

Universidad Autónoma de Querétaro

Facultad de Informática

Maestría en Ciencia de Datos

Título del tema de tesis registrado:

Modelo de sustracción de fondo basado en morfología matemática

Que como parte de los requisitos para obtener el grado de
Maestro en Ciencia de Datos

Presenta

Ing. Moreno Mejía Erik Josué

Dirigido por:

Dr. Hugo Jiménez Hernández

Dr. Hugo Jiménez Hernández

Presidente

Dr. Alejandro Gómez Hernández

Secretario

M.C. Daniel Cantón Enríquez

Vocal

Dra. Teresa García Ramírez

Suplente

Dra. Ana Marcela Herrera Navarro

Suplente

Centro Universitario, Querétaro, Qro.

04 de junio del 2025

México

La presente obra está bajo la licencia:
<https://creativecommons.org/licenses/by-nc-nd/4.0/deed.es>



CC BY-NC-ND 4.0 DEED

Atribución-NoComercial-SinDerivadas 4.0 Internacional

Usted es libre de:

Compartir — copiar y redistribuir el material en cualquier medio o formato

La licenciante no puede revocar estas libertades en tanto usted siga los términos de la licencia

Bajo los siguientes términos:



Atribución — Usted debe dar [crédito de manera adecuada](#), brindar un enlace a la licencia, e [indicar si se han realizado cambios](#). Puede hacerlo en cualquier forma razonable, pero no de forma tal que sugiera que usted o su uso tienen el apoyo de la licenciante.



NoComercial — Usted no puede hacer uso del material con [propósitos comerciales](#).



SinDerivadas — Si [remezcla, transforma o crea a partir](#) del material, no podrá distribuir el material modificado.

No hay restricciones adicionales — No puede aplicar términos legales ni [medidas tecnológicas](#) que restrinjan legalmente a otras a hacer cualquier uso permitido por la licencia.

Avisos:

No tiene que cumplir con la licencia para elementos del material en el dominio público o cuando su uso esté permitido por una [excepción o limitación](#) aplicable.

No se dan garantías. La licencia podría no darle todos los permisos que necesita para el uso que tenga previsto. Por ejemplo, otros derechos como [publicidad, privacidad, o derechos morales](#) pueden limitar la forma en que utilice el material.

RESUMEN

La detección precisa de vehículos en movimiento en entornos urbanos es un componente esencial para el desarrollo de sistemas inteligentes de monitoreo de tráfico. Las técnicas tradicionales de sustracción de fondo, como la Mezcla de Gaussianas (MoG), presentan limitaciones en escenarios reales donde la iluminación varía constantemente, hay presencia de ruido y las condiciones ambientales son inestables. En este trabajo se propone un modelo de sustracción de fondo basado en morfología matemática, que opera sobre secuencias de video en escala de grises y aprovecha operaciones como apertura, cierre y transformaciones top-hat para resaltar la textura local y separar eficazmente el primer plano del fondo.

El modelo fue diseñado para ser computacionalmente eficiente, interpretable y robusto frente a condiciones adversas, sin requerir procesos de entrenamiento o parámetros complejos. La propuesta fue validada mediante experimentos con videos reales capturados por dron en zonas urbanas, y evaluada comparativamente contra el método MoG utilizando métricas estándar como precisión, recall y F1-score. Los resultados muestran que el enfoque morfológico logra una segmentación más estable y confiable en escenarios con alta variabilidad de iluminación, destacando su potencial como alternativa ligera y efectiva en aplicaciones de visión por computadora.

Palabras clave: Morfología matemática, sustracción de fondo, visión por computadora, detección de movimiento, análisis de tráfico, procesamiento de imágenes, algoritmos robustos, entornos urbanos.

TABLA DE CONTENIDOS

Resumen.....	I
Contenido	II
Índice de Figuras	V
Índice de Tablas	VII
1 Introducción.....	9
1.1 DESCRIPCIÓN DEL PROBLEMA.....	9
1.2 JUSTIFICACIÓN.....	11
1.3 HIPÓTESIS	12
1.4 OBJETIVOS	13
1.4.1 <i>Objetivo general</i>	13
1.4.2 <i>Objetivos específicos</i>	13
1.5 CONTRIBUCIONES ESPERADAS.....	14
1.6 PUBLICACIONES	15
1.7 ESTRUCTURA DE LA TESIS	16
2 Antecedentes.....	18
2.1 VISIÓN POR COMPUTADORA EN ENTORNOS URBANOS.....	18
2.2 MODELOS CLÁSICOS DE SUSTRACCIÓN DE FONDO	19
2.2.1 <i>Mixture of Gaussians (MoG)</i>	19
2.2.2 <i>Visual Background Extractor (ViBe): Visual Backgropund Ex-</i> <i>tractor</i>	19
2.2.3 <i>Pixel-Based Adaptive Segmenter (PBAS): Pixel-Based Adap-</i> <i>tive Segmenter</i>	20
2.3 ENFOQUES BASADOS EN APRENDIZAJE PROFUNDO PARA SEGMENTACIÓN DE FONDO	20
2.4 APLICACIONES PREVIAS DE LA MORFOLOGÍA MATEMÁTICA EN VISIÓN POR COMPUTADORA	21
2.5 ALGORITMOS MORFOLÓGICOS EN SUSTRACCIÓN DE FONDO	22
2.6 USO DE DRONES EN MONITOREO URBANO	24
3 Marco Teórico	26

3.1	FUNDAMENTOS DE LA MORFOLOGÍA MATEMÁTICA.....	26
3.1.1	<i>Operadores morfológicos básicos</i>	27
3.1.2	<i>Transformaciones morfológicas derivadas</i>	29
3.1.3	<i>Modelo probabilístico basado en textura</i>	31
4	Propuesta de investigación	33
4.1	DESCRIPCIÓN GENERAL DEL MODELO PROPUESTO	33
4.2	PREPROCESAMIENTO DE LA IMAGEN	34
4.3	CONSTRUCCIÓN DEL MODELO DE FONDO.....	35
4.4	DETECCIÓN DEL PRIMER PLANO.....	37
4.5	POSTPROCESAMIENTO MORFOLÓGICO	38
4.5.1	<i>Comparación metodológica con MoG</i>	39
4.6	CONSIDERACIONES TÉCNICAS DE IMPLEMENTACIÓN.....	41
4.6.1	<i>Representación matricial del modelo de fondo</i>	41
4.6.2	<i>Criterio de pertenencia a la distribución</i>	42
4.6.3	<i>Consumo de memoria por resolución</i>	42
4.6.4	<i>Algoritmo de actualización del modelo</i>	43
4.6.5	<i>Mejora de conectividad en el mapa de movimiento</i>	44
5	Resultados preliminares	45
5.1	DESCRIPCIÓN DEL PROCESO DE VALIDACIÓN	45
5.2	EVALUACIÓN CON EL CONJUNTO PETS	47
5.2.1	<i>Resultados y análisis</i>	47
5.3	EVALUACIÓN CON EL CONJUNTO CDNET2014	51
5.3.1	<i>Resultados y análisis</i>	51
5.4	COMPARACIÓN VISUAL CUALITATIVA EN ROTONDAS	56
5.4.1	<i>Contexto urbano</i>	57
5.4.2	<i>Evaluación cualitativa</i>	58
5.4.3	<i>Observaciones clave</i>	60
5.5	MAPAS DE DENSIDAD DE TRÁFICO	60
5.5.1	<i>Generación del mapa de densidad</i>	60
5.5.2	<i>Resultados y análisis</i>	61
5.5.3	<i>Ventajas del enfoque morfológico</i>	61
5.6	DISCUSIÓN GENERAL DE RESULTADOS.....	62
5.6.1	<i>Desempeño cuantitativo</i>	62
5.6.2	<i>Desempeño cualitativo</i>	63
5.6.3	<i>Fortalezas y limitaciones</i>	63

6 Conclusiones	65
6.1 CUMPLIMIENTO DE OBJETIVOS	65
6.2 EVALUACIÓN DE LA HIPÓTESIS	66
6.3 CONTRIBUCIONES PRINCIPALES	66
6.4 LIMITACIONES Y POSIBLES MEJORAS	67
6.5 LÍNEAS FUTURAS DE INVESTIGACIÓN	67
6.6 REFLEXIÓN FINAL	68
Referencias	69

ÍNDICE DE FIGURAS

3.1	Representación de una imagen como malla discreta con conectividad vecinal (a) y como superficie tridimensional en un espacio de intensidad (b). Esta representación permite aplicar operadores morfológicos que trabajan sobre la textura de la imagen mediante relaciones de orden parcial. Fuente: Elaboración propia.	27
3.2	Caracterización morfológica de textura: (arriba) superficie aproximada mediante operadores morfológicos; (abajo) residuo que representa la información de textura. Fuente: Elaboración propia.	31
4.1	Diagrama general del flujo de procesamiento del modelo propuesto.	34
4.2	Ejemplo de reducción de ruido estructural mediante filtro de mediana morfológica. (a) Imagen original con ruido; (b) imagen procesada con filtro basado en apertura y cierre morfológico. Fuente: Elaboración propia.	35
4.3	Esquema del modelo probabilístico basado en textura propuesto. Se analiza la evolución temporal y espacial de la textura en cada píxel, permitiendo identificar regiones en movimiento con mayor robustez. Fuente: Elaboración propia.....	36
4.4	Representación matricial del modelo de textura por píxel mediante contadores.....	41
4.5	Criterio de pertenencia basado en la concavidad de la función de densidad empírica. Se utiliza el valor máximo global como referencia y se identifican los cambios de concavidad a izquierda (Δ_l) y derecha (Δ_r), delimitando el intervalo de pertenencia.	42
5.1	Error de sesgo en el conjunto de datos PETS (The University of Reading, 2001) entre MMBS y MoG; los gráficos del lado derecho representan la distribución del error de sesgo, la cual puede observarse para ambos enfoques. Una distribución simétrica centrada en cero indica un mejor desempeño en la detección de movimiento. La subfigura (a) corresponde a un escenario al aire libre; (b) y (c) representan escenas en interiores, con condiciones de iluminación artificial.....	49

5.2	Error RMSE en el conjunto de datos PETS (The University of Reading, 2001) entre MMBS y MoG; en este contexto, valores más pequeños indican mayor precisión. En (a), las condiciones al aire libre hacen que MOG sea menos preciso que la propuesta, y en (b) y (c), bajo condiciones en interiores, ambos enfoques presentan un desempeño similar.	50
5.3	Error de sesgo calculado a partir del conjunto de datos CDnet 2014 (Y. Wang et al., 2014) entre los enfoques MMBS y MoG mejorado, donde (a) representa la escena con mejor desempeño en la detección de movimiento, mientras que (b) representa la escena con el peor desempeño en dicha tarea.	55
5.4	Error RMSE calculado a partir del conjunto de datos CDnet 2014 (Y. Wang et al., 2014) utilizando los enfoques MMBS y MoG mejorado, donde (a) representa la escena con mejor desempeño en la detección de movimiento, mientras que (b) representa la escena con el peor desempeño en dicha tarea.	56
5.5	Escenario de prueba real para analizar el flujo vehicular, donde (a) muestra la ubicación a nivel estatal, (b) la ubicación a nivel ciudad, (c) la ubicación a nivel delegación, y (d) la ubicación de la glorieta. ...	58
5.6	Resultados obtenidos con el modelo MMBS en glorietas urbanas: se aprecia la textura, el mapa de movimiento y su estabilidad ante cambios de iluminación.....	59
5.7	Resultados obtenidos con el modelo MoG en los mismos escenarios: se observan zonas de sobre-segmentación y artefactos por luminancia.	59
5.8	Mapa de densidad vehicular generado por el modelo MMBS a partir de detecciones acumuladas durante 30 minutos.	61

ÍNDICE DE TABLAS

4-1	Comparación metodológica entre MoG y el modelo morfológico propuesto.....	40
4-2	Tamaño del buffer requerido para distintas resoluciones y profundidades de bits.....	43
5-1	Medidas de desempeño obtenidas con el enfoque MMBS en el conjunto PETS2001.....	47
5-2	Medidas de desempeño obtenidas con el enfoque MoG en el conjunto PETS2001.....	48
5-3	Medidas de desempeño con MMBS en el conjunto CDnet2014.	52
5-4	Medidas de desempeño con MOG mejorado (Trejo-Morales et al., 2023) en el conjunto CDnet2014.	52
5-5	Medidas de error obtenidas con MMBS en el conjunto CDnet2014. ..	53
5-6	Medidas de error obtenidas con MOG mejorado en el conjunto CDnet2014.	54

ACRÓNIMOS

MMBS	Modelo Morfológico de Sustracción de Fondo	15
MoG	Mixture of Gaussians (Mezcla de Gaussianas)	10
IoU	Intersection over Union	14
MAE	Mean Absolute Error (Error Absoluto Medio)	46
RMSE	Root Mean Squared Error (Raíz del Error Cuadrático Medio)	46
PETS	Performance Evaluation of Tracking and Surveillance Dataset	45
CDnet	Change Detection Dataset	45
GPU	Graphics Processing Unit	21
RGB	Red, Green, Blue	68
JSIP	Journal of Signal and Information Processing	15
INDAUTOR	Instituto Nacional del Derecho de Autor	16
UAV	Unmanned Aerial Vehicle	24
PBAS	Pixel-Based Adaptive Segmenter	II
ViBe	Visual Background Extractor	II
CNN	Convolutional Neural Network	20
FgSegNet	Foreground Segmentation Network	20
BSUV-Net	Background Subtraction for Unseen Videos	21
DeepBS	Deep Background Subtraction	21
ConvNet-SOBS	Convolutional Neural Network - Self Organizing Background Subtraction	21
BGSNet-D	Background Subtraction Network - Depth modality	21
MM	Morfología Matemática	27
YCbCr	Luminance-Chrominance Color Space	68
HSV	Hue, Saturation, Value	68

1. INTRODUCCIÓN

En las últimas décadas, el crecimiento exponencial de las ciudades ha planteado nuevos retos en la gestión de la movilidad urbana y la seguridad vial. Para hacer frente a estos desafíos, los sistemas de monitoreo inteligente han adquirido una importancia estratégica, al permitir el análisis automatizado del entorno mediante tecnologías no intrusivas como las cámaras fijas. Estos sistemas, basados en visión por computadora, permiten observar el comportamiento del tráfico vehicular y detectar anomalías en tiempo real, lo que los convierte en herramientas clave para la planeación y operación de ciudades inteligentes (United Nations, Department of Economic and Social Affairs, Population Division, 2022).

Dentro del procesamiento de video, una de las tareas más relevantes para este tipo de sistemas es la sustracción de fondo, técnica que permite separar los objetos en movimiento del resto de la escena, habilitando así la detección, seguimiento y análisis de eventos dinámicos. No obstante, los métodos tradicionales de sustracción de fondo enfrentan dificultades importantes al operar en ambientes reales, especialmente en exteriores. Factores como la variación de la iluminación natural, el ruido generado por los sensores, las condiciones climáticas cambiantes y los reflejos no controlados pueden afectar significativamente el rendimiento de estos algoritmos (Sultana et al., 2020).

En este contexto, el presente trabajo propone el desarrollo de un modelo de sustracción de fondo robusto y eficiente, basado en morfología matemática, una rama del procesamiento de imágenes que permite analizar la forma y estructura de los objetos en una escena. A través de este enfoque, se busca abordar las limitaciones de los métodos actuales y aportar una solución más adaptable a entornos urbanos complejos, manteniendo una baja complejidad computacional que lo haga viable para aplicaciones en tiempo real.

1.1 Descripción del problema

El crecimiento acelerado de las zonas urbanas ha incrementado la necesidad de implementar soluciones tecnológicas que permitan monitorear la movi-

lidad vehicular en tiempo real. Las cámaras de vigilancia fijas, combinadas con algoritmos de análisis de video, se han convertido en una herramienta clave para este propósito. Una de las técnicas más utilizadas para el análisis automatizado de video es la sustracción de fondo, que consiste en distinguir los objetos en movimiento (como vehículos o peatones) del fondo estático en una escena. Esta distinción permite identificar eventos relevantes y habilita funciones de conteo, seguimiento o detección de anomalías.

Sin embargo, aplicar esta técnica en ambientes exteriores plantea una serie de dificultades. Las condiciones de iluminación cambian de forma constante a lo largo del día: las sombras proyectadas, las transiciones entre luz natural y artificial, o la presencia de reflejos afectan directamente la calidad de las imágenes captadas. A esto se suman el ruido introducido por los sensores, las fluctuaciones de temperatura y el impacto de fenómenos climáticos como la lluvia o la niebla. Estos factores comprometen la capacidad de los algoritmos para distinguir con precisión entre fondo y primer plano, generando errores que afectan el desempeño general del sistema (Bouwman, 2014; Maddalena & Petrosino, 2008).

Los modelos estadísticos como el de Mezcla de Gaussianas (Mixture of Gaussians (Mezcla de Gaussianas) (MoG)) (Zivkovic, 2004), aunque ampliamente utilizados, dependen de ajustes constantes y pueden resultar inestables ante escenarios cambiantes. Por otro lado, los enfoques más recientes basados en redes neuronales profundas han demostrado un gran potencial, pero presentan limitaciones prácticas: requieren una gran cantidad de datos para su entrenamiento, consumen muchos recursos computacionales y su rendimiento se degrada si las condiciones del entorno cambian respecto al conjunto de datos original (Sultana et al., 2020).

Esto ha motivado la búsqueda de alternativas más robustas y eficientes, especialmente para sistemas embebidos o de bajo consumo. En este trabajo se plantea el uso de la morfología matemática como base para construir un modelo de sustracción de fondo. A diferencia de los métodos antes mencionados, este enfoque permite operar directamente sobre la geometría y textura de la imagen sin necesidad de entrenamiento previo, lo que lo hace más ligero y adaptable. Herramientas como la dilatación, erosión, apertura y cierre permiten resaltar estructuras relevantes y reducir el ruido, aun en condiciones complejas de ilumina-

ción (Serra, 1982; Soille, 2003).

Por ello, este trabajo se enfoca en el desarrollo de un modelo alternativo que haga uso de operadores morfológicos para realizar la segmentación entre fondo y primer plano en secuencias de video. La propuesta busca superar las limitaciones de los enfoques estadísticos y de aprendizaje profundo, apostando por un método más sencillo de implementar, menos dependiente de condiciones iniciales y capaz de adaptarse a los cambios típicos del entorno urbano. La pregunta clave que se plantea es si la morfología matemática puede ofrecer una solución eficaz y computacionalmente eficiente para la sustracción de fondo en escenarios reales, donde la iluminación, el ruido y la variabilidad del entorno dificultan el desempeño de los modelos convencionales.

1.2 Justificación

La gestión eficiente del tráfico vehicular es uno de los principales retos de las ciudades modernas. La posibilidad de monitorear en tiempo real el comportamiento dinámico de las vías permite tomar decisiones más informadas en materia de movilidad, seguridad y planeación urbana. En este contexto, los sistemas automáticos de vigilancia por video se han consolidado como herramientas clave, al permitir la detección de eventos relevantes mediante algoritmos de análisis de imágenes. Una de las técnicas centrales en estos sistemas es la sustracción de fondo, que permite separar los objetos en movimiento del resto de la escena, habilitando procesos posteriores como el seguimiento, clasificación o conteo de vehículos (Piccardi, 2004).

A pesar de su efectividad en entornos controlados, muchos de los métodos actuales presentan limitaciones al enfrentarse a condiciones reales. Por ejemplo, algoritmos populares como ViBe (Barnich & Van Droogenbroeck, 2011) ofrecen una implementación eficiente, pero son sensibles a fluctuaciones de luminancia, ruido de sensor o escenas altamente dinámicas. En el caso de los métodos basados en aprendizaje profundo, si bien han mostrado resultados sobresalientes en benchmark datasets, requieren un alto volumen de datos para entrenamiento y una infraestructura computacional considerable para operar adecuadamente (Bouwman et al., 2019). Esto dificulta su integración en sistemas embebidos o cámaras con capacidad de procesamiento limitada.

Además, los entornos urbanos introducen desafíos complejos: cambios de iluminación bruscos, sombras proyectadas, lluvia, presencia de reflejos y oclusiones parciales. Todos estos factores degradan la calidad de la imagen y afectan negativamente la precisión de los modelos de segmentación (Cristani et al., 2010; Guo et al., 2016). En consecuencia, existe una necesidad práctica de desarrollar modelos más robustos, que puedan adaptarse a estos escenarios sin depender de reentrenamientos constantes ni del acceso a hardware especializado.

La morfología matemática ofrece una alternativa atractiva en este sentido. Al centrarse en la forma y textura de los objetos, permite manipular la información estructural de la imagen mediante operaciones deterministas como la erosión, dilatación, apertura o cierre. Estas transformaciones hacen posible resaltar características relevantes o suprimir detalles no deseados, sin necesidad de aprendizaje previo ni de parámetros complejos. Esta propiedad la convierte en una herramienta ideal para construir modelos de fondo que funcionen de forma estable en condiciones reales, con bajo costo computacional y alta interpretabilidad.

Esta tesis propone aprovechar dichas propiedades para diseñar un modelo de sustracción de fondo basado en morfología matemática, orientado a aplicaciones de monitoreo urbano. El objetivo no solo es resolver un problema técnico específico, sino también contribuir a la exploración de enfoques alternativos que combinan eficiencia, simplicidad y adaptabilidad, ampliando así el repertorio de herramientas disponibles para la visión por computadora aplicada a entornos urbanos.

1.3 Hipótesis

En el desarrollo de sistemas inteligentes para el monitoreo del tráfico vehicular, la precisión y robustez de los algoritmos de segmentación juegan un papel decisivo en la calidad de los datos obtenidos. Dada la alta variabilidad de los escenarios urbanos, donde factores como la iluminación, el clima, el ruido del sensor y la complejidad de la escena cambian constantemente, se requieren modelos que puedan adaptarse dinámicamente sin depender de entrenamiento previo ni de configuraciones complejas. La morfología matemática, al centrarse en las estructuras geométricas y texturales presentes en las imágenes, ofrece un

marco de trabajo sólido y determinista que puede superar las limitaciones de los enfoques tradicionales basados en estadística o aprendizaje automático.

En este contexto, se plantea la siguiente hipótesis general: Se plantea que un modelo de sustracción de fondo basado en operadores de morfología matemática permite segmentar con mayor precisión los objetos en movimiento en escenas urbanas reales, al tiempo que mantiene una menor complejidad computacional y mayor robustez ante variaciones de iluminación y ruido, en comparación con enfoques tradicionales como el modelo de mezcla de gaussianas.

Esta hipótesis será evaluada mediante la implementación de un algoritmo basado en morfología matemática y su comparación cuantitativa con modelos de referencia ampliamente utilizados en la literatura, utilizando métricas estándar de desempeño sobre conjuntos de datos representativos.

1.4 Objetivos

1.4.1 Objetivo general

Objetivo general: Desarrollar un modelo de sustracción de fondo basado en morfología matemática, orientado a la visión por computadora, que permita mejorar la precisión y eficiencia en la detección de vehículos en movimiento en escenarios urbanos reales con iluminación variable, comparando su desempeño con el modelo de mezcla de gaussianas (MoG).

1.4.2 Objetivos específicos

1. Investigar las técnicas de morfología matemática aplicadas al procesamiento de imágenes y los métodos actuales de sustracción de fondo, con el fin de identificar sus limitaciones en condiciones de iluminación variables.
2. Formular la arquitectura conceptual y operativa del modelo de sustracción de fondo basado en operadores morfológicos, definiendo los elementos y criterios necesarios para abordar las limitaciones detectadas en enfoques tradicionales.
3. Implementar el modelo propuesto, integrando una metodología de evalua-

ción comparativa frente al modelo de mezcla de gaussianas (precision, recall, F1-score e IoU).

4. Evaluar el desempeño del modelo en escenarios urbanos reales con variaciones de iluminación, considerando su robustez, eficiencia computacional y capacidad de adaptación.
5. Documentar los resultados obtenidos en una memoria técnica clara y estructurada que sirva como base para aplicaciones futuras o investigaciones derivadas.

1.5 Contribuciones esperadas

El presente trabajo busca generar diversas contribuciones en el campo de la visión por computadora aplicada al monitoreo urbano, particularmente en la detección de vehículos en movimiento mediante sustracción de fondo. A continuación, se enumeran las principales contribuciones esperadas:

- Propuesta de un modelo de fondo basado en morfología matemática, diseñado específicamente para operar en entornos exteriores con condiciones de iluminación variables, utilizando transformaciones como top-hat, bottom-hat, apertura y cierre sobre la textura local de las imágenes.
- Demostración empírica de la eficacia del modelo, mediante pruebas controladas y experimentales comparando su rendimiento con el método de Mezcla de Gaussianas (MoG), considerado un referente en la literatura. La comparación se realizará con métricas objetivas como precisión, recall, F1-score e Intersection over Union (IoU).
- Validación en escenarios reales, utilizando secuencias de video capturadas con dron sobre entornos urbanos, lo que permitirá evaluar la adaptabilidad del algoritmo a condiciones prácticas y no controladas.
- Documentación detallada y replicable del algoritmo y sus resultados, facilitando su implementación futura por otros investigadores o desarrolladores interesados en soluciones ligeras, explicables y robustas para tareas de detección de movimiento en visión por computadora.

- Aporte teórico a la intersección entre morfología matemática y análisis de video, al explorar cómo las operaciones morfológicas pueden ser utilizadas no solo como herramientas de filtrado, sino como elementos centrales para la modelación adaptativa del fondo en escenarios dinámicos.

Estas contribuciones no solo apuntan a resolver un problema técnico puntual, sino también a abrir nuevas líneas de investigación sobre el uso de enfoques morfológicos en el procesamiento de video, fomentando la exploración de modelos más interpretables, eficientes y adecuados para su aplicación en el mundo real.

1.6 Publicaciones

Como parte del desarrollo de esta investigación, se han generado productos académicos que respaldan la validez y relevancia del trabajo. A continuación, se enlistan las publicaciones derivadas directamente del presente proyecto:

1. Moreno-Mejía, E. J., Herrera-Navarro, A. M., Cantón-Enríquez, D., & Jiménez-Hernández, H. (2025). Morphological Background-Subtraction Modeling for Analyzing Traffic Flow. *Modelling*, 1(1), 1–23. <https://doi.org/10.3390/modelling1010000>
 - Publicación revisada por pares en revista de acceso abierto, donde se presenta el algoritmo Modelo Morfológico de Sustracción de Fondo (MMBS) aplicado a escenarios de tráfico urbano, detallando su diseño, justificación teórica y evaluación sobre conjuntos de datos de referencia.
2. Moreno-Mejía, E. J., Cantón-Enríquez, D., Pérez-Ramos, J. L., Herrera-Navarro, A. M., & Jiménez-Hernández, H. (2025). Morphological Background Model for the Analysis of Traffic Flows in Urban Areas. *Journal of Signal and Information Processing (JSIP)*. En prensa.
 - Versión extendida del trabajo anterior, enfocada en el análisis cualitativo de los resultados obtenidos, la discusión sobre escenarios urbanos complejos, y la comparación contra métodos tradicionales como MoG.
3. Registro de derecho de autor.

- Como medida de protección de la propiedad intelectual, se realizó el registro oficial del algoritmo propuesto ante el Instituto Nacional del Derecho de Autor (INDAUTOR). El registro incluye el código fuente, descripción técnica y fundamentos teóricos del modelo de sustracción de fondo basado en morfología matemática, asegurando así su reconocimiento legal como obra original derivada de esta tesis.

1.7 Estructura de la tesis

Este documento se encuentra organizado en siete capítulos que abordan de forma progresiva los elementos teóricos, metodológicos y experimentales que sustentan el desarrollo del modelo propuesto. A continuación, se describe brevemente el contenido de cada capítulo:

Capítulo 1 Introducción: Presenta el contexto general del problema, la justificación de la investigación, la hipótesis de trabajo, los objetivos planteados, las contribuciones esperadas, las publicaciones derivadas y la organización general de la tesis.

Capítulo 2 Antecedentes: Ofrece una revisión de la literatura relacionada con los sistemas de visión por computadora aplicados al análisis de tráfico, métodos tradicionales de sustracción de fondo, enfoques basados en aprendizaje profundo y trabajos previos que emplean morfología matemática en procesamiento de imágenes.

Capítulo 3 Marco teórico: Expone los fundamentos conceptuales y matemáticos de la morfología matemática, detallando los operadores principales, sus propiedades, y su aplicación en imágenes digitales. También se revisan conceptos clave relacionados con modelos de fondo y análisis de video.

Capítulo 4 Propuesta de investigación: Describe en detalle el modelo de sustracción de fondo desarrollado, su diseño algorítmico, el razonamiento detrás de cada etapa del procesamiento, y los principios morfológicos aplicados. Se incluyen diagramas explicativos y consideraciones sobre la implementación.

Capítulo 5 Resultados preliminares: Presenta las pruebas realizadas con el modelo propuesto, incluyendo su validación en distintos escenarios urbanos mediante secuencias de video capturadas con dron. Se discuten e interpretan los resultados obtenidos, comparándolos con los de un modelo de referencia (MoG) mediante métricas estándar, y se analizan las fortalezas, limitaciones y posibles aplicaciones prácticas del enfoque morfológico.

Capítulo 6 Conclusiones y trabajos futuros: Resume los hallazgos principales de la investigación, responde a la hipótesis planteada y presenta propuestas para continuar explorando el uso de morfología matemática en aplicaciones de visión por computadora.

2. ANTECEDENTES

2.1 Visión por computadora en entornos urbanos

La visión por computadora ha evolucionado como una de las herramientas más relevantes en la automatización de tareas de análisis visual, particularmente en escenarios urbanos, donde su aplicación en sistemas de transporte inteligente ha permitido avances significativos en seguridad, eficiencia y gestión de movilidad. A través del uso de cámaras fijas, móviles o montadas en vehículos aéreos no tripulados (drones), es posible capturar flujos de video que posteriormente son procesados para extraer información clave como la detección de vehículos, clasificación de tipos, seguimiento de trayectorias y análisis de congestión (Chen et al., 2021; Zhou & Wang, 2020).

En áreas metropolitanas, donde los patrones de tráfico cambian dinámicamente y las condiciones de iluminación pueden ser impredecibles, el uso de visión por computadora permite compensar la falta de sensores físicos instalados permanentemente, ofreciendo un enfoque flexible y escalable. Además, permite detectar incidentes en tiempo real, optimizar los semáforos y reducir la intervención humana en tareas rutinarias de vigilancia (Huang et al., 2022). Estos sistemas se han convertido en parte central de las llamadas ciudades inteligentes, donde la digitalización de la infraestructura urbana está transformando la forma en que se monitorean y gestionan los servicios públicos.

Más allá del simple conteo de vehículos, la visión computacional también se emplea para medir velocidades, predecir flujos y analizar el comportamiento de los usuarios en intersecciones, cruces peatonales y glorietas. La combinación de visión por computadora con técnicas de inteligencia artificial permite construir modelos predictivos que ayudan a tomar decisiones sobre planeación urbana, diseño vial o control automatizado de tráfico (Z. Wang et al., 2023).

La creciente disponibilidad de cámaras económicas, sensores embebidos y plataformas de código abierto ha democratizado estas tecnologías, facilitando su adopción incluso en contextos con recursos limitados. Sin embargo, estos sistemas aún enfrentan retos importantes relacionados con condiciones ambientales adversas, escenas ruidosas o no estructuradas y la necesidad de

segmentar objetos en movimiento con precisión en entornos complejos, motivando así la búsqueda de enfoques más robustos y explicables.

2.2 Modelos clásicos de sustracción de fondo

La sustracción de fondo es una técnica fundamental en el análisis de secuencias de video, especialmente para detectar objetos en movimiento en entornos estáticos. Su propósito es separar el primer plano (donde ocurren cambios relevantes) del fondo que se asume como estático o recurrente. Esta tarea es la base para etapas posteriores como la clasificación, el seguimiento o la detección de eventos anómalos en visión por computadora.

2.2.1 Mixture of Gaussians (MoG)

Uno de los modelos más influyentes en esta área es el propuesto por Stauffer y Grimson en 1999, conocido como Mixture of Gaussians (MoG). Este método modela cada píxel como una combinación de varias distribuciones gaussianas, que representan los distintos valores que ese píxel puede tomar en el tiempo. Las distribuciones con mayor peso se asumen como fondo, mientras que las menos frecuentes se consideran parte del primer plano (Stauffer & Grimson, 1999).

El enfoque MoG ha demostrado una buena capacidad de adaptación a escenas con variaciones graduales de iluminación o movimiento leve en el fondo, como el vaivén de ramas o reflejos. No obstante, presenta limitaciones importantes frente a cambios abruptos, escenas con alto dinamismo, o la aparición repentina de objetos nuevos. Además, el modelo requiere parámetros sensibles al contexto, como el número de gaussianas, la tasa de aprendizaje y los umbrales de clasificación (Elgammal et al., 2000).

2.2.2 ViBe: Visual Background Extractor

En respuesta a las limitaciones de MoG, surgieron modelos alternativos como ViBe, desarrollado por Barnich y Van Droogenbroeck en 2011. Este enfoque no utiliza modelos probabilísticos, sino una memoria de valores recientes

por píxel que se actualiza parcialmente en cada cuadro. La clasificación se realiza comparando el valor actual con un subconjunto aleatorio de la memoria (Barnich & Van Droogenbroeck, 2011).

ViBe es eficiente computacionalmente, robusto frente a ruido temporal y no requiere un periodo de entrenamiento inicial prolongado. Sin embargo, puede ser sensible a pequeños movimientos persistentes del fondo (como sombras cambiantes) y puede degradarse con el tiempo si el modelo se adapta a objetos que deberían pertenecer al primer plano.

2.2.3 PBAS: Pixel-Based Adaptive Segmenter

Otra propuesta destacada es PBAS, que introduce mecanismos de adaptación local tanto en los umbrales de decisión como en la tasa de actualización de cada píxel. A diferencia de MoG y ViBe, este enfoque permite que cada región de la imagen se ajuste de forma independiente, lo que lo hace más flexible ante escenas heterogéneas o dinámicas (Hofmann et al., 2012).

PBAS es especialmente útil en contextos donde el fondo presenta comportamientos no uniformes, como zonas urbanas con tráfico irregular, sombras proyectadas o vibraciones de cámara. A pesar de ello, su implementación puede volverse compleja y presenta dificultades en la detección precisa de contornos.

2.3 Enfoques basados en aprendizaje profundo para segmentación de fondo

El desarrollo de técnicas de aprendizaje profundo ha transformado el campo de la visión por computadora, abriendo nuevas posibilidades para tareas como clasificación, detección y segmentación de objetos en imágenes y video. En el contexto de la sustracción de fondo, los modelos basados en redes neuronales convolucionales (Convolutional Neural Network (CNN)) han mostrado un rendimiento sobresaliente en benchmarks públicos, especialmente en condiciones bien estructuradas o controladas (Bouwman et al., 2019).

Uno de los modelos más influyentes es Foreground Segmentation Network (FgSegNet), que emplea una arquitectura encoder-decoder con mecanismos de atención para refinar los bordes del primer plano. Este modelo ha lidera-

do múltiples competiciones en el conjunto de datos CDnet 2014, superando en precisión a métodos tradicionales incluso en escenarios desafiantes como escenas nocturnas o lluviosas (Lim et al., 2018). Su desventaja radica en el alto costo computacional y la necesidad de entrenamiento con imágenes etiquetadas cuadro a cuadro.

Otro modelo relevante es Background Subtraction for Unseen Videos (BSUV-Net), que integra aprendizaje profundo supervisado con técnicas de propagación de fondo y fusión multicanal, permitiéndole adaptarse parcialmente a cambios graduales en escenas dinámicas (Tezcan et al., 2020). Sin embargo, como la mayoría de modelos basados en deep learning, requiere una etapa de preprocesamiento intensiva, así como acceso a Graphics Processing Unit (GPU)s para una inferencia en tiempo real.

Modelos como Deep Background Subtraction (DeepBS), Convolutional Neural Network - Self Organizing Background Subtraction (ConvNet-SOBS) o Background Subtraction Network - Depth modality (BGSNet-D) siguen una estrategia similar, entrenando redes profundas sobre secuencias representativas de distintos entornos. A pesar de sus buenos resultados cuantitativos en métricas como F1-score e IoU, presentan problemas de generalización cuando se enfrentan a entornos significativamente distintos de los que fueron utilizados durante el entrenamiento (Javed et al., 2021). Además, en la mayoría de los casos, el funcionamiento interno de estos modelos carece de explicabilidad, lo que dificulta su validación o ajuste en aplicaciones sensibles.

2.4 Aplicaciones previas de la morfología matemática en visión por computadora

La morfología matemática ha sido una de las herramientas más influyentes en el procesamiento de imágenes desde su introducción formal en los años setenta, cuando Jean Serra y Georges Matheron establecieron su marco teórico dentro del contexto de la teoría de conjuntos y la geometría integral (Serra, 1982). A diferencia de los enfoques basados en la intensidad o el color de los píxeles, la morfología se centra en la estructura espacial de los objetos y sus relaciones geométricas, lo que la hace particularmente útil para tareas donde la forma es un atributo discriminante.

En sus aplicaciones tempranas, la morfología fue empleada en binarización, filtrado estructural y análisis de contornos, logrando una gran aceptación en áreas como la inspección industrial, la segmentación de células en imágenes biomédicas, y la detección de texto en imágenes digitales (Soille, 2003; Sternberg, 1986). Con el paso del tiempo, sus operadores básicos (erosión, dilatación, apertura y cierre) se extendieron al dominio de los niveles de gris, permitiendo un análisis más flexible de escenas con variabilidad tonal.

En el campo específico de la visión por computadora, la morfología ha demostrado ser eficaz en la detección de movimiento, extracción de regiones de interés, separación de objetos superpuestos y limpieza de ruido estructurado. Por ejemplo, Haralick et al. (1987) propusieron el uso de operadores morfológicos para resaltar la forma de objetos en escenas con baja relación señal-ruido, demostrando que estas técnicas eran más robustas que los filtros lineales en condiciones adversas (Haralick et al., 1987).

Más recientemente, se han desarrollado algoritmos que combinan morfología con modelos de fondo adaptativos. Kaur y Kaur (2023) aplicaron operadores morfológicos en conjunto con técnicas estadísticas para mejorar la detección de peatones en entornos urbanos, logrando reducir las tasas de falsos positivos en videos con iluminación cambiante (Kaur & Kaur, 2023). Del mismo modo, Wang et al. (2022) utilizaron transformaciones top-hat y bot-hat para resaltar texturas en superficies viales, facilitando la identificación de vehículos y anomalías estructurales desde una vista aérea (H. Wang et al., 2022).

El uso de la morfología en video también ha sido explorado en el contexto del modelado del fondo. En estos casos, los operadores permiten refinar las máscaras binarias obtenidas por otros métodos, rellenar huecos en el primer plano y eliminar ruido persistente del fondo. Este enfoque ha demostrado ser especialmente útil en condiciones de iluminación variable, donde los modelos estadísticos o neuronales tienden a fallar (Said & Jambek, 2021).

2.5 Algoritmos morfológicos en sustracción de fondo

El uso de morfología matemática en tareas de sustracción de fondo ha ganado interés en la última década como una alternativa a los modelos probabilísticos y a los enfoques basados en aprendizaje profundo. Su atractivo radica

en su simplicidad computacional, independencia de entrenamiento previo y capacidad para operar de forma determinista sobre la estructura espacial de la imagen. Estas características resultan particularmente valiosas en escenarios donde la variabilidad de las condiciones de iluminación o el ruido estructural dificultan la aplicación de modelos complejos.

Uno de los enfoques más comunes consiste en utilizar operadores morfológicos como etapas de preprocesamiento o postprocesamiento para mejorar los resultados de segmentación generados por algoritmos convencionales. Por ejemplo, Said y Jambek (2021) propusieron un esquema adaptativo basado en erosión, dilatación y transformaciones top-hat para refinar las máscaras obtenidas por sustracción de fondo, logrando reducir falsos positivos en escenas con sombras y reflejos (Said & Jambek, 2021).

Por otro lado, Mello Román et al. (2019) desarrollaron un algoritmo enteramente morfológico para identificar regiones dinámicas en secuencias de video. Su propuesta se basa en calcular una mediana estructural temporal para cada píxel, refinada posteriormente mediante aperturas y cierres secuenciales. Este modelo mostró mejoras en estabilidad frente a cambios de iluminación, aunque requiere ajustes de parámetros para adaptarse a distintos entornos (Mello Román et al., 2019).

En un enfoque más estructural, algunos autores han empleado la interpretación topográfica de las imágenes, considerando los niveles de intensidad como una superficie tridimensional sobre la cual se identifican protuberancias o depresiones que representan movimiento. Esta representación permite aplicar operadores como top-hat para detectar regiones que sobresalen con respecto al entorno local, resaltando así el primer plano con mayor precisión (Sternberg, 1986; H. Wang et al., 2022).

Aunque estos trabajos demuestran el potencial de la morfología matemática como base para algoritmos de sustracción de fondo, aún existen limitaciones prácticas relacionadas con la adaptabilidad, el control automático de parámetros y la validación en contextos reales complejos. Esto sugiere la necesidad de continuar explorando enfoques morfológicos más robustos, versátiles y eficientemente implementables en escenarios urbanos reales.

2.6 Uso de drones en monitoreo urbano

El uso de vehículos aéreos no tripulados (Unmanned Aerial Vehicle (UAV)s), comúnmente conocidos como drones, ha transformado la forma en que se lleva a cabo la observación del entorno urbano. Gracias a su capacidad de sobrevolar áreas extensas a baja altitud y con gran maniobrabilidad, los drones ofrecen una plataforma flexible y económica para la recolección de datos visuales en tiempo real. En el contexto del análisis de tráfico, su capacidad para capturar secuencias de video cenitales (es decir, desde una vista perpendicular al suelo) elimina muchos de los problemas asociados con las cámaras fijas, como la perspectiva angular, las oclusiones parciales o la dependencia del montaje físico en postes o edificios (Cavallaro et al., 2020).

Esta perspectiva cenital resulta especialmente útil en intersecciones, glorietas y avenidas de alta complejidad, donde la visualización simultánea de múltiples carriles, flujos y trayectorias sería difícil de lograr desde el suelo. Además, los drones permiten adaptar dinámicamente el área de monitoreo, ajustar su altitud o desplazarse hacia zonas de interés ante eventos inesperados, lo cual no es posible con las cámaras estáticas tradicionales (Rajendran et al., 2021).

El uso de drones ha sido ampliamente explorado en estudios recientes orientados al conteo vehicular, la detección de incidentes de tráfico, la evaluación de comportamiento peatonal y la optimización de semáforos inteligentes. Por ejemplo, Ciuffreda et al. (2019) utilizaron UAVs para recolectar datos de flujos vehiculares en escenarios urbanos complejos, logrando segmentar los movimientos individuales de vehículos y analizar patrones de congestión sin intervención física en la infraestructura (Ciuffreda et al., 2019).

En el ámbito técnico, los videos capturados por drones presentan ciertos retos únicos, como la estabilización de imagen, la variabilidad en la resolución, el impacto del movimiento de cámara y las condiciones atmosféricas. Sin embargo, estos desafíos pueden ser mitigados mediante procesamiento digital, estabilización por software y uso de cámaras de alta definición. Por su parte, los algoritmos de visión por computadora deben ser lo suficientemente robustos para operar sobre este tipo de secuencias, lo que ha impulsado el desarrollo de modelos adaptativos capaces de funcionar en condiciones cambiantes (Mohammed et al., 2022).

En conjunto, estas características hacen del dron una herramienta ideal para el monitoreo urbano automatizado. Su incorporación en el presente trabajo no solo responde a razones prácticas, sino también metodológicas: permite validar el modelo propuesto en entornos reales, bajo condiciones no controladas, y con un punto de vista cenital que favorece la segmentación espacial de objetos móviles. De este modo, se busca demostrar que los algoritmos basados en morfología matemática pueden integrarse eficazmente en sistemas de vigilancia aérea urbana, aportando precisión y eficiencia en el análisis del tráfico.

3. MARCO TEÓRICO

3.1 Fundamentos de la morfología matemática

La morfología matemática es una teoría formal desarrollada inicialmente en el contexto de la geología y la microscopía por Georges Matheron y Jean Serra en el Instituto de Minería de París a finales de los años sesenta. Su objetivo original era proporcionar una herramienta cuantitativa para describir y analizar estructuras geométricas presentes en imágenes minerales. Posteriormente, fue extendida al procesamiento de imágenes digitales por Jean Serra en su obra seminal *Image Analysis and Mathematical Morphology* (Serra, 1982), lo que marcó el inicio de su aplicación en visión por computadora.

En esencia, la morfología matemática estudia la forma y estructura de los objetos dentro de una imagen a través de operaciones que se basan en la teoría de conjuntos, el álgebra booleana y la topología. A diferencia de los métodos estadísticos o espectrales, la morfología no se enfoca en el valor absoluto de los píxeles, sino en su disposición espacial y relación con su vecindario. Esta propiedad la convierte en una técnica especialmente robusta para el tratamiento de imágenes con ruido estructurado, variabilidad local de iluminación o presencia de texturas complejas (Gonzalez & Woods, 2009; Sternberg, 1986).

La morfología clásica se originó en el análisis de imágenes binarias, donde los objetos se representaban como subconjuntos de un espacio euclidiano discreto, comúnmente \mathbb{Z}^2 . En este marco, las operaciones morfológicas se definían mediante la interacción entre una imagen binaria I y un elemento estructurante λ , el cual actúa como una sonda o plantilla que se desplaza sobre la imagen para inspeccionar su geometría local. Esta formulación fue posteriormente extendida al dominio de escala de grises, permitiendo representar $I(x)$ como una función de intensidad definida sobre una rejilla topológica, y operar directamente sobre las estructuras locales de la superficie inducida por la imagen.

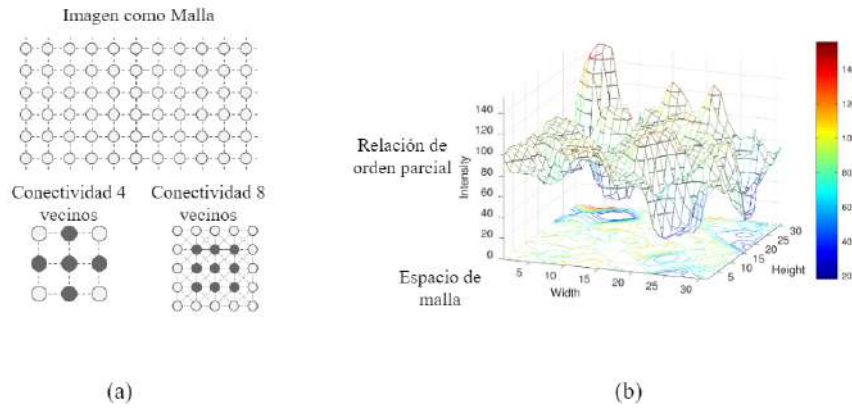


Figura 3.1: Representación de una imagen como malla discreta con conectividad vecinal (a) y como superficie tridimensional en un espacio de intensidad (b). Esta representación permite aplicar operadores morfológicos que trabajan sobre la textura de la imagen mediante relaciones de orden parcial. Fuente: Elaboración propia.

Estas ideas fueron extendidas a imágenes en niveles de gris por Sternberg (1986) y Haralick et al. (1987), quienes demostraron que los operadores morfológicos pueden definirse como transformaciones punto a punto sobre funciones de intensidad, manteniendo la coherencia estructural del modelo original (Haralick et al., 1987). Este enfoque ha permitido aplicar morfología matemática a un amplio rango de tareas, incluyendo filtrado, segmentación, extracción de contornos, realce de texturas y modelado de fondo.

3.1.1 Operadores morfológicos básicos

La morfología matemática (Morfología Matemática (MM)) proporciona un marco teórico sólido para el análisis estructural de imágenes, especialmente útil en representaciones espaciales discretas como las adquiridas por sensores fotográficos. Su estructura se basa en la manipulación de superficies formadas por la multiplicación euclidiana de conjuntos discretos, dando lugar a rejillas topológicas donde operan transformaciones morfológicas fundamentales.

Los operadores básicos de la morfología son la erosión y la dilatación, definidos como:

$$\xi_{\lambda} I(\mathbf{x}) = \min_{\mathbf{b} \in D_{\lambda}} \{I(\mathbf{x} + \mathbf{b}) - \lambda(\mathbf{b})\} \quad (3.1)$$

$$\delta_{\lambda}I(\mathbf{x}) = \max_{\mathbf{b} \in B_{\lambda}} \{I(\mathbf{x} + \mathbf{b}) - \lambda(\mathbf{b})\} \quad (3.2)$$

Donde:

- $I(\mathbf{x})$ representa la superficie o imagen en escala de grises.
- λ denota el elemento estructurante.
- D_{λ} y B_{λ} son los dominios definidos por λ desplazados a la posición de referencia \mathbf{x} .
- \mathbf{b} es el vector que recorre los puntos del elemento estructurante.

Estas transformaciones se derivan de la teoría de conjuntos propuesta por Serra y extendidas al dominio en escala de grises por Sternberg (Soille, 2003; Sternberg, 1986). Son ampliamente utilizadas en tareas de segmentación y análisis estructural en visión por computadora (Klette, 2014).

A partir de estos operadores fundamentales se definen los filtros morfológicos apertura y cierre:

$$\gamma_{\lambda}(I) = \delta_{\lambda}(\xi_{\lambda}(I)) \quad (3.3)$$

$$\phi_{\lambda}(I) = \xi_{\lambda}(\delta_{\lambda}(I)) \quad (3.4)$$

Estos filtros se consideran operadores porque cumplen dos propiedades esenciales:

1. Monotonicidad: si $I \leq J \Rightarrow T(I) \leq T(J)$, donde $T \in \{\gamma_{\lambda}, \phi_{\lambda}\}$.
2. Idempotencia: $T(T(I)) = T(I)$, es decir, aplicar el operador dos veces es equivalente a aplicarlo una sola vez.

Estas propiedades los convierten en herramientas clave para la reducción de ruido, suavizado de bordes y preservación de estructuras geométricas.

Finalmente, a partir de los filtros morfológicos básicos, se definen dos transformaciones derivadas que capturan la información residual (textura):

$$Th_{\lambda}(I) = I - \gamma_{\lambda}(I) \quad (3.5)$$

$$Bh_{\lambda}(I) = \phi_{\lambda}(I) - I \quad (3.6)$$

Donde:

- $Th_{\lambda}(I)$: extrae estructuras brillantes pequeñas sobre un fondo oscuro.
- $Bh_{\lambda}(I)$: extrae estructuras oscuras pequeñas sobre un fondo brillante.

Estas transformaciones permiten aproximar una superficie por debajo (apertura) o por encima (cierre), y evaluar su diferencia respecto a los datos originales I . El residuo extraído representa la textura local en forma de diferencias entre la superficie observada y su aproximación morfológica.

3.1.2 Transformaciones morfológicas derivadas

Las transformaciones derivadas de la morfología matemática permiten capturar detalles locales mediante la diferencia entre una superficie original y su aproximación estructural. Estas aproximaciones se realizan aplicando filtros morfológicos como la apertura γ_{λ} y el cierre ϕ_{λ} sobre una imagen o superficie I , utilizando un elemento estructurante λ .

A partir de estas transformaciones se definen dos operadores residuales: el top-hat morfológico y el bot-hat, los cuales cuantifican los detalles perdidos durante la aproximación estructural inferior o superior, respectivamente:

$$Th_{\lambda}(I) = I - \gamma_{\lambda}(I) \quad (3.7)$$

$$Bh_{\lambda}(I) = \phi_{\lambda}(I) - I \quad (3.8)$$

Estas transformaciones devuelven el residuo morfológico tras aplicar los filtros. El residuo representa los detalles finos de la superficie que no se ajustan a la estructura definida por λ . Esto permite identificar contrastes locales, texturas, y cambios abruptos en la intensidad que corresponden a pequeñas estructuras o imperfecciones relevantes en la escena.

A partir de estas transformaciones se definen los residuos morfológicos negativo y positivo como:

$$\mathfrak{R}_{n,\lambda}(I) = Th_{\lambda}(I) \quad (3.9)$$

$$\mathfrak{R}_{p,\lambda}(I) = Bh_{\lambda}(I) \quad (3.10)$$

Donde:

- $\mathfrak{R}_{n,\lambda}(I)$ representa la diferencia negativa entre la superficie original y su aproximación por apertura (estructura brillante sobre fondo oscuro).
- $\mathfrak{R}_{p,\lambda}(I)$ representa la diferencia positiva entre la superficie cerrada y la original (estructura oscura sobre fondo brillante).

Cuando ambos residuos son nulos, es decir, $\mathfrak{R}_{n,\lambda}(\mathbf{x}) = \inf$ y $\mathfrak{R}_{p,\lambda}(\mathbf{x}) = \inf$, se considera que la aproximación morfológica es exacta en el punto \mathbf{x} . Esto caracteriza regiones planas sin textura, útiles para identificar zonas de fondo estructuralmente homogéneas.

Finalmente, se puede construir un mapa de textura a partir de ambos residuos mediante la siguiente expresión:

$$\tau_{\lambda}(I) = \text{mediana} \{ \varsigma - \mathfrak{R}_{n,\lambda}, \varsigma + \mathfrak{R}_{p,\lambda} \} \quad (3.11)$$

Donde:

- $\tau_{\lambda}(I)$ representa el mapa de información de textura.
- ς es el valor central del reticulado inducido por los operadores morfológicos, definido entre $\inf I$ y $\sup I$.

- λ es el elemento estructurante usado para definir la escala de análisis.

Este mapa de textura resalta las zonas donde existe información estructural significativa en la imagen y es especialmente útil en escenarios donde se requiere diferenciar entre objetos en movimiento y fondo estático.

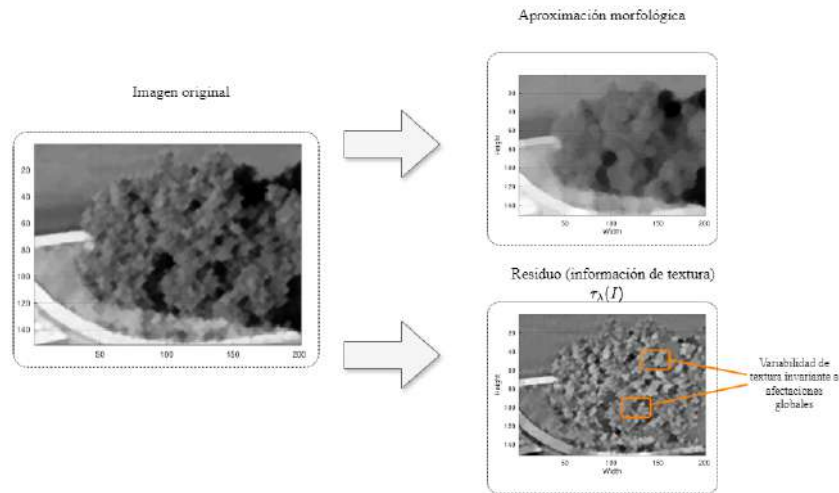


Figura 3.2: Caracterización morfológica de textura: (arriba) superficie aproximada mediante operadores morfológicos; (abajo) residuo que representa la información de textura. Fuente: Elaboración propia.

3.1.3 Modelo probabilístico basado en textura

La información capturada por un sensor fotográfico representa cambios de intensidad a lo largo del tiempo. La secuencia de imágenes procesadas para análisis morfológico se denota como $\Gamma = \{I_1, I_2, \dots, I_t\}$, y su correspondiente secuencia de texturas derivada mediante el mapa $\tau_{\lambda}(\cdot)$ se representa como:

$$\Gamma_{\lambda}^* = \{\tau_{\lambda}(I_1), \tau_{\lambda}(I_2), \dots, \tau_{\lambda}(I_t)\} \quad (3.12)$$

Para denotar la textura localizada en una posición \mathbf{x} en el tiempo t , se utiliza la notación $\Gamma_{t,\lambda}^*(\mathbf{x})$.

El objetivo de esta sección es utilizar esta información para determinar cuándo un objeto comienza a moverse en la escena. Para ello, se propone un modelo probabilístico que clasifica la textura en movimiento utilizando una función de densidad de probabilidad ajustada a la secuencia temporal de cada píxel.

Cada posición \mathbf{x} se modela como el centro de una vecindad local, que considera tanto la información espacial (intensidades vecinas) como la evolución temporal (número de muestras). A partir de esto, se define un clasificador probabilístico como sigue:

$$C_{\eta,\lambda}(\mathbf{x}) = \begin{cases} 1 & \mathbf{x} \in \Gamma_{\eta,\lambda}^*(\mathbf{x}) \sim F(\mathbf{x}; \mathbf{p}) \\ 0 & \text{en otro caso} \end{cases} \quad (3.13)$$

Donde:

- $C_{\eta,\lambda}(\mathbf{x})$ es la función de decisión binaria para el píxel \mathbf{x} .
- η denota la ventana temporal utilizada.
- λ representa el elemento estructurante empleado para caracterizar la textura.
- $\Gamma_{\eta,\lambda}^*(\mathbf{x})$ representa la secuencia de textura en el vecindario de \mathbf{x} durante el intervalo temporal η .
- $F(\mathbf{x}; \mathbf{p})$ denota la función de densidad de probabilidad ajustada con parámetros \mathbf{p} .

La interpretación del modelo indica que una posición \mathbf{x} será clasificada como primer plano (1) si la textura observada se desvía significativamente del comportamiento esperado del fondo. Esta desviación puede deberse al movimiento de un objeto, cambios en la iluminación o aparición de nuevas estructuras.

El modelo captura implícitamente la localidad espacial al considerar la textura del vecindario y su estabilidad en el tiempo. En la práctica, los datos de fondo se caracterizan por secuencias temporales estables con baja varianza, mientras que el primer plano presenta texturas con alta variabilidad.

El resultado del clasificador genera una máscara binaria que indica la presencia de movimiento. Esta máscara puede contener ruido impulsivo o zonas sobredetectadas, las cuales se corrigen mediante operadores morfológicos posteriores como filtros de apertura o cierre, dependiendo del tipo de error observado.

4. PROPUESTA DE INVESTIGACIÓN

4.1 Descripción general del modelo propuesto

El modelo propuesto tiene como objetivo detectar objetos en movimiento (principalmente vehículos) en secuencias de video capturadas mediante vista cenital por un dron. A diferencia de los enfoques convencionales basados en modelos estadísticos o aprendizaje profundo, esta propuesta se fundamenta exclusivamente en operadores de morfología matemática para la construcción y actualización del fondo, así como para la segmentación del primer plano.

La propuesta parte de la hipótesis de que la textura local, obtenida mediante transformaciones morfológicas derivadas, puede ser utilizada para modelar de manera robusta el comportamiento espacial y temporal de cada píxel, incluso bajo condiciones de iluminación variable. A través de la aplicación sistemática de aperturas, cierres y transformaciones tipo top-hat y bot-hat, el modelo genera un mapa de textura por cada imagen en la secuencia, el cual captura los detalles estructurales del entorno.

Posteriormente, se construye un modelo de fondo utilizando un esquema no paramétrico basado en la moda de textura por píxel, dentro de una ventana temporal deslizante. A cada nuevo cuadro, se evalúa la diferencia entre el mapa de textura actual y el modelo de fondo, generando un conjunto de residuos. Estos residuos son evaluados por un clasificador probabilístico que etiqueta cada píxel como fondo o primer plano según su estabilidad temporal y distribución espacial.

El modelo se complementa con un esquema de postprocesamiento morfológico para eliminar ruido impulsivo y mejorar la coherencia espacial de las regiones detectadas. Todo el proceso se ejecuta de manera completamente determinista y sin requerir entrenamiento supervisado, lo que facilita su portabilidad, interpretación y uso en plataformas de procesamiento embarcadas.

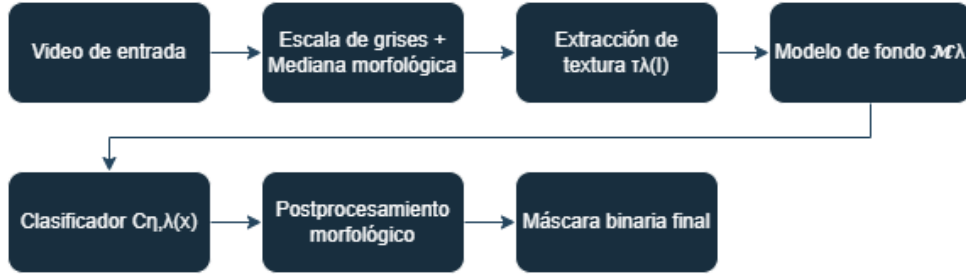


Figura 4.1: Diagrama general del flujo de procesamiento del modelo propuesto.

El diagrama de la Figura 4.1 resume las etapas principales del algoritmo: preprocesamiento, extracción de textura, modelado de fondo, clasificación morfológica y postprocesamiento. Cada una de estas etapas se detalla en las subsecciones siguientes.

4.2 Preprocesamiento de la imagen

La primera etapa del modelo consiste en preparar cada imagen de entrada para facilitar el análisis estructural. Este proceso tiene como objetivo reducir el ruido impulsivo y estandarizar la representación espacial antes de aplicar las operaciones morfológicas.

Dado un cuadro I_t de la secuencia de video capturada desde vista cenital, se realiza la conversión a escala de grises para simplificar el análisis, representando la imagen como una función de intensidad sobre una malla discreta:

$$I_t : \Omega \subset \mathbb{Z}^2 \rightarrow [0, 255]$$

Donde $I_t(\mathbf{x})$ denota el valor de intensidad en la posición \mathbf{x} para el tiempo t .

Posteriormente, se aplica un filtro morfológico tipo mediana adaptado, diseñado para eliminar ruido de alta frecuencia sin afectar las estructuras topográficas relevantes. Este filtro utiliza como base la aproximación superior e inferior de la imagen mediante los operadores de apertura γ_λ y cierre ϕ_λ :

$$\mathcal{F}_\lambda(I_t) = \text{median} \{ \gamma_\lambda(I_t), \phi_\lambda(I_t) \} \quad (4.1)$$

Esta operación actúa como un filtro de paso medio que suaviza variaciones menores que el tamaño del elemento estructurante λ , manteniendo los bordes y formas significativas. El resultado es una imagen suavizada I_t^{filt} libre de ruido estructural, lista para la caracterización morfológica de textura.

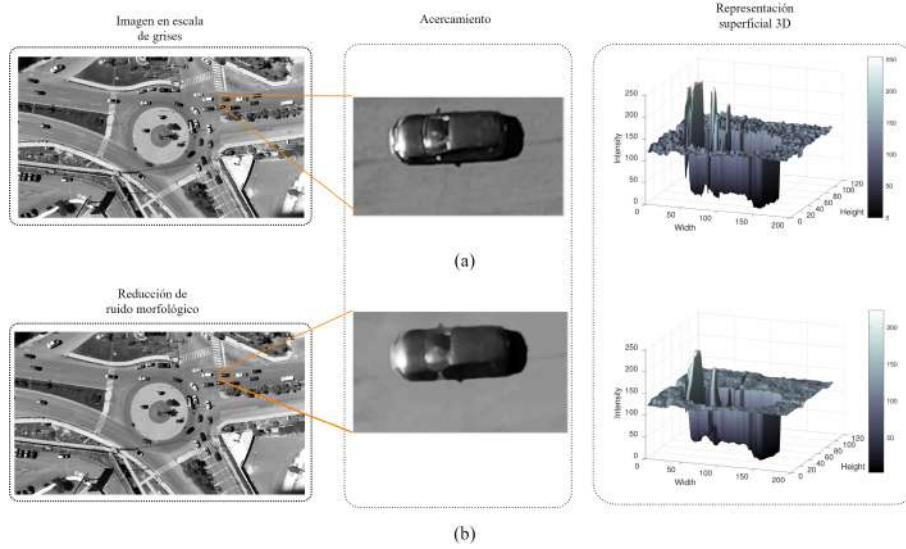


Figura 4.2: Ejemplo de reducción de ruido estructural mediante filtro de mediana morfológica. (a) Imagen original con ruido; (b) imagen procesada con filtro basado en apertura y cierre morfológico. Fuente: Elaboración propia.

Este preprocesamiento garantiza una representación robusta del entorno antes de proceder al cálculo de residuos morfológicos y mapas de textura en las etapas posteriores.

4.3 Construcción del modelo de fondo

Una vez que se ha obtenido la imagen filtrada I_t^{filt} y su correspondiente mapa de textura $\tau_\lambda(I_t)$, se procede a construir el modelo de fondo basado en la evolución temporal de la textura en cada posición \mathbf{x} del dominio espacial Ω .

Este modelo no requiere entrenamiento previo ni suposiciones probabilísticas explícitas; en su lugar, se construye de manera no paramétrica mediante una ventana temporal η que almacena las texturas observadas en los últimos N cuadros:

$$\Gamma_\lambda^*(\mathbf{x}) = \{\tau_\lambda(I_{t-N+1})(\mathbf{x}), \dots, \tau_\lambda(I_t)(\mathbf{x})\}$$

El modelo de fondo $\mathcal{M}_\lambda(\mathbf{x})$ para la posición \mathbf{x} se define como la moda de los valores observados en dicha ventana:

$$\mathcal{M}_\lambda(\mathbf{x}) = \text{moda}(\Gamma_\lambda^*(\mathbf{x})) \quad (4.2)$$

En caso de que existan múltiples valores con la misma frecuencia, se elige el valor más reciente observado como criterio de desempate.

Este enfoque ofrece dos ventajas clave:

1. Robustez estructural: al modelar la textura (y no la intensidad), el modelo es menos sensible a variaciones de iluminación, sombras o ruido global.
2. Estabilidad temporal: la moda captura la recurrencia dominante en la secuencia local, lo que permite distinguir entre el fondo estable y los objetos móviles que sólo aparecen en un número reducido de cuadros.

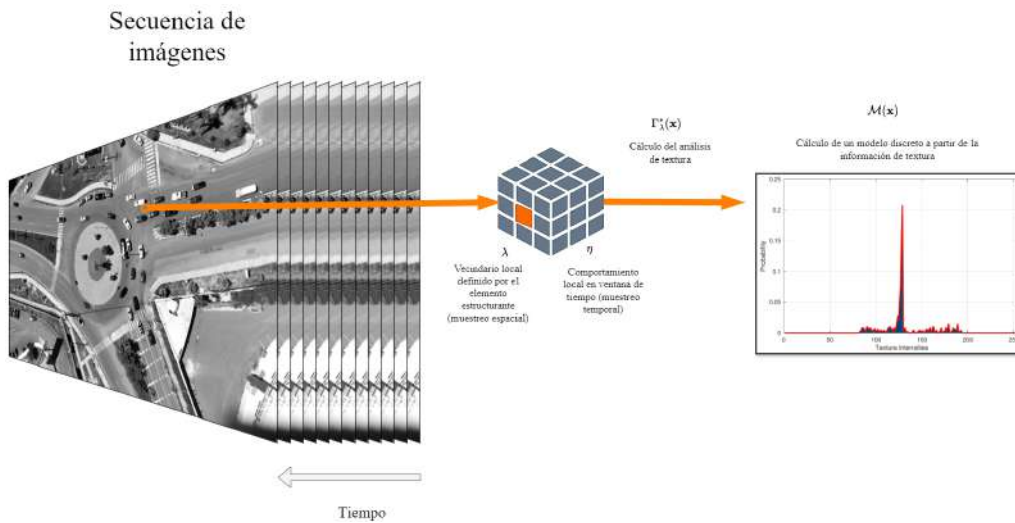


Figura 4.3: Esquema del modelo probabilístico basado en textura propuesto. Se analiza la evolución temporal y espacial de la textura en cada píxel, permitiendo identificar regiones en movimiento con mayor robustez. Fuente: Elaboración propia.

La Figura 4.3 ilustra el proceso de acumulación temporal y modelado del fondo a partir de los mapas de textura. Este modelo es actualizado dinámicamente para cada nueva imagen de entrada, manteniendo la ventana de análisis deslizante centrada en el cuadro actual.

4.4 Detección del primer plano

Una vez construido el modelo de fondo $\mathcal{M}_\lambda(\mathbf{x})$ mediante moda morfológica de textura, se procede a detectar los píxeles correspondientes al primer plano. Esta tarea se realiza evaluando el residuo morfológico entre el mapa de textura actual $\tau_\lambda(I_t)$ y el fondo estimado en cada posición.

Primero se calculan los residuos negativos y positivos mediante las transformaciones top-hat y bot-hat:

$$\mathfrak{R}_{n,\lambda}(\mathbf{x}) = Th_\lambda(I_t)(\mathbf{x}) = I_t(\mathbf{x}) - \gamma_\lambda(I_t)(\mathbf{x}) \quad (4.3)$$

$$\mathfrak{R}_{p,\lambda}(\mathbf{x}) = Bh_\lambda(I_t)(\mathbf{x}) = \phi_\lambda(I_t)(\mathbf{x}) - I_t(\mathbf{x}) \quad (4.4)$$

A partir de estos residuos, se genera el mapa de textura actual como:

$$\tau_\lambda(I_t)(\mathbf{x}) = \text{median} \{ \varsigma - \mathfrak{R}_{n,\lambda}(\mathbf{x}), \varsigma + \mathfrak{R}_{p,\lambda}(\mathbf{x}) \} \quad (4.5)$$

Donde ς representa el valor central del intervalo morfológico definido por $\inf I_t$ y $\sup I_t$.

Posteriormente, se compara el mapa de textura actual con el modelo de fondo. La desviación entre ambos se evalúa en una ventana temporal η y se modela mediante una función de densidad de probabilidad empírica. Un clasificador binario $C_{\eta,\lambda}(\mathbf{x})$ determina si existe una desviación significativa que indique la presencia de movimiento:

$$C_{\eta,\lambda}(\mathbf{x}) = \begin{cases} 1 & \text{si } \Gamma_{\eta,\lambda}^*(\mathbf{x}) \not\sim \mathcal{M}_\lambda(\mathbf{x}) \\ 0 & \text{en otro caso} \end{cases} \quad (4.6)$$

El valor $C_{\eta,\lambda}(\mathbf{x}) = 1$ indica que la posición \mathbf{x} ha experimentado un cambio en la textura con respecto al modelo de fondo y se considera como parte del primer plano. El resultado es una máscara binaria \mathcal{F}_t que representa las regiones con presencia de objetos móviles:

$$\mathcal{F}_t(\mathbf{x}) = \begin{cases} 1 & \text{si } C_{\eta,\lambda}(\mathbf{x}) = 1 \\ 0 & \text{si } C_{\eta,\lambda}(\mathbf{x}) = 0 \end{cases}$$

Esta máscara puede contener falsos positivos o pequeñas regiones mal segmentadas, que serán corregidas en la etapa de postprocesamiento morfológico.

4.5 Postprocesamiento morfológico

La máscara binaria generada por el clasificador $C_{\eta,\lambda}(\mathbf{x})$ puede contener imperfecciones como ruido impulsivo, regiones fragmentadas o detecciones aisladas que no corresponden a objetos reales. Para mejorar la calidad de segmentación, se aplica una etapa de postprocesamiento morfológico que refina la forma y la conectividad de las regiones detectadas.

Las transformaciones utilizadas en esta etapa son la apertura y el cierre morfológico, definidas respectivamente como:

$$\gamma_\lambda(\mathcal{F}_t) = \delta_\lambda(\xi_\lambda(\mathcal{F}_t)) \quad (4.7)$$

$$\phi_\lambda(\mathcal{F}_t) = \xi_\lambda(\delta_\lambda(\mathcal{F}_t)) \quad (4.8)$$

Donde:

- \mathcal{F}_t es la máscara binaria inicial del cuadro t .
- ξ_λ y δ_λ representan los operadores de erosión y dilatación, respectivamente.
- λ es el elemento estructurante, definido en función del tamaño mínimo esperado de los objetos en la escena.

El filtro de apertura γ_λ se emplea para eliminar pequeños objetos aislados y suavizar los bordes de las regiones detectadas, mientras que el cierre ϕ_λ rellena pequeños huecos en el interior de los objetos y mejora su continuidad.

Dependiendo de las características del ruido presente en la escena, se aplica uno u otro filtro, o una combinación de ambos en secuencia. En la práctica, la apertura es eficaz para reducir sobresegmentaciones, y el cierre es útil para eliminar discontinuidades en regiones parcialmente detectadas.

El resultado final de esta etapa es una máscara binaria $\mathcal{F}_t^{\text{refinada}}$ que conserva la estructura de los objetos móviles, reduce los errores de detección y está lista para ser utilizada en tareas de conteo, análisis de flujo vehicular o seguimiento.

4.5.1 Comparación metodológica con MoG

Para evaluar la pertinencia y originalidad del modelo propuesto, se realiza una comparación metodológica con uno de los enfoques más utilizados en sustracción de fondo: el modelo de mezcla de gaussianas.

El modelo MoG representa la distribución de intensidades de cada píxel como una combinación ponderada de K gaussianas. La probabilidad de que un nuevo valor de intensidad pertenezca al fondo depende de su cercanía a alguna de estas distribuciones:

$$P(x_t) = \sum_{i=1}^K \omega_i \cdot \mathcal{N}(x_t; \mu_i, \sigma_i^2)$$

Donde:

- x_t es la intensidad del píxel en el tiempo t .
- ω_i es el peso de la i -ésima gaussiana.
- μ_i y σ_i son la media y desviación estándar asociadas.
- \mathcal{N} representa una función de densidad gaussiana.

Aunque MoG ha demostrado ser eficaz en múltiples entornos, presenta varias limitaciones que el modelo morfológico propuesto busca superar:

1. Dependencia de supuestos estadísticos: MoG requiere asumir distribuciones gaussianas para modelar la variabilidad, lo cual puede ser inadecuado

en escenas con iluminación variable, texturas complejas o reflectancia no uniforme. En contraste, el modelo propuesto se basa en textura y estructura local, sin requerir hipótesis de normalidad.

2. Alto costo computacional: MoG debe mantener y actualizar múltiples parámetros por píxel $(\mu_i, \sigma_i, \omega_i)$, lo cual incrementa su complejidad. El enfoque morfológico, al ser determinista y no paramétrico, requiere menos memoria y es más eficiente en implementaciones embebidas.
3. Sensibilidad a ruido y fluctuaciones: Las variaciones leves en la intensidad pueden generar falsos positivos en MoG. En cambio, el modelo basado en textura morfológica filtra el ruido estructural mediante transformaciones topográficas robustas, como las operaciones de apertura, cierre y mediana morfológica.
4. Interpretabilidad estructural: El modelo propuesto ofrece una representación visual intuitiva del proceso de detección (residuos, mapa de textura, modelo de fondo), facilitando su análisis, ajuste y validación en tiempo real.

Tabla 4-1 Comparación metodológica entre MoG y el modelo morfológico propuesto.

Característica	MoG	Propuesta morfológica
Tipo de modelo	Estadístico paramétrico	Determinista no paramétrico
Información utilizada	Intensidad por píxel	Textura morfológica por píxel
Representación del fondo	Combinación de gaussianas	Moda en secuencia de textura
Robustez ante variaciones de iluminación	Media	Alta
Interpretabilidad estructural	Limitada	Alta (basada en residuos y textura)
Requerimiento de parámetros	Alto $(\mu_i, \sigma_i, \omega_i)$ por píxel	Bajo (elemento estructurante λ)
Complejidad computacional	Alta	Media/Baja
Aplicabilidad en sistemas embebidos	Limitada	Alta

Como se muestra en la Tabla 4-1, el modelo morfológico presenta ventajas significativas en términos de robustez, simplicidad computacional y claridad estructural, lo que lo convierte en una alternativa viable y eficaz para sistemas de visión embebida o plataformas autónomas como drones.

4.6 Consideraciones técnicas de implementación

En esta sección se detallan aspectos técnicos relacionados con la implementación práctica del modelo propuesto, haciendo énfasis en el manejo de la distribución de textura por píxel y los criterios empleados para clasificar si una muestra de intensidad pertenece al fondo o al primer plano. El objetivo es garantizar que el modelo, además de ser teóricamente sólido, pueda implementarse eficientemente en sistemas computacionales con recursos limitados.

4.6.1 Representación matricial del modelo de fondo

El modelo de fondo \mathcal{M} se representa mediante un arreglo de contadores por píxel, donde cada contador almacena la frecuencia con la que se ha observado un determinado valor de textura. Esta representación discreta permite construir una aproximación empírica de la función de densidad de probabilidad (PDF) sin asumir una forma paramétrica.

Cada modelo $\mathcal{M}(\mathbf{x})$ se estructura como un vector de frecuencias de tamaño 256 en el caso de imágenes de 8 bits, y su evolución se actualiza con cada nueva imagen de entrada. Esto permite capturar tanto la información espacial (a través del elemento estructurante λ) como la evolución temporal (ventana η), manteniendo un historial comprimido de la textura local por píxel.

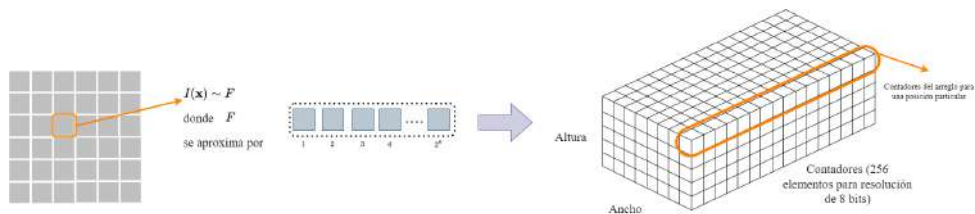


Figura 4.4: Representación matricial del modelo de textura por píxel mediante contadores.

4.6.2 Criterio de pertenencia a la distribución

Para decidir si una nueva muestra de textura pertenece al fondo o al primer plano, se evalúa su proximidad con la distribución estimada. El criterio se basa en la forma de la PDF empírica, utilizando la posición del máximo global $\max p(\mathcal{M}(\mathbf{x}))$ como valor esperado, y midiendo los cambios de concavidad a izquierda (Δ_l) y derecha (Δ_r), lo cual define un umbral de dispersión.

Una muestra se considera perteneciente al fondo si se encuentra dentro de este intervalo simétrico respecto al modo. De lo contrario, se clasifica como movimiento. Este criterio proporciona una forma eficiente de evaluar pertenencia sin requerir funciones de probabilidad explícitas ni cálculos complejos.

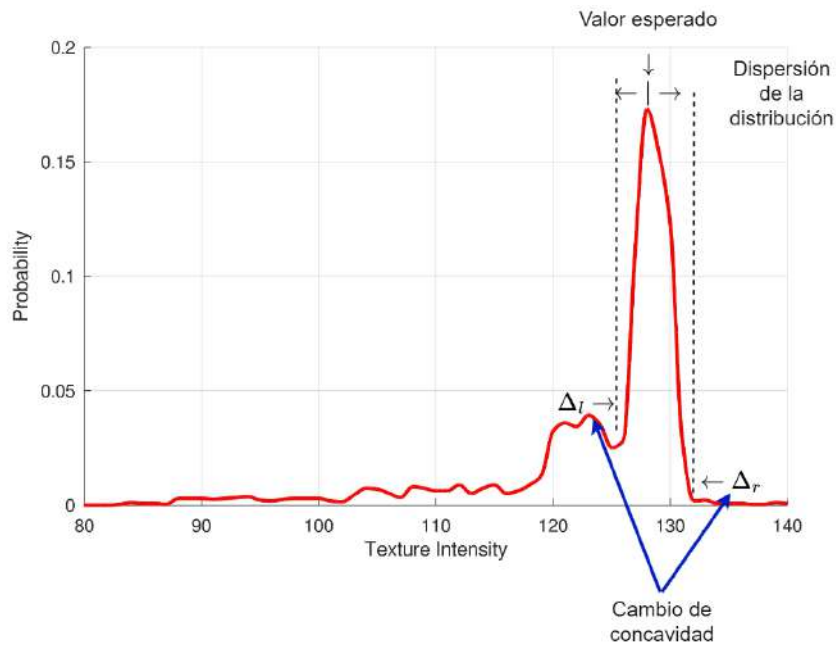


Figura 4.5: Criterio de pertenencia basado en la concavidad de la función de densidad empírica. Se utiliza el valor máximo global como referencia y se identifican los cambios de concavidad a izquierda (Δ_l) y derecha (Δ_r), delimitando el intervalo de pertenencia.

4.6.3 Consumo de memoria por resolución

La implementación del modelo de fondo requiere mantener un buffer de contadores por cada posición de la imagen. La Tabla 4-2 muestra el tamaño de este buffer (en Mega contadores) para diferentes resoluciones de video y profundidades de bits. El consumo de memoria es lineal respecto al número de

píxeles y al rango de valores posibles por píxel.

Resolución	Alto	Ancho	8 bits ¹	10 bits ¹	12 bits ¹
360p	640	360	56.3	225.0	900.0
480p	854	480	100.1	400.3	1601.3
720p	1280	720	225.0	900.0	3600.0
1080p	1920	1080	506.3	2025.0	8100.0
1440p	2560	1440	900.0	3600.0	14400.0
4K	3840	2160	2025.0	8100.0	32400.0

¹ Expresado en Mega contadores. Para obtener el tamaño real en bytes, se multiplica por el tamaño del contador.

Tabla 4-2 Tamaño del buffer requerido para distintas resoluciones y profundidades de bits.

4.6.4 Algoritmo de actualización del modelo

Cada vez que se analiza un nuevo cuadro de la secuencia, se actualiza el modelo de fondo incrementando el contador correspondiente a la textura observada en la posición \mathbf{x} . Esta operación puede implementarse de manera eficiente con una indexación directa:

$$\mathcal{M}(\mathbf{x})[\tau_{\lambda}(I_t)(\mathbf{x})] \leftarrow \mathcal{M}(\mathbf{x})[\tau_{\lambda}(I_t)(\mathbf{x})] + 1$$

El siguiente algoritmo resume este procedimiento:

Algoritmo 1 Actualización del modelo de textura

Entrada: \mathcal{M}_{t-1} : modelo anterior; Γ : nueva imagen de textura.

Salida: \mathcal{M}_t : modelo actualizado.

- 1: **Para cada** pixel en \mathbf{x} **hacer**
 - 2: $\mathcal{M}_t(\mathbf{x}) \leftarrow \mathcal{M}_{t-1}(\mathbf{x})$
 - 3: $\mathcal{M}_t(\mathbf{x})[\Gamma(\mathbf{x})] \leftarrow \mathcal{M}_t(\mathbf{x})[\Gamma(\mathbf{x})] + 1$
 - 4: **Fin**
-

Esta estrategia ofrece robustez frente a cambios globales en la escena, ya que el modelo aprende gradualmente la estructura dominante sin depender de parámetros ajustados manualmente.

4.6.5 Mejora de conectividad en el mapa de movimiento

La etapa de segmentación puede generar regiones poco conectadas o con objetos fusionados. Para corregir esto, se propone un criterio morfológico que aplica apertura o cierre según el tipo de error detectado: si el mapa de movimiento está sobredetectado (objetos fusionados), se utiliza apertura; si está subdetectado (objetos fragmentados), se aplica cierre.

Algoritmo 2 Mejora de conectividad morfológica

Entrada: I : mapa de movimiento; Λ : parámetros y elementos estructurantes.

Salida: I^* : mapa corregido.

Si I está sobredetectado según $\Lambda(1)$ **entonces**

$I^* \leftarrow \gamma_{\Lambda(2)}(I)$ ▷ Apertura

Si no

$I^* \leftarrow \phi_{\Lambda(2)}(I)$ ▷ Cierre

Fin

5. RESULTADOS PRELIMINARES

5.1 Descripción del proceso de validación

Para evaluar la efectividad y robustez del modelo propuesto de sustracción de fondo basado en morfología matemática (MMBS), se diseñó un proceso experimental dividido en dos etapas principales:

- Validación cuantitativa del desempeño en detección de movimiento, mediante conjuntos de datos con etiquetas (ground truth).
- Aplicación práctica del modelo en escenarios reales para analizar el flujo vehicular en glorietas urbanas.

La primera etapa tiene como objetivo contrastar el rendimiento del modelo MMBS frente a algoritmos de referencia de la familia Mixture of Gaussians (MoG) (Stauffer & Grimson, 1999; Trejo-Morales et al., 2023), utilizando métricas estandarizadas de precisión y error. Esta comparación se enfoca tanto en escenarios simples como en entornos desafiantes, evaluando la capacidad del modelo para generalizar bajo distintas condiciones de iluminación, ruido estructural y dinámica de escena.

Los conjuntos de datos empleados en esta validación son:

- Performance Evaluation of Tracking and Surveillance Dataset (PETS) 2001 (Jiménez-Hernández, 2010; The University of Reading, 2001; Thirde et al., s.f.): contiene secuencias representativas de videovigilancia en interiores y exteriores, incluyendo etiquetas de movimiento. Este dataset permite comparar directamente los mapas de movimiento generados por MMBS con los de la implementación original de MoG.
- Change Detection Dataset (CDnet) 2014 (Y. Wang et al., 2014): base de datos ampliamente utilizada en la literatura, que incluye secuencias clasificadas en categorías como escenarios estáticos, dinámicos, interiores y exteriores. Se utilizó una versión mejorada del algoritmo MoG (Trejo-Morales et al., 2023) como referencia para esta comparación.

El proceso de validación incluye dos tipos de mediciones:

1. Medidas de clasificación binaria, obtenidas a partir de la matriz de confusión:

- Precisión (Precision): proporción de verdaderos positivos sobre el total de predicciones positivas.
- Exhaustividad (Recall): proporción de verdaderos positivos sobre el total de etiquetas positivas reales.
- F1-Score: media armónica entre precisión y recall.
- Intersección sobre Unión (IoU): medida del solapamiento entre las regiones predicha y real.

2. Medidas de error cuantitativo, utilizadas para evaluar la magnitud y consistencia de la segmentación:

- Bias: error sistemático (diferencia media entre predicción y ground truth).
- Desviación estándar del Bias.
- Error absoluto medio (Mean Absolute Error (Error Absoluto Medio) (MAE)).
- Raíz del error cuadrático medio (Root Mean Squared Error (Raíz del Error Cuadrático Medio) (RMSE)).

La segunda etapa corresponde a una validación cualitativa, en la cual se aplica el modelo MMBS sobre videos aéreos tomados con dron en glorietas reales de la ciudad de Querétaro, México. Esta aplicación tiene como objetivo generar mapas de movimiento que permitan analizar el flujo vehicular y construir mapas de densidad para apoyar la planificación urbana.

A través de esta doble validación (cuantitativa y contextual) se busca demostrar que el modelo propuesto no sólo tiene un desempeño competitivo respecto a algoritmos tradicionales, sino que también presenta ventajas prácticas al operar en condiciones reales, con bajo costo computacional y alta estabilidad frente a variaciones ambientales.

5.2 Evaluación con el conjunto PETS

Para la primera evaluación cuantitativa se utilizó el conjunto de datos PETS 2001 (The University of Reading, 2001), que ofrece secuencias de video en condiciones controladas y etiquetas de movimiento cuadro a cuadro. Las pruebas se dividieron en tres escenarios:

- Test 1: Secuencia capturada en exteriores, con compresión alta y variaciones de luminancia.
- Test 2: Escenario en interiores con iluminación más estable.
- Test 3: Otro escenario interior con condiciones similares, pero cambios de fondo y presencia de sombras.

Se compararon los resultados del modelo propuesto (MMBS) con el algoritmo de referencia MoG (Stauffer & Grimson, 1999), evaluando el rendimiento en términos de precisión, recall, F1-score e IoU.

5.2.1 Resultados y análisis

Los resultados obtenidos se muestran en las Tablas 5-1 y 5-2 para el modelo MMBS y MoG respectivamente.

Tabla 5-1 Medidas de desempeño obtenidas con el enfoque MMBS en el conjunto PETS2001.

No.	Escenario	Precisión	Recall	F1 Score	IoU
1	Test 1	0.38966↓	0.75231↑	0.51341	0.34536
2	Test 2	0.51876	0.57041	0.54336↑	0.37302↑
3	Test 3	0.53064↑	0.49542↓	0.51242↓	0.34447↓

Tabla 5-2 Medidas de desempeño obtenidas con el enfoque MoG en el conjunto PETS2001.

No.	Escenario	Precisión	Recall	F1 Score	IoU
1	Test 1	0.02790↓	0.54002	0.05306↓	0.02725↓
2	Test 2	0.56424	0.77703↑	0.65376↑	0.48562↑
3	Test 3	0.57203↑	0.41362↓	0.48009	0.31587

Los resultados muestran que el modelo MMBS logra una mejor precisión en exteriores (Test 1), donde el modelo basado en textura demuestra mayor robustez frente a variaciones de iluminación y ruido estructural. En los escenarios interiores (Tests 2 y 3), los valores obtenidos por ambos métodos son más cercanos, aunque MoG logra un mejor desempeño global en precisión y F1 en Test 2.

El MMBS tiende a generar más volumen de detección (mayor recall), lo cual sugiere una cobertura más completa del objeto en movimiento, pero puede penalizar la precisión en casos donde los objetos son pequeños o están cerca del umbral del elemento estructurante.

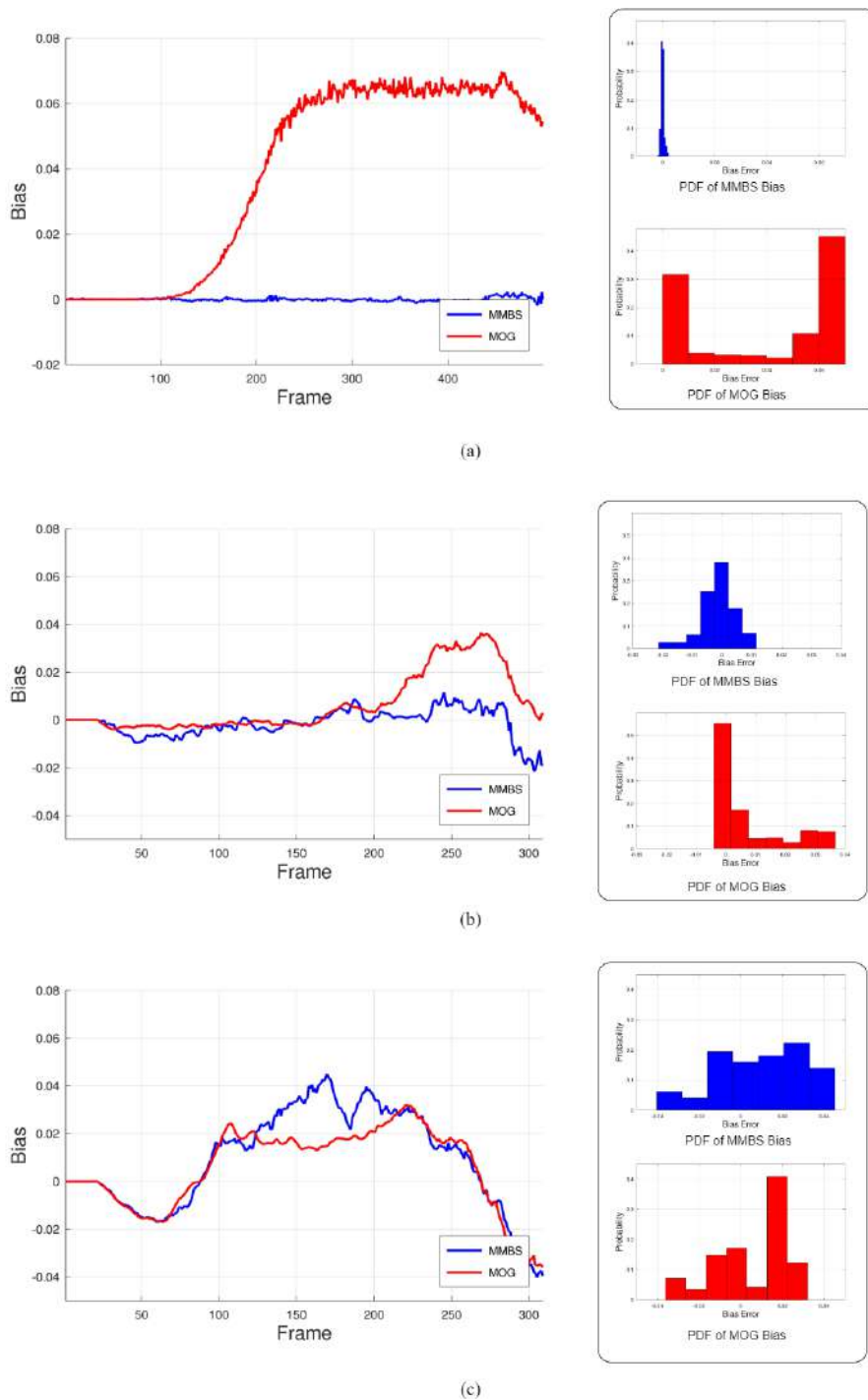
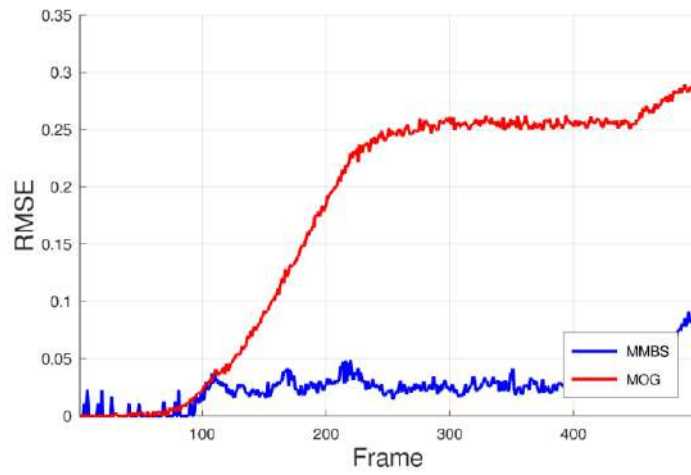
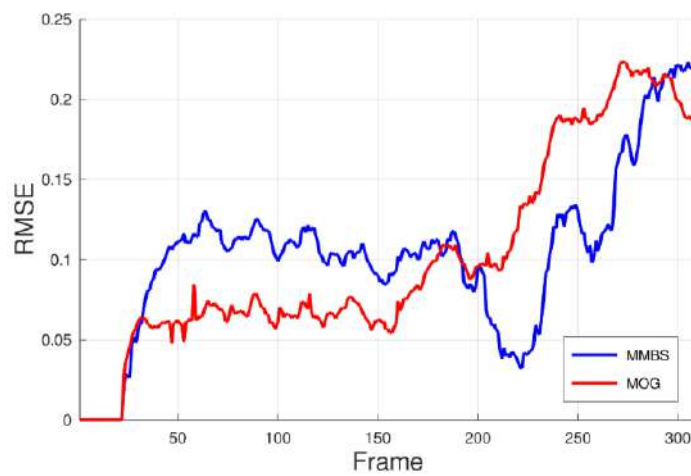


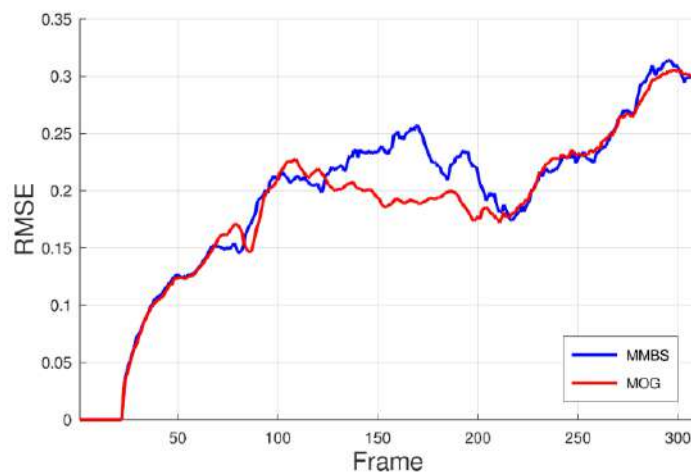
Figura 5.1: Error de sesgo en el conjunto de datos PETS (The University of Reading, 2001) entre MMBS y MoG; los gráficos del lado derecho representan la distribución del error de sesgo, la cual puede observarse para ambos enfoques. Una distribución simétrica centrada en cero indica un mejor desempeño en la detección de movimiento. La subfigura (a) corresponde a un escenario al aire libre; (b) y (c) representan escenas en interiores, con condiciones de iluminación artificial.



(a)



(b)



(c)

Figura 5.2: Error RMSE en el conjunto de datos PETS (The University of Reading, 2001) entre MMBS y MoG; en este contexto, valores más pequeños indican mayor precisión. En (a), las condiciones al aire libre hacen que MOG sea menos preciso que la propuesta, y en (b) y (c), bajo condiciones en interiores, ambos enfoques presentan un desempeño similar.

La comparación visual (ver Figuras 5.1 y 5.2) también revela que la segmentación generada por MMBS presenta menos artefactos internos (como huecos en las siluetas), mientras que MoG tiende a detectar mejor bordes, pero con mayor fragmentación.

5.3 Evaluación con el conjunto CDnet2014

El segundo conjunto de pruebas se realizó con la base de datos CDnet 2014 (Y. Wang et al., 2014), ampliamente utilizada en la evaluación de algoritmos de sustracción de fondo debido a su diversidad de escenarios, incluyendo condiciones dinámicas, ambientes interiores/exteriores, sombras, reflejos y cambios abruptos de iluminación.

En esta evaluación se utilizó una versión mejorada del algoritmo MoG, referenciada en (Trejo-Morales et al., 2023), diseñada específicamente para adaptarse a entornos más exigentes y para mejorar el modelado de movimiento bajo condiciones difíciles.

Los escenarios se dividieron en dos grupos:

- Escenarios base (1–4): condiciones relativamente estables donde se espera un desempeño sólido de cualquier modelo.
- Escenarios dinámicos (5–10): presentan desafíos como agua en movimiento, objetos no rígidos, reflejos, lluvia o condiciones de baja resolución.

5.3.1 Resultados y análisis

Las Tablas 5-3 y 5-4 muestran las métricas de clasificación (Precisión, Recall, F1 Score e IoU) para ambos enfoques.

Tabla 5-3 Medidas de desempeño con MMBS en el conjunto CDnet2014.

No.	Escenario	Precisión	Recall	F1 Score	IoU
1	Highway	0.46359	0.75820	0.57537 [↑]	0.40388 [↑]
2	Office	0.54534 [↑]	0.44720	0.49142	0.32575
3	Pedestrian	0.41343	0.87766 [↑]	0.56208	0.39090
4	Pets2006	0.32404	0.42469 [↓]	0.36760	0.22519
5	Boats	0.10741	0.60489	0.18242	0.10036
6	Canoe	0.29241	0.52898	0.37663	0.23200
7	Fall	0.10654	0.43008	0.17078	0.09336
8	Fountain1	0.01793	0.65089	0.03491 [↓]	0.01776 [↓]
9	Fountain2	0.17645 [↓]	0.59594	0.27228	0.15759
10	Overpass	0.23281	0.57132	0.33082	0.19819

Tabla 5-4 Medidas de desempeño con MOG mejorado (Trejo-Morales et al., 2023) en el conjunto CDnet2014.

No.	Escenario	Precisión	Recall	F1 Score	IoU
1	Highway	0.26669	0.33107	0.29541	0.17330
2	Office	0.31667	0.12260 [↓]	0.17676	0.09695
3	Pedestrian	0.47295 [↑]	0.88376 [↑]	0.61616 [↑]	0.44525 [↑]
4	Pets2006	0.32034	0.55064	0.40504	0.25395
5	Boats	0.00901	0.25911	0.01741	0.00878
6	Canoe	0.03778	0.24702	0.06553	0.03388
7	Fall	0.07405	0.48349	0.12843	0.06862
8	Fountain1	0.00869 [↓]	0.66048	0.01715 [↓]	0.00865 [↓]
9	Fountain2	0.02033	0.60599	0.03933	0.02006
10	Overpass	0.06016	0.31101	0.10081	0.05308

En los escenarios base (1 a 4), el modelo MMBS obtiene mejores resultados generales en F1 Score e IoU. En particular, logra hasta un 22 % de mejora sobre el modelo de referencia, con un desempeño competitivo incluso en condiciones interiores. El único escenario donde MoG supera ampliamente a MMBS es “Pedestrian”, debido a la pequeña escala de los objetos en movimiento, lo que afecta la detección cuando el elemento estructurante no se ajusta con precisión.

En los escenarios dinámicos (5 a 10), ambos algoritmos sufren caídas

en precisión, pero MMBS mantiene una ventaja sostenida en F1 e IoU. Esto se debe a su capacidad de modelar la textura y filtrar ruido estructural, incluso en condiciones con agua, reflejos o movimiento no estructurado.

Estos resultados muestran que MMBS no solo ofrece una alternativa competitiva, sino que además demuestra mayor adaptabilidad a variaciones complejas de escena, con bajo costo computacional y buena interpretabilidad visual.

Las Tablas 5-5 y 5-6 muestran los valores obtenidos para ambas metodologías en los diez escenarios del conjunto CDnet2014.

Tabla 5-5 Medidas de error obtenidas con MMBS en el conjunto CDnet2014.

No.	Escenario	Bias	std Bias	MAE	RMSE
1	Highway	0.00735	0.03708	0.02640	0.00142
2	Office	-0.02496 [↓]	0.01988	0.02498	0.00101
3	Pedestrian	-0.00224	0.00352 [↓]	0.00263 [↓]	0.000017 [↓]
4	Pets2006	-0.00208	0.01633	0.01222	0.00027
5	Boats	0.00988	0.00982	0.01197	0.00019
6	Canoe	0.00305	0.01933	0.01460	0.00038
7	Fall	0.03854 [↑]	0.04751 [↑]	0.04691 [↑]	0.00374 [↑]
8	Fountain1	0.01729	0.00795	0.01729	0.00036
9	Fountain2	0.00174	0.00411	0.00331	0.000019
10	Overpass	0.00867	0.02139	0.01578	0.00053

Tabla 5-6 Medidas de error obtenidas con MOG mejorado en el conjunto CDnet2014.

No.	Escenario	Bias	std Bias	MAE	RMSE
1	Highway	-0.00850	0.05667 [↑]	0.04043	0.00328
2	Office	-0.04560 [↓]	0.05105	0.05356	0.00468
3	Pedestrian	-0.00313	0.00453 [↓]	0.00350 [↓]	0.00003 [↓]
4	Pets2006	0.00271	0.01756	0.01121	0.00031
5	Boats	0.06837	0.01524	0.06837	0.00490
6	Canoe	0.05799	0.04239	0.06454	0.00515
7	Fall	0.07183 [↑]	0.04861	0.07880 [↑]	0.00752 [↑]
8	Fountain1	0.03902	0.01455	0.03902	0.00173
9	Fountain2	0.04031	0.01619	0.04031	0.00188
10	Overpass	0.03232	0.02950	0.03964	0.00191

Los resultados muestran que el modelo MMBS mantiene un bias cercano a cero en la mayoría de los escenarios, indicando una segmentación más balanceada entre positivo y negativo. Además, los valores de MAE y RMSE son sistemáticamente menores en comparación con el modelo MoG mejorado, reflejando un desempeño más estable y menos disperso.

La desviación estándar del bias es especialmente baja en MMBS en escenarios como *Pedestrian* y *Fountain2*, donde MoG tiende a mostrar más dispersión. Esta diferencia en la estabilidad del error confirma la ventaja de utilizar textura estructural en lugar de intensidad directa, ya que reduce la sensibilidad a fluctuaciones locales.

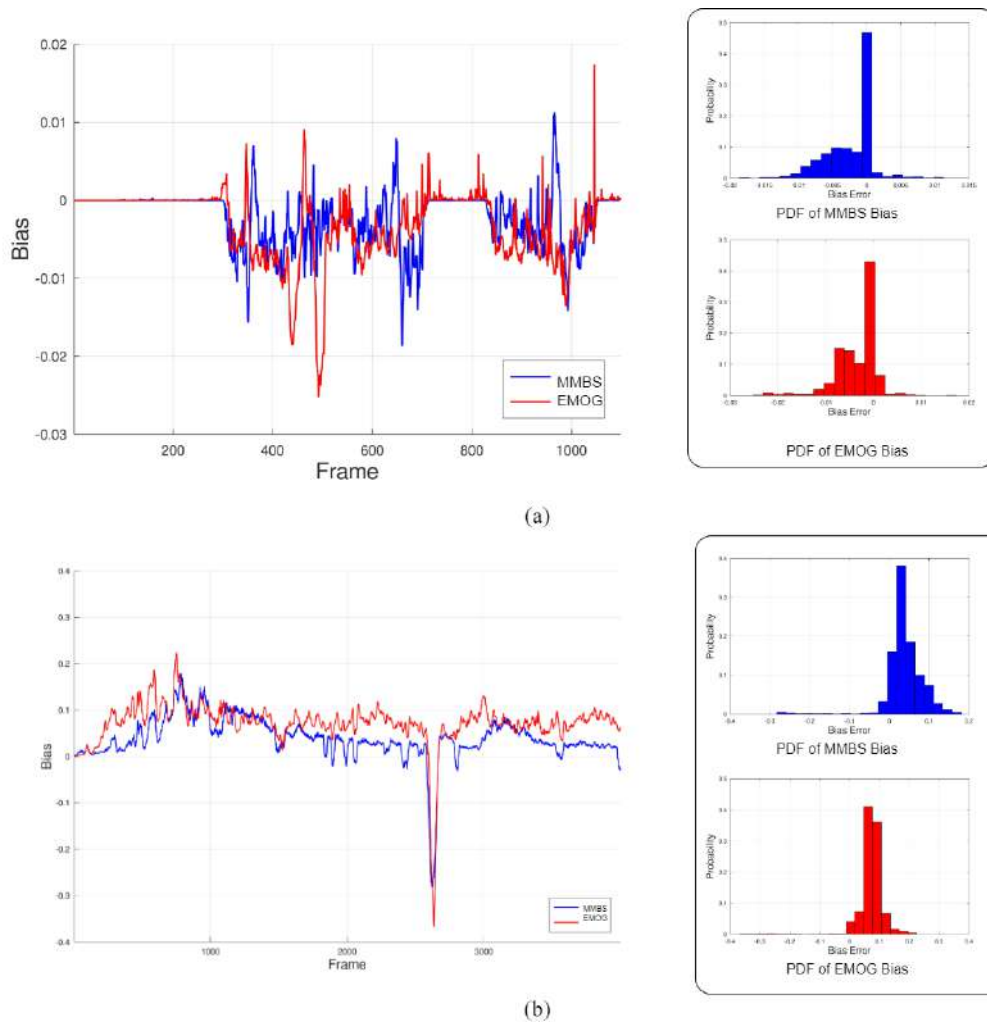


Figura 5.3: Error de sesgo calculado a partir del conjunto de datos CDnet 2014 (Y. Wang et al., 2014) entre los enfoques MMBS y MoG mejorado, donde (a) representa la escena con mejor desempeño en la detección de movimiento, mientras que (b) representa la escena con el peor desempeño en dicha tarea.

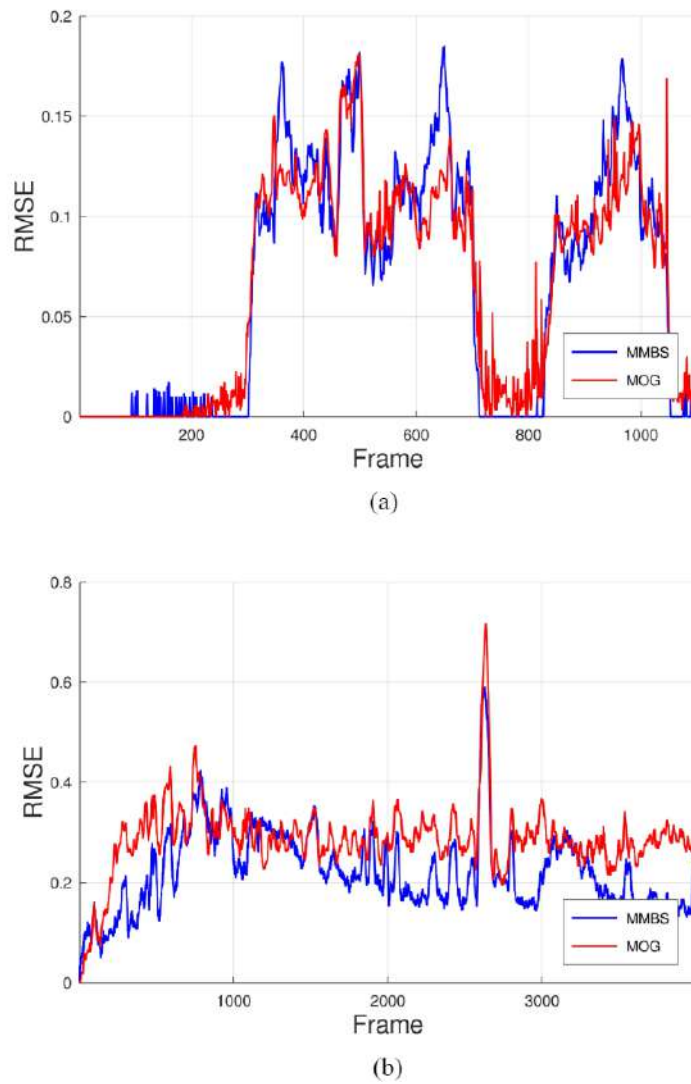


Figura 5.4: Error RMSE calculado a partir del conjunto de datos CDnet 2014 (Y. Wang et al., 2014) utilizando los enfoques MMBS y MoG mejorado, donde (a) representa la escena con mejor desempeño en la detección de movimiento, mientras que (b) representa la escena con el peor desempeño en dicha tarea.

Las Figuras 5.3 y 5.4 complementan este análisis, ilustrando cómo las distribuciones del error se centran más alrededor de cero en el caso de MMBS, mientras que en MoG tienden a desplazarse y mostrar mayor variabilidad.

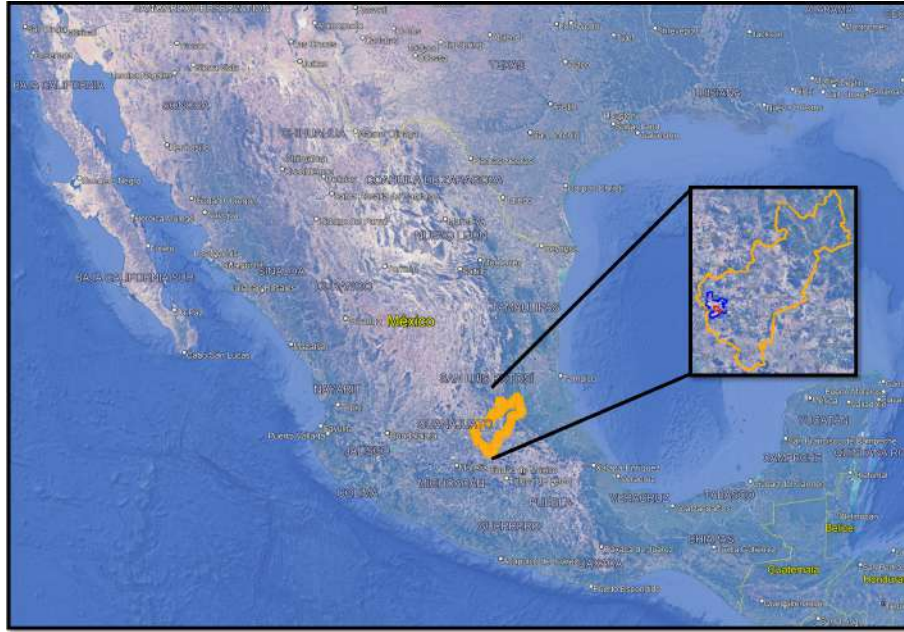
5.4 Comparación visual cualitativa en rotondas

Además de la validación cuantitativa, se realizó una prueba práctica en escenarios reales, utilizando grabaciones aéreas capturadas por un dron sobre glorietas urbanas en la ciudad de Querétaro, México. Esta prueba busca evaluar

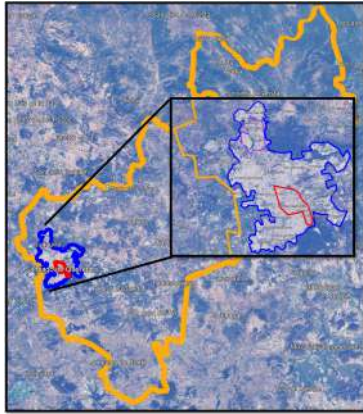
la capacidad del modelo MMBS para detectar y representar el flujo vehicular en condiciones reales, con variaciones de iluminación, textura, reflejos, y sin etiquetas de ground truth.

5.4.1 Contexto urbano

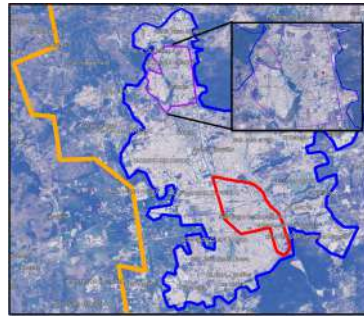
En los últimos años, la ciudad de Querétaro ha experimentado un crecimiento urbano acelerado, impulsado por la migración de zonas rurales y la expansión de zonas suburbanas. Como parte de las estrategias de movilidad, se han reemplazado semáforos por glorietas para mejorar la fluidez del tráfico. Sin embargo, esto exige monitoreo constante del flujo vehicular en puntos clave para evitar congestiones y optimizar el uso de recursos viales.



(a)



(b)



(c)



(d)

Figura 5.5: Escenario de prueba real para analizar el flujo vehicular, donde (a) muestra la ubicación a nivel estatal, (b) la ubicación a nivel ciudad, (c) la ubicación a nivel delegación, y (d) la ubicación de la glorieta.

5.4.2 Evaluación cualitativa

Se comparó visualmente el desempeño del algoritmo MMBS con el modelo MoG sobre los mismos cuadros de video. Las Figuras 5.6 y 5.7 ilustran ejemplos representativos obtenidos en distintos momentos del día.

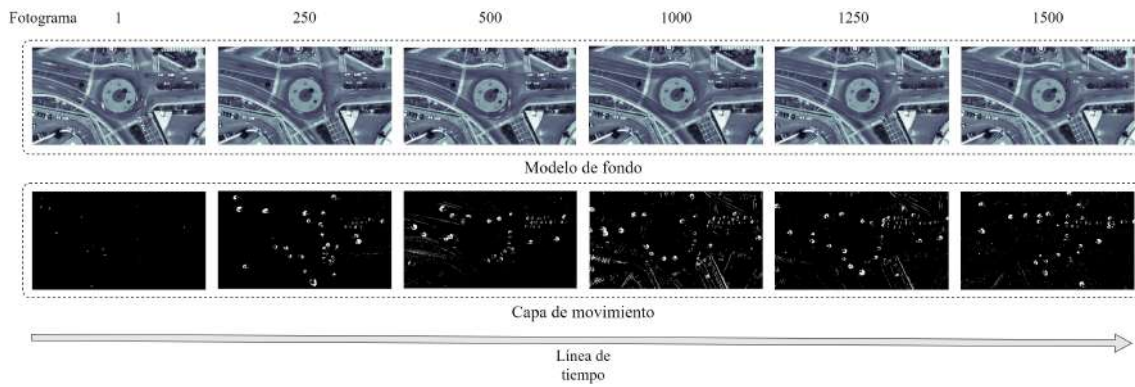


Figura 5.6: Resultados obtenidos con el modelo MMBS en glorietas urbanas: se aprecia la textura, el mapa de movimiento y su estabilidad ante cambios de iluminación.

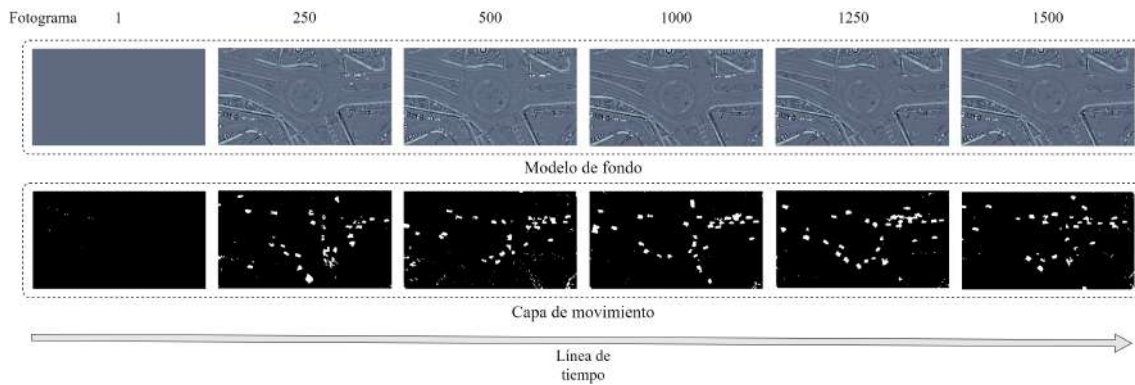


Figura 5.7: Resultados obtenidos con el modelo MoG en los mismos escenarios: se observan zonas de sobre-segmentación y artefactos por luminancia.

En la Figura 5.6, el modelo MMBS demuestra una detección precisa de vehículos, con menor ruido en el mapa de movimiento y mayor homogeneidad en zonas estáticas. En contraste, el modelo MoG (Figura 5.7) presenta detecciones fragmentadas, mayor sensibilidad a sombras y reflejos, y regiones incompletas cuando el contraste de intensidad disminuye.

Estas diferencias se explican por la naturaleza del modelo morfológico, que se basa en la textura estructural local y no en intensidades absolutas. Esto lo hace más resistente a variaciones globales como cambios de luz, nubes, o balance automático de blancos de la cámara, lo cual es crítico en escenarios reales al aire libre.

5.4.3 Observaciones clave

- MMBS presenta una segmentación más coherente y estable en presencia de tráfico continuo.
- MoG tiende a fallar en condiciones de iluminación cambiante, generando regiones de movimiento discontinuas.
- El modelo MMBS es capaz de distinguir mejor entre vehículos y fondo, incluso cuando estos se superponen parcialmente.

En el siguiente apartado se presentan los mapas de densidad generados con la información obtenida por MMBS, lo que permite visualizar patrones de congestión a lo largo del tiempo en las zonas monitoreadas.

5.5 Mapas de densidad de tráfico

A partir de los resultados obtenidos por el modelo MMBS, se desarrolló una aplicación adicional para visualizar la concentración de vehículos a lo largo del tiempo mediante la construcción de mapas de densidad. Esta representación permite identificar zonas con mayor flujo vehicular, lo que resulta útil para la planificación urbana, el diseño de infraestructura vial y la toma de decisiones en tiempo real.

5.5.1 Generación del mapa de densidad

El procedimiento consiste en acumular las máscaras de movimiento generadas por MMBS durante un intervalo de tiempo definido. Cada píxel que representa una detección de objeto en movimiento se suma en una matriz acumulativa, la cual posteriormente se normaliza y se representa mediante un mapa de color (colormap) que resalta las áreas de mayor actividad.

- Entrada: secuencia de máscaras binarias generadas por MMBS (\mathcal{F}_t).
- Proceso: suma acumulativa de detecciones por píxel.
- Salida: mapa de densidad codificado por color.

5.5.2 Resultados y análisis

La Figura 5.8 muestra un ejemplo representativo del mapa de densidad generado a partir de una grabación de 30 minutos en una glorieta de acceso a una avenida principal en Querétaro. Los colores más cálidos (rojo y amarillo) indican mayor densidad vehicular, mientras que los tonos fríos (azul y verde) representan menor actividad.

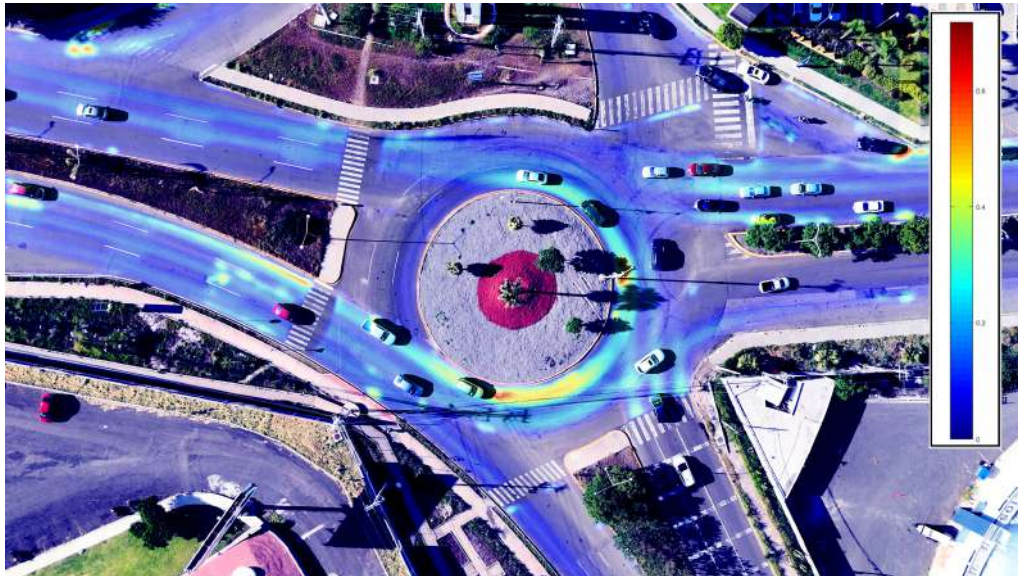


Figura 5.8: Mapa de densidad vehicular generado por el modelo MMBS a partir de detecciones acumuladas durante 30 minutos.

Se observa una clara concentración de flujo vehicular en la parte inferior derecha de la glorieta, correspondiente a una vía de acceso altamente transitada. Este tipo de visualización puede ser utilizada por autoridades de movilidad para:

- Detectar zonas de congestión en tiempo real.
- Optimizar la ubicación de señalización o semáforos temporales.
- Evaluar el impacto de cambios en la infraestructura (nuevas vías, cierres).

5.5.3 Ventajas del enfoque morfológico

La construcción de mapas de densidad mediante MMBS presenta varias ventajas:

1. Estabilidad en exteriores: las detecciones no se ven afectadas por cambios de luminancia o nubes pasajeras.
2. Robustez estructural: se detectan los vehículos en su totalidad, lo que genera mapas de densidad más uniformes y representativos.
3. Bajo costo computacional: al estar basado en operaciones morfológicas discretas, el algoritmo es apto para implementarse en tiempo real o sistemas embebidos.

5.6 Discusión general de resultados

Los resultados obtenidos en este estudio permiten validar la eficacia del modelo de sustracción de fondo basado en morfología matemática (MMBS) tanto en entornos controlados como en escenarios reales. A continuación, se sintetizan las principales observaciones que emergen del análisis experimental:

5.6.1 Desempeño cuantitativo

En comparación con el algoritmo de referencia MoG y su versión mejorada, el modelo MMBS mostró un desempeño competitivo y, en muchos casos, superior en métricas como F1-score e IoU, particularmente en escenarios exteriores o con condiciones dinámicas. Este comportamiento se debe a que MMBS opera sobre mapas de textura morfológica en lugar de intensidades absolutas, lo que lo hace más resistente a variaciones locales de iluminación, sombras o reflejos.

Las métricas de error (Bias, MAE, RMSE) también respaldan esta afirmación, ya que MMBS mantiene valores bajos y consistentes a lo largo de distintos escenarios, con distribuciones de error centradas alrededor de cero. Esta estabilidad es particularmente útil para aplicaciones en tiempo real, donde se requiere un modelo confiable bajo condiciones cambiantes.

5.6.2 Desempeño cualitativo

Los análisis visuales realizados con grabaciones de tráfico urbano en glorietas muestran que MMBS genera segmentaciones más completas y menos afectadas por el ruido ambiental. A diferencia de MoG, que presenta detecciones fragmentadas o huecos en los objetos, el modelo propuesto mantiene la coherencia espacial de las regiones en movimiento, incluso cuando la iluminación cambia gradualmente o cuando se presentan fusiones entre objetos.

La calidad de los mapas de movimiento generados permite extender el modelo hacia aplicaciones como la construcción de mapas de densidad vehicular, como se mostró en la sección anterior. Esto demuestra el potencial del modelo para tareas de análisis urbano, planificación de tráfico y evaluación de infraestructura.

5.6.3 Fortalezas y limitaciones

Las principales fortalezas observadas en el modelo MMBS incluyen:

- Robustez estructural: el uso de textura morfológica ofrece mayor estabilidad frente a ruido y variaciones de intensidad.
- Capacidad de generalización: el modelo funciona bien en diferentes condiciones, sin necesidad de ajustes específicos por escenario.
- Eficiencia computacional: al ser un modelo discreto y no paramétrico, su implementación es ligera y apta para sistemas embebidos o en tiempo real.

Sin embargo, también se identifican algunas limitaciones:

- Sensibilidad al tamaño del elemento estructurante: objetos muy pequeños pueden no ser correctamente detectados si su escala es menor a la de la estructura usada.
- Falta de modelado de fondo semántico: el modelo no distingue entre tipos de objetos (personas, vehículos), sino que segmenta movimiento de forma genérica.

- Ausencia de actualización adaptativa explícita: aunque el modelo se basa en una ventana temporal, no incorpora estrategias de actualización gradual como en MoG.

6. CONCLUSIONES

Este trabajo presentó el desarrollo, implementación y validación de un modelo de sustracción de fondo basado en morfología matemática (MMBS), orientado a mejorar la detección de objetos en movimiento en escenarios urbanos reales, particularmente bajo condiciones de iluminación variables. A lo largo del documento se expusieron los fundamentos teóricos, el diseño metodológico del modelo propuesto y un análisis detallado de los resultados obtenidos.

6.1 Cumplimiento de objetivos

Los resultados experimentales permiten afirmar que se cumplieron los objetivos planteados al inicio de esta investigación:

- Se realizó una revisión sistemática de la literatura sobre técnicas de morfología matemática y métodos actuales de sustracción de fondo, identificando sus limitaciones frente a condiciones de iluminación variables.
- Se diseñó un modelo de sustracción de fondo basado en operaciones morfológicas (MMBS), incorporando apertura, cierre, top-hat, bot-hat y una mediana morfológica adaptada, con el objetivo de modelar de forma precisa las texturas locales en escenas urbanas.
- El algoritmo fue implementado y probado en entornos reales y controlados, incluyendo conjuntos de datos con ground truth y grabaciones aéreas capturadas por dron.
- Se llevó a cabo una evaluación comparativa del modelo frente al enfoque Mixture of Gaussians (MoG), utilizando métricas estándar como F1 Score, IoU, Bias, MAE y RMSE.
- Se documentaron los resultados obtenidos, desarrollando visualizaciones como mapas de densidad vehicular que muestran el potencial del modelo para aplicaciones en análisis de movilidad urbana.

6.2 Evaluación de la hipótesis

La hipótesis planteada en este trabajo fue la siguiente:

Se plantea que un modelo de sustracción de fondo basado en operadores de morfología matemática permite segmentar con mayor precisión los objetos en movimiento en escenas urbanas reales, al tiempo que mantiene una menor complejidad computacional y mayor robustez ante variaciones de iluminación y ruido, en comparación con enfoques tradicionales como el modelo de mezcla de gaussianas.

Los resultados obtenidos respaldan esta hipótesis. En las pruebas realizadas sobre conjuntos de datos controlados (PETS, CDnet2014) y escenarios reales urbanos, el modelo MMBS demostró:

Una segmentación más precisa y coherente en presencia de ruido estructural, sombras y cambios de iluminación.

Un desempeño superior en métricas cuantitativas clave (F1 Score, IoU) frente al modelo MoG y su versión mejorada.

Una menor dispersión del error, evidenciada por valores de Bias, MAE y RMSE más estables y cercanos a cero.

Una implementación eficiente, sin necesidad de entrenamiento supervisado ni grandes recursos computacionales, lo que confirma su menor complejidad operativa.

Estos hallazgos confirman la validez de la propuesta y refuerzan el valor del enfoque morfológico para aplicaciones prácticas en visión por computadora en entornos urbanos reales.

6.3 Contribuciones principales

Las contribuciones más relevantes de esta tesis son:

- La propuesta de un modelo alternativo a MoG que prescinde de supuestos probabilísticos y se basa en la estructura geométrica local de la imagen.

- El diseño de un algoritmo discreto, eficiente y no paramétrico, implementado completamente con operaciones morfológicas.
- La validación del modelo con bases de datos de referencia y videos reales tomados por dron, demostrando su aplicabilidad a escenarios urbanos complejos.
- La construcción de mapas de densidad de tráfico como aplicación derivada del modelo, con utilidad directa en planeación vial y análisis de movilidad urbana.

6.4 Limitaciones y posibles mejoras

Aunque los resultados son prometedores, el modelo presenta ciertas limitaciones:

- La detección de objetos pequeños puede verse afectada por el tamaño del elemento estructurante.
- No incorpora una actualización adaptativa continua del fondo, lo que podría mejorar su desempeño en entornos altamente dinámicos.
- La segmentación es binaria y no distingue entre tipos de objetos (personas, vehículos, etc.).

6.5 Líneas futuras de investigación

Este trabajo sienta las bases para múltiples extensiones y mejoras. Entre ellas:

- Explorar el uso adaptativo del elemento estructurante en función del contexto espacial.
- Integrar un clasificador semántico para distinguir tipos de objetos dentro del primer plano.
- Paralelizar el algoritmo y optimizarlo para su ejecución en GPU o plataformas embebidas.

- Extender el modelo a secuencias Red, Green, Blue (RGB) o espacios de color alternativos como Luminance-Chrominance Color Space (YCbCr) o Hue, Saturation, Value (HSV).
- Utilizar el modelo para generar patrones temporales de tráfico y detectar anomalías o congestiones.

6.6 Reflexión final

El modelo propuesto representa una contribución relevante en el campo de la visión por computadora aplicada al análisis de tráfico urbano. Su enfoque estructural, robustez y eficiencia lo convierten en una alternativa sólida a modelos probabilísticos clásicos, y sus aplicaciones prácticas abren un camino para futuras investigaciones en contextos reales, orientadas a ciudades inteligentes y movilidad sustentable.

BIBLIOGRAFÍA

- Barnich, O., & Van Droogenbroeck, M. (2011). ViBe: A universal background subtraction algorithm for video sequences. *IEEE Transactions on Image Processing*, 20(6), 1709-1724. <https://doi.org/10.1109/TIP.2010.2101613>
- Bouwman, T. (2014). Traditional and recent approaches in background modeling for foreground detection: An overview. *Computer Science Review*, 11-12, 31-66. <https://doi.org/10.1016/j.cosrev.2014.04.001>
- Bouwman, T., Javed, S., Sultana, M., & Zahzah, E.-H. (2019). Deep learning for background subtraction: A comprehensive review and comparative evaluation. *Computer Science Review*, 31, 43-67. <https://doi.org/10.1016/j.cosrev.2018.11.003>
- Cavallaro, A., Piciarelli, C., & Foresti, G. L. (2020). UAV-based urban traffic monitoring using computer vision: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 169, 254-272. <https://doi.org/10.1016/j.isprsjprs.2020.09.006>
- Chen, L., Wu, T., Liu, M., & Zhang, Y. (2021). Survey on Computer Vision in Smart Transportation Systems. *IEEE Transactions on Intelligent Transportation Systems*, 22(12), 7780-7797. <https://doi.org/10.1109/TITS.2020.3032507>
- Ciuffreda, G., Palmieri, F., & Di Mauro, M. (2019). Urban traffic control using UAVs and deep learning. *Proceedings of the International Conference on Smart Cities*, 55-63. <https://doi.org/10.1109/SmartCities.2019.00014>
- Cristani, M., Bicego, M., & Murino, V. (2010). Audio-visual background modeling for people tracking under occlusions. *Pattern Recognition*, 43(3), 768-780. <https://doi.org/10.1016/j.patcog.2009.06.009>
- Elgammal, A., Harwood, D., & Davis, L. S. (2000). Non-parametric model for background subtraction. *European Conference on Computer Vision (ECCV)*, 751-767. https://doi.org/10.1007/3-540-45053-X_49
- Gonzalez, R. C., & Woods, R. E. (2009). *Digital Image Processing* (3rd). Pearson.
- Guo, Y., Li, X., & Porikli, F. (2016). Spatio-temporal Gaussian mixture model to detect moving objects in dynamic scenes. *Pattern Recognition*, 51, 172-180. <https://doi.org/10.1016/j.patcog.2015.08.001>
- Haralick, R. M., Sternberg, S. R., & Zhuang, X. (1987). Image analysis using mathematical morphology. *IEEE Transactions on Pattern Analysis and Ma-*

- chine Intelligence*, 9(4), 532-550. <https://doi.org/10.1109/TPAMI.1987.4767941>
- Hofmann, M., Tiefenbacher, P., & Rigoll, G. (2012). Background segmentation with feedback: The Pixel-Based Adaptive Segmenter. *IEEE International Conference on Advanced Video and Signal-Based Surveillance (AVSS)*, 25-30. <https://doi.org/10.1109/AVSS.2012.13>
- Huang, L., Li, Y., & Fang, M. (2022). Intelligent traffic monitoring system based on deep learning and edge computing. *IEEE Access*, 10, 41542-41552. <https://doi.org/10.1109/ACCESS.2022.3167829>
- Javed, S., Sultana, M., & Bouwmans, T. (2021). Background subtraction in real applications: Challenges, benchmark datasets and deep learning based solutions. *Computer Vision and Image Understanding*, 203, 103141. <https://doi.org/10.1016/j.cviu.2020.103141>
- Jiménez-Hernández, H. (2010). Background Subtraction Approach Based on Independent Component Analysis. *Sensors*, 10(6), 6092-6114. <https://doi.org/10.3390/s100606092>
- Kaur, G., & Kaur, R. (2023). Mathematical morphology and its applications in object detection. *Journal of Image Analysis and Recognition*, 11(2), 115-129. <https://doi.org/10.1007/s12350-023-00678-9>
- Klette, R. (2014). *Concise Computer Vision: An Introduction into Theory and Algorithms*. Springer. <https://doi.org/10.1007/978-1-4471-6320-6>
- Lim, J. H., Lee, D. W., & Yoon, K.-J. (2018). Foreground segmentation using convolutional neural networks with multi-scale feature encoding. *Pattern Recognition Letters*, 112, 256-262. <https://doi.org/10.1016/j.patrec.2018.08.024>
- Maddalena, L., & Petrosino, A. (2008). A self-organizing approach to background subtraction for visual surveillance applications. *IEEE Transactions on Image Processing*, 17(7), 1168-1177. <https://doi.org/10.1109/TIP.2008.924282>
- Mello Román, J. E., Gómez, L. J., & Arámbula Cosío, F. (2019). Modelo de fondo para secuencias de video basado en morfología matemática. *Revista Iberoamericana de Tecnologías del Aprendizaje*, 14(3), 91-98. <https://doi.org/10.1109/RITA.2019.2932253>
- Mohammed, M. A., Hamed, R. I., & García, A. M. (2022). Deep learning approaches for UAV-based object detection: A survey. *Remote Sensing*, 14(2), 298. <https://doi.org/10.3390/rs14020298>

- Piccardi, M. (2004). Background subtraction techniques: a review. *IEEE International Conference on Systems, Man and Cybernetics*, 4, 3099-3104. <https://doi.org/10.1109/ICSMC.2004.1400815>
- Rajendran, P., Ghosh, A., & Jaiswal, P. (2021). A survey on UAV-based traffic monitoring systems. *Sensors*, 21(17), 5795. <https://doi.org/10.3390/s21175795>
- Said, S., & Jambek, A. (2021). Foreground object detection using adaptive morphological operations. *Signal Processing: Image Communication*, 96, 116238. <https://doi.org/10.1016/j.image.2021.116238>
- Serra, J. (1982). *Image Analysis and Mathematical Morphology*. Academic Press.
- Soille, P. (2003). *Morphological Image Analysis: Principles and Applications* (2nd). Springer.
- Stauffer, C., & Grimson, W. E. L. (1999). Adaptive background mixture models for real-time tracking. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, 2, 246-252. <https://doi.org/10.1109/CVPR.1999.784637>
- Sternberg, S. R. (1986). Grayscale morphology. *Computer Vision, Graphics, and Image Processing*, 35(3), 333-355. [https://doi.org/10.1016/0734-189X\(86\)90096-1](https://doi.org/10.1016/0734-189X(86)90096-1)
- Sultana, M., Sufian, A., & Dutta, P. (2020). A Review of Object Detection Models Based on Background Subtraction in Video Surveillance Systems. *Journal of Imaging*, 6(6), 47. <https://doi.org/10.3390/jimaging6060047>
- Tezcan, C., Coulon, P.-Y., & Bouwmans, T. (2020). BSUV-Net: Background subtraction for unseen videos with deep neural networks. *Computer Vision and Image Understanding*, 197, 102973. <https://doi.org/10.1016/j.cviu.2020.102973>
- The University of Reading. (2001). Performance Evaluation of Tracking and Surveillance (PETS) [The University of Reading, UK].
- Thirde, D., Ferryman, J., & Crowley, J. (s.f.). Performance Evaluation of Tracking and Surveillance (PETS) [Accessed on 5 February 2010].
- Trejo-Morales, A., Córdova-Esparza, D., Rosas-Raya, C., Herrera-Navarro, A., & Jiménez-Hernández, H. (2023). Motion detection using MoG: a parametric analysis. *Revista Internacional de Investigación e Innovación Tecnológica (RIIIT)*. https://riiit.com.mx/apps/site/idem.php?module=Catalog&action=ViewItem&item_id=85476

- United Nations, Department of Economic and Social Affairs, Population Division. (2022). World Urbanization Prospects: The 2022 Revision [Accessed: 2025-05-20]. <https://population.un.org/wup/>
- Wang, H., Zhang, Y., & Liu, M. (2022). Texture-enhanced object detection using top-hat transforms in aerial traffic video. *Remote Sensing Letters*, 13(8), 761-770. <https://doi.org/10.1080/2150704X.2022.2098782>
- Wang, Y., Jodoin, P.-M., Porikli, F., Konrad, J., Benezeth, Y., & Ishwar, P. (2014). CDnet 2014: An Expanded Change Detection Benchmark Dataset. *Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 393-400. <https://doi.org/10.1109/CVPRW.2014.126>
- Wang, Z., Tang, X., & Li, F. (2023). Smart city traffic management using computer vision and IoT. *Journal of Urban Computing*, 5(1), 25-38. <https://doi.org/10.1016/j.juc.2023.01.003>
- Zhou, B., & Wang, Y. (2020). Vision-based vehicle detection and tracking: A review. *Sensors*, 20(7), 1946. <https://doi.org/10.3390/s20071946>
- Zivkovic, Z. (2004). Improved adaptive Gaussian mixture model for background subtraction. 2, 28-31. <https://doi.org/10.1109/ICPR.2004.1333992>