Este trabajo se organiza en cuatro secciones principales: en la primera sección se abarca la introducción y conceptos básicos del artículo; la segunda sección contiene la metodología de análisis; la tercera sección comprende los resultados de la revisión y la descripción de la arquitectura propuesta; finalmente, la última sección contiene la conclusión y el trabajo a futuro.

## METODOLOGÍA DE ANÁLISIS

La Figura 1 muestra la metodología de análisis utilizada, se encontraron 32 artículos relacionados con CBVR en las principales bibliotecas digitales de Ciencias de la Computación (IEEE Xplore, ACM digital library, SpringerLink, ScienceDirect, entre otras). Posteriormente, los trabajos se analizaron con base en seis criterios: 1) El tipo de recuperación de videos que consideran; 2) El tipo de consulta que realizan: por video, por imagen, entre otras; 3) Las tecnologías que utilizan; 4) El conjunto de datos que emplean; 5) Los descriptores usados, y 6) Las medidas de similitud. Solo se eligieron trabajos publicados en los últimos cinco años (2018-2022) en revistas o memorias de congresos y se excluyeron tesis de maestría o doctorado, libros y artículos que no están escritos en inglés.

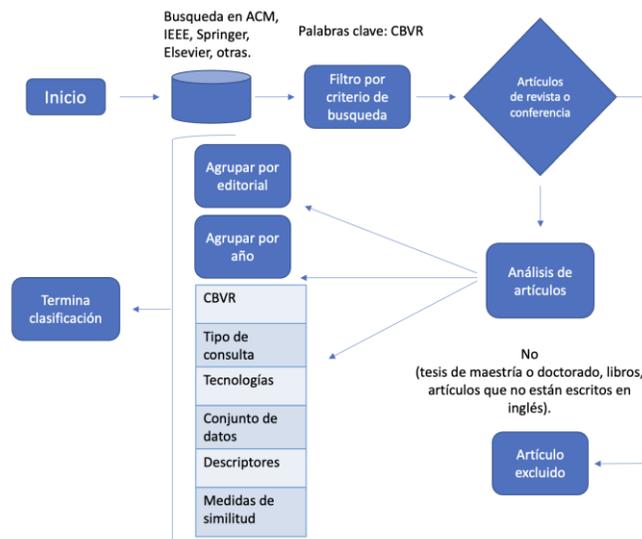


Figura 1 Metodología de análisis.

La Tabla 1 muestra la comparación de los trabajos analizados con la propuesta del módulo CBVR que se presenta en este artículo (por cuestiones de espacio solo se muestran 21 artículos). El módulo CBVR propuesto tomará una imagen como entrada para recuperar los videos más similares, utilizará el descriptor SURF (*Speeded Up Robust Features*, Características Robustas Aceleradas), empleará el Sistema Gestor de Bases de Datos (SGBD) multimedia Cottontail DB, usará la distancia euclidiana y considera datos reales de la base de datos HITO. Por lo tanto, se tomaron en cuenta seis criterios: 1) Recuperación de video basada en contenido; 2) Ejecución de consultas por imagen; 3) Utilización del descriptor SURF; 4) Empleo de un SGBD multimedia; 5) Uso de la distancia Euclidiana, y 6) Consideración de datos reales.

*Tabla 1. Análisis comparativo de los trabajos analizados*

Artículo	1	2	3	4	5	6
Kumar et al. [1]	No	Sí	No	No	No	No
Saodi y Jai Andaloussi [2]	Sí	No	No	No	No	No
Ram et al. [4]	Sí	Sí	No	No	No	Sí
Araujo y Girod [5]	Sí	Sí	No	No	No	No
Hamad et al. [6]	Sí	No	No	No	Sí	Si
Mühling et al. [7]	Sí	No	No	No	Sí	Si
Kaliaperumal et al. [8]	Sí	No	No	No	No	Sí
Chen et al. [9]	Sí	No	No	No	No	No
Zhang et al. [10]	Sí	Sí	No	No	No	Sí
Prathiba y Kumari [11]	Sí	No	No	No	No	No
Tseytlin y Makarov [12]	Sí	No	No	No	No	Sí
Mounika y Khare [13]	Sí	Sí	No	No	No	Sí
Potluri et al. [14]	Sí	Sí	Sí	No	No	No
Shao et al. [15]	Sí	No	No	No	No	No
Jo et al. [16]	Sí	No	No	No	No	No
Zhou et al. [17]	No	Si	No	No	No	No
Soniminde y Thepade [18]	Sí	No	No	No	No	No
Yasin et al. [19]	Sí	No	No	No	No	Sí
Shao et al. [20]	No	Sí	No	No	No	No
Zhang et al. [21]	No	No	No	No	No	Sí
Guisepe Amato et al. [22]	Sí	No	No	No	No	No
Este trabajo	Sí	Sí	Sí	Sí	Sí	Sí

Como se observa en la Tabla 1, no se encontró algún trabajo que cumpla con los seis criterios considerados, la mayoría de los métodos CBVR no consideran un SGBD especializado en la administración de datos multimedia y utilizan datos sintéticos, mientras que el módulo propuesto permitirá recuperar videos con base en contenido de eventos relevantes del Instituto Tecnológico de Orizaba y utilizará el SGBD multimedia Cottontail DB.

## RESULTADOS

En esta sección se presentan los resultados de la revisión y la arquitectura propuesta del módulo CBVR.

**Tipo de recuperación.** 22 artículos analizados usaron como tipo de recuperación CBVR. Otro tipo de recuperación considerada fue recuperación de video a gran escala.

**Editorial.** En la Figura 2 se observa que se encontraron más artículos en la editorial IEEE, la mayoría utilizan un video como entrada para recuperar los videos similares.

**Artículos por año.** La Figura 3 muestra que en 2019 y 2021 se publicaron más artículos, de los 8 artículos del 2019 al menos 5 usaron como descriptor características CNN (Redes Neuronales Convolucionales) y el resto utilizaron HOG (Histogramas de Gradientes Orientados) y SURF, mientras que en el 2021 solo usaron características CNN.

**Tipo de consulta.** En la Figura 4 se muestra los tipos de consulta empleados en los artículos analizados, donde el tipo de consulta más frecuente es por video.

**Tecnologías.** La Figura 5 muestra las tecnologías utilizadas que presentaron los trabajos analizados, donde se observa que CNN fue la más empleada.

**Conjuntos de datos.** 21 de 32 métodos CBVR usaron *benchmarks* en su validación. La Figura 6 presenta los *benchmarks* utilizados.

**Medida de similitud.** En la Figura 7, el gráfico demuestra que las medidas de similitud más usadas en los trabajos analizados fueron: Distancia Euclidiana, Similitud Coseno y Distancia Hamming.

**Descriptor.** En la Figura 8 se presentan los descriptores más empleados en los trabajos analizados. El descriptor más utilizado son las características CNN.



Figura 2 Artículos por editorial.



Figura 3 Artículos por año.

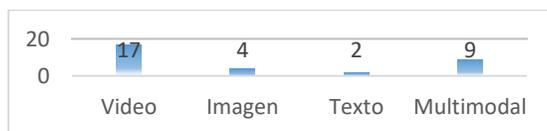


Figura 4 Tipo de consulta.

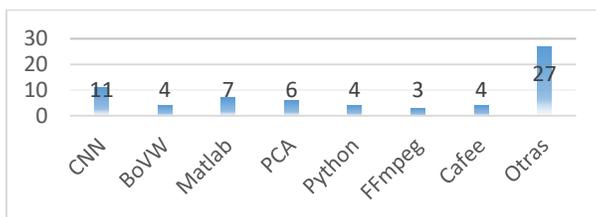


Figura 5 Tecnologías.

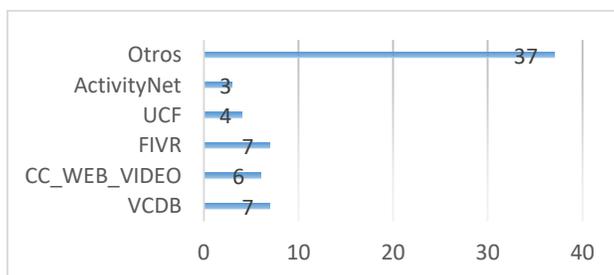


Figura 6 Benchmarks.

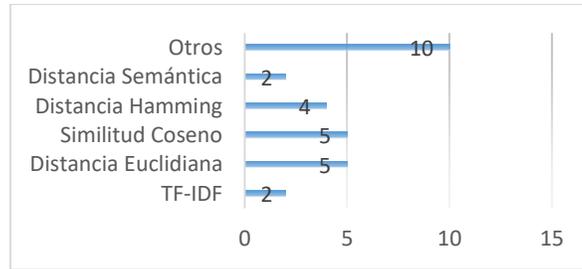


Figura 7 Medida de similitud.

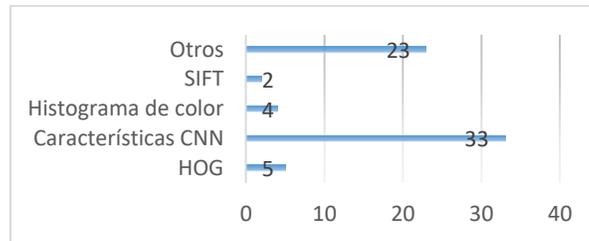


Figura 8 Descriptores.

## Arquitectura

La arquitectura del sistema se diseñó con base en el patrón arquitectónico MVC (*model view control*, modelo vista controlador), este distribuye los componentes del sistema de forma que facilite su mantenimiento y se representa de manera abstracta en la Figura 9. Los componentes de la arquitectura propuesta que se describen a continuación.

**Modelo:** Agrupa la lógica del sistema representada en los *beans* administrados que tienen acceso a los componentes de la interfaz y pasan información a los *beans* de modelo, estos últimos representan las clases importantes del dominio y usan las bibliotecas, como la de OpenCV para realizar la búsqueda basada en contenido (videos) y de JDBC (*Java Database Connectivity*, Conectividad a bases de datos de Java) para controlar el acceso al sistema gestor de bases de datos Cottontail DB y manipular la información.

**Vista:** Mediante archivos XHTML (*eXtensible HyperText Markup Language*, lenguaje de marcado de hipertexto extensible) se representa el modelo y maneja la interacción con el usuario; para este propósito se utilizan etiquetas propias de JSF (*JavaServer Faces*), en conjunto con hojas de estilo (CSS) para dar una mejor visualización al usuario.

**Controlador:** El *Servlet* de JSF es el vínculo entre el modelo y la vista. Se ocupa de gestionar las peticiones de los recursos accediendo al modelo requerido en la petición del usuario y seleccionando la vista adecuada para representarlo.

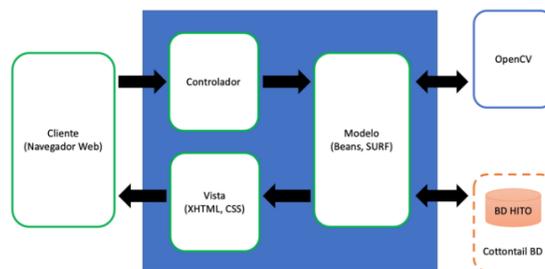


Figura 9 Arquitectura propuesta.

Para el diseño de la arquitectura se seleccionó el SGBD Cottontail DB [23] por su capacidad para gestionar y consultar grandes colecciones multimedia. Se optó por JSF [24] por la flexibilidad para crear aplicaciones con el lenguaje puro de Java [25], su facilidad de desarrollo y que también es gratuito. Se eligió NetBeans [26] como entorno de desarrollo integrado por la facilidad para desarrollar aplicaciones con el marco de trabajo seleccionado. Se decide usar la metodología de desarrollo UWE (*Unified Modeling Language Web Engineering*, Ingeniería Web del Lenguaje Unificado de Modelado) [27] porque se especializa en el desarrollo de aplicaciones web y multimedia. Por último, se seleccionó OpenCV (*Open Computer Vision*, Visión Artificial Abierta) [28] porque es una biblioteca libre de visión artificial y que tiene aplicaciones en detección de movimiento, reconocimiento de objetos, reconstrucción 3D a partir de imágenes, entre otros.

## CONCLUSIONES

La recuperación de videos con base en contenido (CBVR) es relevante en las bases de datos multimedia, ya que permite obtener videos similares proporcionando una imagen o video como entrada, esto logra resultados más precisos si se compara con las consultas basadas en texto. Por tal motivo, es este proyecto se presentaron los resultados de la revisión de 32 artículos relacionados con CBVR y se propuso una arquitectura de un módulo CBVR que se agregará al sistema de gestión de datos multimedia HITO.

Con este trabajo se beneficiarán en gran medida los usuarios del Instituto Tecnológico de Orizaba, provocando un impacto social, que pretende crear conciencia sobre la historia del instituto. También generará un impacto económico, ya que, al utilizar un método CBVR se obtendrán mejores resultados en cuanto a consultas de videos, permitiendo el acceso al contenido multimedia solicitado por los usuarios (profesores, estudiantes, entre otros) de una manera más efectiva.

El trabajo a futuro es diseñar el módulo de acuerdo con la metodología para desarrollo de software, desarrollar el módulo del sistema para recuperar videos de HITO con base en contenido y validar el módulo por medio de un caso de estudio.

## AGRADECIMIENTOS.

Los autores de este trabajo agradecen el apoyo por parte del consejo Nacional de Ciencia y Tecnología (CONACyT) por el soporte financiero al proyecto A1-S-51808 y al Tecnológico Nacional de México por permitir el desarrollo de este proyecto.

## REFERENCIAS BIBLIOGRÁFICAS

- [1] V. Kumar, V. Tripathi, and B. Pant, "Content based Surgical Video Retrieval via Multi-Deep Features Fusion," in 2021 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT), 2021, pp. 1–5, doi: 10.1109/conecct52877.2021.9622562.
- [2] E. M. Saoudi and S. Jai-Andaloussi, "A distributed Content-Based Video Retrieval system for large datasets," J. Big Data, vol. 8, no. 1, 2021, doi: 10.1186/s40537-021-00479-x.
- [3] M. J. Rodríguez-Arauz, L. Rodríguez Mazahua, M. L. Arrijoja-Rodríguez, M. A. Abud-Figueroa, S. G. Peláez-Camarena, and L. del C. Martínez-Méndez, "Design of a Multimedia Data Management System that Uses Horizontal Fragmentation to Optimize Content-based Queries," in IMM 2020: The Tenth International Conference on Advances in Information Mining and Management, 2020, pp. 15–21.
- [4] R. S. Ram, S. A. Prakash, M. Balaanand, and C. B. Sivaparthipan, "Colour and orientation of pixel based video retrieval using IHBM similarity measure," Multimed. Tools Appl., vol. 79, no. 15–16, pp. 10199–10214, 2020, doi: 10.1007/s11042-019-07805-9.
- [5] A. Araujo and B. Girod, "Large-Scale Video Retrieval Using Image Queries," IEEE Trans.

Circuits Syst. Video Technol., vol. 28, no. 6, pp. 1406–1420, 2018, doi: 10.1109/TCSVT.2017.2667710.

[6] S. Hamad, A. S. Farhan, and D. Y. Khudhur, “Content based video retrieval using discrete cosine transform,” *Indones. J. Electr. Eng. Comput. Sci.*, vol. 21, no. 2, pp. 839–845, 2020, doi: 10.11591/ijeecs.v21.i2.pp839-845.

[7] M. Mühling et al., “Content-based video retrieval in historical collections of the German Broadcasting Archive,” *Int. J. Digit. Libr.*, vol. 20, no. 2, pp. 167–183, 2019, doi: 10.1007/s00799-018-0236-z.

[8] N. Kaliaperumal, A. Das, and V. Balakrishnan, “A Content-Based Retrieval Model with Combinational Features and Indexing for Distributed Video Objects,” *Int. J. Eng. Res. Technol.*, vol. 13, no. 12, pp. 5142–5148, 2020, [Online]. Available: [http://www.irphouse.com/ijert20/ijertv13n12\\_148.pdf](http://www.irphouse.com/ijert20/ijertv13n12_148.pdf).

[9] H. Chen et al., “A supervised video hashing method based on a deep 3d convolutional neural network for large-scale video retrieval,” *Sensors*, vol. 21, no. 9, pp. 1–15, 2021, doi: 10.3390/s21093094.

[10] C. Zhang, Y. Lin, L. Zhu, A. Liu, Z. Zhang, and F. Huang, “CNN-VWII: An efficient approach for large-scale video retrieval by image queries,” *Pattern Recognit. Lett.*, vol. 123, pp. 82–88, 2019, doi: 10.1016/j.patrec.2019.03.015.

[11] T. Prathiba and R. S. S. Kumari, “Content based video retrieval system based on multimodal feature grouping by KFCM clustering algorithm to promote human–computer interaction,” *J. Ambient Intell. Humaniz. Comput.*, vol. 12, no. 6, pp. 6215–6229, 2021, doi: 10.1007/s12652-020-02190-w.

[12] B. Tseytlin and I. Makarov, “Content based video retrieval system for distorted video queries,” *CEUR Workshop Proc.*, vol. 2795, pp. 99–107, 2020.

[13] R. M. Bommisetty and A. Khare, “Content based video retrieval using histogram of gradients and frame fusion,” no. January, p. 65, 2020, doi: 10.1117/12.2558559.

[14] T. Potluri and N. Ganeswara Rao, “Content Based Video Retrieval Using SURF, BRISK and HARRIS Features for Query-by-image,” *Commun. Comput. Inf. Sci.*, vol. 1035, pp. 265–276, 2019, doi: 10.1007/978-981-13-9181-1\_24.

[15] J. Shao, X. Wen, B. Zhao, and X. Xue, “Context Encoding for Video Retrieval with Contrastive Learning,” *Proc. - 2021 IEEE Winter Conf. Appl. Comput. Vision, WACV 2021*, no. August, pp. 3267–3277, 2021, doi: 10.1109/WACV48630.2021.00331.

[16] W. Jo, G. Lim, J. Kim, J. Yun, and Y. Choi, “Exploring the Temporal Cues to Enhance Video Retrieval on Standardized CDVA,” *IEEE Access*, vol. 10, pp. 38973–38981, 2022, doi: 10.1109/ACCESS.2022.3165177.

[17] Y. Zhou, M. Wang, R. Wang, and S. Huo, “Graph Neural Network for Video Relocalization.” *arXiv*, 2020, doi: 10.48550/ARXIV.2007.09877.

[18] N. V. Soniminde and S. D. Thepade, “Performance Improvement of Content-Based Video Retrieval using Thepade’s Sorted n-ary Block Truncation Coding(TSnBTC) with Various Color Spaces,” in *2021 International Conference on Communication information and Computing Technology (ICCICT)*, 2021, pp. 1–5, doi: 10.1109/iccict50803.2021.9509934.

[19] D. Yasin, A. Sohail, and I. Siddiqi, “Semantic Video Retrieval using Deep Learning Techniques,” *Proc. 2020 17th Int. Bhurban Conf. Appl. Sci. Technol. IBCAST 2020*, pp. 338–343, 2020, doi: 10.1109/IBCAST47879.2020.9044601.

- [20] J. Shao, X. Wen, B. Zhao, and X. Xue, “Temporal Context Aggregation for Video Retrieval with Contrastive Learning,” 2021, doi: 10.48550/ARXIV.2008.01334.
- [21] H. Zhang, Y. Ji, W. Huang, and L. Liu, “Sitcom-star-based clothing retrieval for video advertising: a deep learning framework,” *Neural Comput. Appl.*, vol. 31, no. 11, pp. 7361–7380, 2019, doi: 10.1007/s00521-018-3579-x.
- [22] G. Amato et al., “The vision video search system: Exploiting off-the-shelf text search engines for large-scale video retrieval,” *J. Imaging*, vol. 7, no. 5, 2021, doi: 10.3390/jimaging7050076.
- [23] R. Gasser, L. Rossetto, S. Heller, and H. Schuldt, “Cottontail DB: An Open Source Database System for Multimedia Retrieval and Analysis,” in *Proceedings of the 28th ACM International Conference on Multimedia*, 2020, pp. 4465–4468, doi: 10.1145/3394171.3414538.
- [24] “JavaServer Faces Technology.” [Online]. Available: <https://www.oracle.com/java/technologies/javaserverfaces.html>.
- [25] “Java | Oracle.” [Online]. Available: <https://www.java.com/es/>.
- [26] “Welcome to Apache NetBeans.” [Online]. Available: <https://netbeans.apache.org/>.
- [27] O. U. Aldana, J. P. U. Pech, and R. A. A. Vera, “Una Herramienta para el Análisis de la Colaboración Diseñada con UWE,” *Rev. Latinoam. Ing. Softw.*, vol. 4, no. 6, pp. 235–242, 2016, doi: 10.18294/RELAIS.2016.235-242.
- [28] “Home - OpenCV.” [Online]. Available: <https://opencv.org/>.