



**Universidad Autónoma de Querétaro**  
Facultad de Informática

Análisis de la dinámica de movimiento de objetos utilizando descriptores  
generales y estructurales

Tesis  
Que como parte de los requisitos para obtener el grado de  
Doctor en Ciencias de la Computación

Presenta  
Julio Alejandro Romero González

Santiago de Querétaro, Marzo, 2022



Universidad Autónoma de Querétaro  
Facultad de Informática  
Doctorado en Ciencias de la Computación

Análisis de la dinámica de movimiento de objetos utilizando descriptores generales y  
estructurales

TESIS

Que como parte de los requisitos para obtener el grado de  
Doctor en Ciencias de la Computación

Presenta:

Julio Alejandro Romero González

Dirigido por:

Dra. Ana Marcela Herrera Navarro

Co-dirigido por:

Dr. Hugo Jiménez Hernández

Dra. Ana Marcela Herrera Navarro

Presidente

Dr. Hugo Jiménez Hernández

Secretario

Dra. Diana Margarita Córdova Esparza

Vocal

Dr. Alberto Vázquez Cervantes

Suplente

Dra. Ma. Teresa García Ramírez

Suplente

Centro Universitario, Querétaro, Qro.

Marzo 2022

México.

---

## Dedicatorias

---

## Agradecimientos

Al término de esta etapa de mi vida, quiero expresar un profundo agradecimiento a quienes, con su ayuda, apoyo, comprensión, amistad, compañía, ánimo, me han ayudado a cruzar con firmeza el camino de la superación, porque con su apoyo y aliento hoy he logrado uno de mis más grandes anhelos. Es cierto que cuando se presenta el momento de decir gracias, hacen falta palabras que realmente siente el corazón, agradezco de todo corazón a Dios y a mis Padres porque a través de ellos me concedió la vida en este mundo, y porque gracias a su apoyo y consejos han guiado mi camino. Así como a mis abuelos que para mí han sido como mis padres y quienes me han formado moralmente como ser humano, quiero agradecerles tantos sacrificios, esfuerzos de su vida para educarme, el amor con el que han entregado su vida para convertirme en un hombre de provecho. A mi tía quien con su ayuda, apoyo y comprensión me ha alentado y motivado a buscar un mejor futuro y edificación de mi vida, por ese cariño que siempre he recibido en los momentos buenos y malos de mi vida, hago este triunfo compartido. A mi mejor amiga a quien jamás encontraré la forma de agradecer su apoyo, comprensión, confianza, sonrisas, espero que comprendas que muchos de mis logros han sido por la fortaleza de tu apoyo, y que mis ideales son inspirados en tu nobleza, a mis amigos quienes directa o indirectamente han sido inspiradores. Finalmente, pero no menos importantes a mi asesor y sinodales quienes se tomaron tiempo para revisar y guiarme en la escritura de esta tesis, a mis maestros por su dedicación para ayudarme a formarme en un ámbito profesional

# Índice general

1. Introducción	4
1.1. Planteamiento del problema . . . . .	6
1.2. Justificación . . . . .	7
1.3. Hipótesis . . . . .	7
2. Antecedentes	8
2.1. Aspectos teóricos . . . . .	11
2.2. Modelado de fondo . . . . .	11
2.3. Descriptores . . . . .	19
2.4. Clasificadores . . . . .	22
2.5. Distancias estadísticas . . . . .	25
3. Metodología	27
3.1. Segmentación . . . . .	27
3.2. Extracción de características . . . . .	29
3.3. Análisis de objetos . . . . .	35
4. Resultados y Discusión	36
4.1. Resultados de la segmentación . . . . .	36
4.2. Resultados de la Extracción de características . . . . .	45

5. Discusión	55
5.1. Discusión . . . . .	55
6. Conclusiones	61

## Índice de cuadros

4.1. Relación de experimentos en bases de datos. Fuente: Elaboración propia	40
4.2. Variabilidad de los valores singulares entre formas de una misma clase Fuente: Elaboración propia . . . . .	46
4.3. Variabilidad de los valores singulares entre formas de una misma clase aplicando rotaciones . . . . .	47
4.4. Variabilidad de los valores singulares entre formas de una misma clase con escalado gradual . . . . .	47
4.5. Error porcentual entre objetos de una misma clase <i>apple</i> – 1. Fuente: Elaboración propia . . . . .	49

## Índice de figuras

2.1. Clasificación de métodos de distancias estadísticas. Fuente: Elaboración propia . . . . .	25
4.1. Escenarios de la Universidad de Edimburgo . jMajecka2009 . . . . .	36
4.2. Escenarios PETS . Fuente: Ferryman y Shahrokni (2009) . . . . .	37
4.3. Escenarios CDNET Fuente: Y. Wang y col. (2014). . . . .	37
4.4. Niveles de Textura de la escena Fuente: Elaboración propia. . . . .	38
4.5. Cuantificación de la textura de la escena Fuente: Elaboración propia. .	39
4.6. Desviación típica de la textura caracterizada en la escena Fuente: Elaboración propia. . . . .	40
4.7. Segmentación de objetos en movimiento . . . . .	40
4.8. Ajuste del valor $\lambda$ para caracterización de la escena . . . . .	41
4.9. Ajuste del valor $\lambda$ para caracterización de las personas en la escena . .	42
4.10. Modelo de fondo en escenario A . . . . .	43
4.11. Modelo de fondo en escenario B . . . . .	44
4.12. Modelo de fondo en escenario C . . . . .	44
4.13. Modelo de fondo en escenario D . . . . .	45
4.14. Escenarios MPEG-7 Fuente: Paramarthalingam y T. T (2021). . . . .	45



4.15. Error porcentual de formas de la clase <i>apple</i> – 1 Fuente: Elaboración propia . . . . .	46
4.16. Error porcentual ante rotaciones de formas de la clase <i>apple</i> – 1 Fuente: Elaboración propia . . . . .	47
4.17. Error porcentual entre objetos de la clase <i>apple</i> – 1 aplicando escalado gradual Fuente: Elaboración propia . . . . .	48
4.18. Error porcentual entre objetos la clase <i>apple</i> – 1 Fuente: Elaboración propia . . . . .	48
4.19. Escenarios Stanford Drone Fuente: Robicquet y col. (2016). . . . .	49
4.20. Variaciones en la forma del objeto . . . . .	50
4.21. Representación de los valores singulares de los vectores de D . . . . .	51
4.22. Vectores singulares de los ejes principales del objeto . . . . .	51
4.23. Influencia de los valores característicos de objetos tipo personas . . . . .	52
4.24. Influencia de los valores característicos de objetos tipo automóviles . . . . .	52
4.25. Clasificación de objetos en movimiento . . . . .	54
5.1. Núcleo de la función de gabor. . . . .	55
5.2. Influencia del parámetro $\lambda$ en el filtro de Gabor. Fuente: Elaboración propia . . . . .	56
5.3. Selección adecuada del valor de $\sigma$ . Fuente: Elaboración propia . . . . .	57
5.4. Desfase del filtro de gabor. Fuente: Elaboración propia . . . . .	58
5.5. Desfase del filtro de Gabor. Fuente: Elaboración propia . . . . .	59

5.6. Textura periódica y no periódica de la escena . . . . . 60

## Abreviaturas

CART	Classification And Regression Trees
CDNET	Change Detection.Net
Drone	Término acuñado al vehículo aéreo no tripulado o VANT
DSTEI	Diference Spatio-Temporal Entropy Image
MPEG7	Moving Picture Experts Group 7

## Resumen

Este trabajo presenta el desarrollo de un algoritmo de descripción de forma que involucra la detección de condiciones atípicas de escenas estructuradas. Para ello, se evalúan objetos en movimiento a través de un modelo de segmentación de fondo basado en filtros Gabor y momentos estadísticos, que se utilizan para cuantificar la información de la escena. Los objetos en movimiento son segmentados y filtrados para obtener su contorno, el cual es evaluado a través de métricas que describen la forma del objeto, estas métricas extraen las características de la forma a través del ángulo con respecto al centroide, la curvatura o el ángulo tangencial. Estas características extraídas se asociarán con un conjunto de descriptores para analizar e inferir la dinámica de los objetos en movimiento. Los resultados obtenidos muestran una fuerte invariancia a transformaciones como escalado, rotación y traslación. Los resultados presentados fueron evaluados cualitativamente y se aplicaron métricas para medir los errores típicos.

**Palabras clave:** descriptores de forma, eigen valores, eigen vectores, segmentación, visión por computador.

## Abstract

This work introduces the development of a shape description algorithm that involves the detection of atypical conditions of structured scenes. To this end, moving objects are evaluated through a background segmentation model based on Gabor filters and statistical moments, which are used to quantify the information of the scene. The moving objects are segmented and filtered to obtain their contour, which is evaluated through metrics that describe the shape of the object, these metrics extract the shape features through the angle with respect to the centroid, the curvature, or the tangential angle. These extracted features will be associated with a set of descriptors to analyze and infer the dynamics of moving objects. The results obtained show a strong invariance to transformations such as scaling, rotation, and translation. The presented results were qualitatively evaluated and metrics were applied to measure typical errors.

**Keywords:** shape descriptors, eigen values, eigen vectors, segmentation, computer vision.

# Capítulo 1 Introducción

En los últimos años, investigadores de diversos campos de visión por computador han centrado sus esfuerzos en desarrollar algoritmos que sean capaces de emular la visión humana, para ello han desarrollado sistemas de visión artificiales, en los que mediante cámaras buscan reconocer patrones, realizar seguimiento, analizar trayectorias o inferir acciones. Dentro de estas actividades, los descriptores de forma juegan un rol importante para describir la topología del objeto y determinar sus características tanto generales como locales.

Un descriptor de forma es una herramienta efectiva para el reconocimiento, clasificación o identificación, debido a que la forma es inherente a la comprensión de una imagen, la cual, se conserva constante a pesar de los cambios en la iluminación, el color o la textura de un objeto. Debido a estas ventajas, las características de forma se han aplicado ampliamente en numerosos métodos como: aproximación de polígonos (Ramaiah y Prasad, 2021), características de interrelación espacial (Mohammed y Hamdi, 2018), momentos (C. Wang y col., 2020; G. Wu y Xu, 2019; Z. Wu y col., 2020), métodos de espacio-escala (Zhou y col., 2019) y transformaciones de forma (da Fontoura y Cesar, 2018). La idea de un descriptor de forma robusto sigue siendo un problema difícil, los desafíos que dificultan el reconocimiento de la forma son debido a transformaciones como rotación, escalamiento, deformaciones provocadas por ruido u oclusiones o efectos

---

no deseados provocados por las sombras.

Básicamente, los descriptores globales solo pueden identificar formas con grandes diferencias, debido a sus características simples. Mientras que, los descriptores locales se centran más en los aspectos de la forma para usarlos como discriminante. Una forma puede describirse por diferentes aspectos como: centro de gravedad, excentricidad, relación de circularidad, varianza elíptica, rectangularidad o relación de área, por mencionar alguno. En general, un descriptor intenta cuantificar la forma dada mediante un conjunto de números que se producen debido a sus aspectos característicos, por lo que un descriptor de forma robusto debe cumplir los siguientes requisitos:

- Preservación de similaridad entre un mismo tipo de formas.
- Invarianza a transformaciones de similitud como: traslación, rotación y escala; y transformaciones afines
- Resistencia al ruido
- Invariancia de ocultación
- Independencia estadística (compacidad)
- Confiabilidad de las características extraídas.

Las técnicas relacionadas a la representación de forma son amplias, y están clasificadas en métodos enfocados en contornos y en regiones. Los métodos con base en contornos están sujetos a la detección de características como puntos, bordes y líneas que se encuentran en los bordes de la forma. Mientras que los métodos basados en regiones obtienen una colección de vectores característicos de las regiones ocupadas por el objeto. El trabajo de esta tesis se centra en la extracción de elementos característicos de

objetos que se encuentran en movimiento, dichos elementos son clasificados mediante un algoritmo de agrupamiento no supervisado, para asociarlos a clases e identificar el tipo de objeto que se está observando.

## 1.1 Planteamiento del problema

Actualmente, el interés por los procesos de reconocimiento de imágenes tiene mucha trascendencia, a medida que la tecnología evoluciona, los datos percibidos por máquinas dedicadas a extraer información de imágenes, requieren de aplicaciones robustas de visión, las cuales presentan problemas en el instante de realizar el análisis e interpretación de la información. Estos problemas se deben al ruido, causado principalmente por las condiciones de iluminación, contraste, ruido inherente de la cámara, perturbaciones ambientales, cambio en la postura de objetos, entre otros. Muchos de los métodos utilizados se centran en la segmentación, análisis de forma, extracción de características, geometría y cuestiones físicas que proporcionen más información sobre las condiciones del objeto, no obstante, aún no existe un método general para reconocimiento de objetos en movimiento, por ejemplo, algunos autores detectan personas como objetos aislados, pero en escenarios reales existen grupos de personas o parejas, que requieren de un análisis más exhaustivo.



## 1.2 Justificación

Diversos investigadores han desarrollado algoritmos propios con la finalidad de mejorar los resultados en la descripción de escenas mediante características invariantes locales y que faciliten el análisis la información relevante para que sirva como base hacia el reconocimiento de objetos.

Derivado de esta investigación se evidencia que un descriptor de características representa una alternativa para llevar a cabo un análisis de forma o geometría y que son ampliamente usados para analizar la dinámica del movimiento del objeto, es decir, trazar trayectorias, seguimiento de personas, mejora de sistemas de vigilancia o conducción de vehículos, crear animaciones, modelado de objetos, entre otros.

## 1.3 Hipótesis

La dinámica de movimiento en un escenario estructurado tiene una entropía baja, por consecuencia, la dinámica atípica representa una pérdida de la estructura del escenario que, mediante la asociación de las características atípicas del escenario a un conjunto de descriptores, permitirá representar una ausencia en la estructura del escenario e inferir la dinámica de los objetos en movimiento.

## Capítulo 2 Antecedentes

La mayoría de los sistemas de visión son capaces de detectar movimiento, tienen como objetivo el reconocimiento y extracción de objetos en movimiento con la finalidad de reconocerlos, realizar seguimiento o para vigilancia. Las técnicas actuales establecen restricciones para evitar la influencia por cambios de iluminación, rotación, escala o efectos provocados por las sombras, que tienen un impacto negativo en el rendimiento del reconocimiento y aunque las técnicas proporcionan resultados alentadores, aún existen problemas como la variación de la velocidad al caminar, la perspectiva, la ropa, el calzado y el transporte de objetos.

En el análisis dinámico de la escena se busca cualquier cambio perceptible y que sea significativo para una aplicación específica, el cuál trata de una secuencia de imágenes en la que cada imagen representa la escena en un instante en particular, la cual generalmente contiene varios objetos. Los cambios que se producen en la escena pueden presentarse debido al movimiento de la cámara, el movimiento de objetos, cambios en la estructura, tamaño, forma de los objetos presentes o debido al ruido; el cual proviene de cambios de iluminación. Dicho análisis requiere de tres fases principales:

- **Explorar las regiones de interés.** En esta fase se extraída la información significativa de las regiones que requieren de un mayor análisis.
- **Fase de análisis.** En esta fase se concentra la información extraída y es procesada

---

para realizar reconocimiento de objetos, análisis de movimiento, seguimiento de trayectorias, segmentación, etc.

- **Fase de aplicación.** En esta fase todo el conocimiento cognitivo generado se aplica para determinar los objetos presentes en la escena y los eventos que en ella tienen lugar.

En la detección de movimiento, la topología de los objetos juega un rol importante para el desarrollo de algoritmos de descripción de escenas, ya que es posible encontrar características locales que faciliten el análisis de la información relevante y que sirva como base hacia el reconocimiento. Un claro ejemplo, es el esqueleto, el cual es una estructura que describe la topología y la geometría de la forma de un objeto, dicha representación permite una forma más simple y computacionalmente efectiva de analizar las propiedades de la forma, las técnicas actuales para análisis de taxonomías, cuentan con propiedades como:

Homotopía, Invarianza y Suavizado (Hu y col., 2020), el objeto puede representarse como una función continua, que es invariante a escala; reconstrucción y detalles (Bertrand y Couprie, 2017), el objeto puede ser reconstruido a partir de su estructura, encontrando los detalles principales de objeto, reduciendo la cantidad de procesamiento requerido para analizarlo. Dentro de este tipo de técnicas, (Bataineh, 2018) obtienen la taxonomía del objeto mediante un proceso iterativo para etiquetar los pixeles vecinos de la frontera que se forma entre el primer plano y el fondo de la imagen, los cuales son eliminados y re-etiquetados para generar una nueva frontera, hasta alcanzar el adelgazamiento a un pixel.

---

De igual manera, (Boudaoud y col., 2018), desarrollaron un algoritmo llamado ZS en el que combinan un enfoque de subdivisión de una imagen en dos conjuntos de coordenadas pares e impares. Mientras tanto (L. Yang y col., 2019), desarrollaron un algoritmo de adelgazamiento paralelo basado en el eje medio para extracción de contorno del cual suprimen el límite exterior hasta conseguir el adelgazamiento de 1 pixel de ancho.

Por su parte, (Ding y col., 2019), proponen un algoritmo de adelgazamiento mejorado para detección del objeto en movimiento, afirmando rapidez con respecto a los algoritmos evaluados en su investigación.

Aunque los esqueletos proporcionan una alternativa para reducir la complejidad del procesamiento y trabajar directamente con la topología del objeto, los métodos no son muy bien vistos debido a la sensibilidad a los detalles, los cuales pueden modificar la taxonomía del objeto, por lo que otras técnicas buscan localizar elementos característicos en la estructura, que definan a los objetos presentes bajo condiciones de variación y que se representen explícitamente las escalas en las que ocurren los diferentes eventos.

La complejidad de estas técnicas, se debe a que dichas condiciones de variación en la representación de la imagen en espacio-escala son causadas por cambios de iluminación, sombras, ruido inherente en la cámara y cambios de escala.

Upadhyay y Chandra (2019), extraen características de grandes estructuras mediante una familia de señales suavizadas, que representan los datos en múltiples escalas, mediante la solución a una ecuación de difusión definida como  $\partial_t L = 0,5 \times \nabla^2 L$ , donde L representa la convolución entre una imagen y un operador gaussiano, con el propósito

de construir una representación de espacio invariante a la escala de la estructura de la imagen.

Por otra parte, Barnouti y col. (2018) analizan la correlación entre las condiciones típicas en el escenario y los patrones de la entropía de una secuencia de imágenes, para proporcionar una abstracción de un gran objeto binario o blob (por sus siglas en inglés, binary large object) de espacio-escala con un blob de nivel de grises en un único nivel de escala, es decir, dicha abstracción implica una proyección desde un blob de espacio de escala de cuatro dimensiones a un blob de nivel de gris tridimensional y una proyección del blob tridimensional a su región de soporte bidimensional.

Para poder analizar la dinámica de los objetos que se encuentran en movimiento, es necesario segmentarlos de la escena. Comúnmente una escena es estacionaria o casi estacionaria, y los objetos que se encuentran en movimiento modifican la topología de su superficie. Por lo cual, un modelo de fondo busca caracterizar los cambios lumínicos y cuantificarlos para discernir entre un objeto estacionario y uno no estacionario.

## 2.1 Aspectos teóricos

## 2.2 Modelado de fondo

La sustracción de fondo emplea técnicas que permiten determinar que secciones de la imagen se mantienen constantes en el tiempo en una escena y sustraer los objetos en primer plano, es decir los objetos que se encuentran en movimiento. Existen múltiples

métodos para modelado de fondo, que se clasifican en paramétricos (S. Zhu y col., 2018) y no paramétricos (Berjón y col., 2018), los modelos paramétricos se caracterizan por estimar un conjunto de parámetros basados en la distribución de color de las imágenes y los no paramétricos se introduce una función matemática para modelar la distribución.

El uso de estos métodos depende de la aplicación ya que la escena se ve afectada por cambios de intensidad, sombras, perspectiva, etc. En esta sección se describirán algunas técnicas de análisis de sustracción de fondo más comunes basadas en diferencia, mezcla de distribuciones Gaussianas y aproximaciones probabilísticas.

### 2.2.0.1 Resta de fondo auto-organizada

En este modelo es creado un mapa neural de  $n$  imágenes  $I$  y cada pixel de la escena es evaluado para adaptarlo a la mejor coincidencia.

$$X_M = \{I_1^{i,j}, \dots, I_n^{i,j}\} \quad (2.1)$$

Cada peso es obtenido, encontrando la mínima distancia entre la intensidad del color actual y la del mapa neural (Lu y Ma, 2020; Maddalena y Petrosino, 2019).

$$B_M = \min(d(X_{n-1} - It)) \quad (2.2)$$

Para discernir los objetos que pertenecen al fondo y al primer plano, la distancia

obtenida es comparada contra un umbral.

$$M_B = \begin{cases} 1 & \text{si } FCN(B_M) \leq 0,5 \\ 0 & \text{en otro caso} \end{cases} \quad (2.3)$$

La ventaja de este modelo, es la adaptación a variaciones graduales de iluminación o fondos dinámicos. Sin embargo, las sombras que llegan a proyectar los objetos son detectadas e incluidas en el modelo de fondo reconstruido.

### 2.2.0.2 Diferencia de imágenes

Este método consiste en obtener la diferencia entre dos imágenes consecutivas y obtener una matriz de distancias, cada elemento de la matriz es comparado contra un umbral  $T_h$  para detectar el movimiento en primer plano (Zeng y M. Zhu, 2018).

$$T_h < |I_k - I_{k+1}| \quad (2.4)$$

### 2.2.0.3 Distribución de probabilidad

El método DSTEI empleado en Bouwmans y col. (2018) y Gao y col. (2020). Observa una distribución de probabilidad de una secuencia de imágenes para cada estado de píxel a lo largo del tiempo. A partir de estas observaciones crea un histograma de tiempo para presentar la distribución del estado en una ventana deslizante. La función de densidad de probabilidad (pdf, probability density function) del estado de los píxeles

se obtiene normalizando el histograma de acuerdo con la [Ecuación \(2.5\)](#)

$$P_{i,j,t} = \frac{H_{i,j,t}}{N} \quad (2.5)$$

Donde  $N$  es el número total de píxeles en el histograma y  $\sum_{t=1}^T P_{i,j,t} = 1$ . Para calcular la variabilidad en la distribución del estado del píxel se calcula la entropía utilizando la [Ecuación \(2.6\)](#).

$$E_{i,j} = -\sum_{t=1}^T P_{i,j,t} \log(P_{i,j,t}) \quad (2.6)$$

Donde  $E_{i,j}$  es la entropía espacio-temporal del píxel (i,j). De esta forma es cuantificado cada nivel de gris en 256 niveles para formar una imagen de energía.

#### 2.2.0.4 Aproximaciones probabilísticas

En este método se estima el fondo obteniendo el promedio de cada píxel en imágenes que se han ido almacenando (García-González y col., 2019). Este método es bastante rápido, pero consume mucha memoria la cual se determina en base a la siguiente [Ecuación 2.7](#)

$$memoria = n * tamao(imagen) \quad (2.7)$$



Para determinar el promediado del fondo se aplica la [Ecuación \(2.8\)](#).

$$B_{(i+1)} = \alpha \cdot F_i + (1 - \alpha) \cdot B_i \quad (2.8)$$

Donde  $B$  es la imagen de fondo,  $F$  la imagen de primer plano,  $\alpha$  la velocidad de aprendizaje (0.05) e  $i$  denota la iteración anterior.

### 2.2.0.5 Mezcla de Gaussianas

Es un algoritmo que representa la distribución de probabilidad de un pixel de ser observado y parte de una estimación  $\theta$  hasta detectar la convergencia, en cada iteración se modela cada pixel con una gaussiana, del cual se determina la probabilidad máxima para calcular los parámetros del modelo:

$$P(X_i) = \sum_{i=1}^k \omega_{i,j} \cdot \eta(X_i \cdot \mu_{i,t}, \Sigma_{i,j}) \quad (2.9)$$

Donde  $P$  es la probabilidad de observación,  $\omega$  el peso asociado a la  $i$  gaussiana. Cada píxel del modelo es actualizado continuamente y entrenado, al estar actualizado deberán ser normalizados los componentes gaussianos (Song y J.-h. Kim, 2019).

$$\omega_{i,j} = (1 - \alpha) \cdot \omega_{i,j} + \alpha \mu_{i,j} = (1 - \rho) \cdot \mu_{i,j} + \rho \quad (2.10)$$

Donde  $\alpha$ ,  $\rho$ , son factores para aumentar o disminuir las velocidades de aprendizaje y  $\Sigma$  es la matriz de covarianza.

### 2.2.0.6 Banco de filtros de Gabor

Este método emplea los filtros para el análisis de bordes y texturas como un modelo espacial y los momentos de orden  $r$  como un modelo temporal. Este método es efectivo para caracterizar regiones casi homogéneas, debido a que el filtro de Gabor está estructurado matemáticamente de tal manera que puede cuidar diferentes formas, tamaños y niveles de suavidad en la imagen.

El banco de filtros de Gabor es una de las funciones que permite caracterizar la densidad (J. Kim y col., 2018), son sensibles a la orientación y se utilizan para el análisis de bordes y texturas (Moreyra2014), el cual es propio para regiones que presentan cambios bruscos o sutiles de intensidad. La función de Gabor 2-D se compone de una función gaussiana llamada envolvente y una función senoidal y cosenoidal llamada portadora (ver Ecuación (2.11)).

$$G_{\lambda,\theta,\phi} = e^{j\omega} \cdot \eta \cdot \exp\left[\left(-\frac{u^2}{2\sigma_x^2} - \frac{v^2}{2\sigma_y^2}\right)\right] \quad (2.11)$$

Donde  $\eta = \frac{1}{2\pi\lambda\sigma_x\sigma_y}$ ,  $\omega = 2\pi\frac{u}{\lambda} + \phi$ , los términos  $\sigma_x$  y  $\sigma_y$  definen la amplitud de la distribución gaussiana, el parámetro  $\lambda$  aumenta o disminuye la frecuencia del filtro,  $\phi$  representa el desfase del término  $e^{j\omega}$  y la tupla  $(u, v)$  son las coordenadas cartesianas

de la frecuencia espacial.

$$u = x \cos(\theta) + y \sin(\theta) \quad (2.12)$$

$$v = y \sin(\theta) - x \cos(\theta)$$

Ahora bien, la transformación de Gabor se muestra en la [Ecuación \(2.13\)](#), la cual es obtenida a partir de la convolución entre la imagen y la función de Gabor.

$$\Upsilon = I_i(x) * G_{\lambda, \theta, \phi}. \quad (2.13)$$

La magnitud de  $M_i$  y  $P_i$  de la función  $\Upsilon$  se expresan en las Ecuaciones (2.14) y (2.15) respectivamente, las cuales representan la parte real ( $\Upsilon_r$ ) y la parte imaginaria del filtro ( $\Upsilon_c$ ).

$$M_i = \sqrt{\Upsilon_r^2 + \Upsilon_c^2} \quad (2.14)$$

$$P_i = \tan^{-1} \left( \frac{\Upsilon_c}{\Upsilon_r} \right) \quad (2.15)$$

Tanto  $M_i$  como  $P_i$ , representan el modelo espacial para la estructura y la textura de la superficie de la escena. Mientras que el modelo de fondo temporal es cuantificado

con momentos estadísticos, [Ecuación \(2.16\)](#).

$$m_r = \sum_{i=1}^N (x_i - \bar{x})^r \cdot P_b(I_i) \quad (2.16)$$

Donde:

- $N$  representa el número total de elementos,
- $x_i$  los valores de la muestra,
- $\bar{x}$  representa la media aritmética de  $x_i$
- $I_i$  denota la intensidad de gris,
- $P_b(I_i)$  es la probabilidad de  $I_i$  y
- $r$  es el momento  $r$  - *simo*

De acuerdo con de la Puente Viedma (2018), los momentos estadísticos representan la probabilidad esperada (cuando  $r = 1$ ), la varianza (cuando  $r = 2$ ), el sesgo (cuando  $r = 3$ ) y la curtosis (cuando  $r = 4$ .)

De acuerdo con este método, para discernir entre objetos en primer plano y los objetos estacionarios es empleado el siguiente umbral:

$$M_B = \begin{cases} B_k & \text{si } m_r \leq \pm k\sigma \\ O_k & \text{en otro caso} \end{cases} \quad (2.17)$$

La segmentación de objetos es una de las partes importantes en el análisis de la dinámica de imágenes, ya que permiten recuperar casi la estructura completa de los objetos para que posteriormente puedan ser descritos de acuerdo con sus característi-

cas. Las características representan un conjunto de elementos de líneas y puntos que presentan similitud en el tiempo. De acuerdo con Rani y col. (2016), la extracción de las características de un objeto es un de las principales etapas para el análisis de escena, reconocimiento de objetos o análisis de su dinámica, pero requieren de un descriptor que permita determinar los invariantes locales.

## 2.3 Descriptores

Un descriptor es un algoritmo utilizado para extraer las características de las imágenes con el fin de diferenciar un tipo de objeto de otro, describir la estructura del objeto, proporcionar una representación del objeto, analizar el movimiento o localización del objeto entre otros. De acuerdo a su aplicación los descriptores se clasifican en; descriptores globales o de forma, de textura, de color o de características. En esta sección se describirán algunos descriptores de forma y de características.

### 2.3.0.1 Descriptores globales o de forma

Un descriptor de forma busca encontrar similitudes entre imágenes y que sean invariantes a rotación, traslación y escala. Las características obtenidas de este tipo de descriptores contemplan firmas de forma, momentos invariantes, matriz de forma, espectrales, etc.

### 2.3.0.2 Descriptores de características

Un descriptor de características busca que puntos como bordes, esquinas o cambios en la estructura de la imagen sean invariantes a rotación, escala o efectos de iluminación, Díaz (2021), dividen los principales métodos para extracción de características en cuatro grupos principales, basados en detectores generales de: bordes, esquinas, de objetos binarios y de regiones. De los cuales solo se describirán los detectores de esquinas de Harris, Laplaciano de Gaussianas y Harris-Laplace.

### 2.3.0.3 Detector de esquinas de Harris

Este algoritmo ha sido desarrollado por Harris y Stephens y ha sido implementado para muchas aplicaciones (Vázquez-Esquivel y col., 2021), consiste en detectar la intersección entre dos ejes mediante una matriz de correlación, la cual describe el cambio de intensidad entre los vecinos de un pixel en específico. Dicha matriz está compuesta por los gradientes horizontales y verticales de la imagen que se está analizando.

$$H = \sum_{xy} w(x, y) \begin{bmatrix} I_x^2 & I_{xy} \\ I_{xy} & I_y^2 \end{bmatrix} \quad (2.18)$$

Donde  $w$  es una función de ventana e  $I$  la intensidad del píxel en  $(x, y)$ , este método es invariante a rotación y translación

### 2.3.0.4 Laplaciano de Gauss

El filtro LOG es un operador que encuentra el cambio de intensidad en todas direcciones, en este método se combina la aplicación de un filtro gaussiano para disminuir la sensibilidad al ruido a la hora y un filtro laplaciano, esto ayuda a detectar los puntos de interés en una imagen destacando los bordes (Hajian y Ramli, 2018).

$$L(x, y, \sigma) = f(x, y) * \left( \frac{1}{\sqrt{2\pi\sigma^2}} e^{\left(-\frac{x^2+y^2}{2\sigma^2}\right)} \right) \quad (2.19)$$

### 2.3.0.5 Detector de Harris-Laplace

Este algoritmo, busca las esquinas en imágenes de múltiples escalas y usa el operador de Laplace para detectar la escala adecuada que ayuda a medir la similitud entre el operador de Harris y la imagen. (Hung y Yen, 2019)

$$L = \det \left( \begin{bmatrix} L_x^2(x, \sigma) & L_{xy}(x, \sigma) \\ I_{xy}(x, \sigma) & I_y^2(x, \sigma) \end{bmatrix} \right) - k * \text{Tr}^2 \left( \begin{bmatrix} L_x^2(x, \sigma) & L_{xy}(x, \sigma) \\ I_{xy}(x, \sigma) & I_y^2(x, \sigma) \end{bmatrix} \right) \quad (2.20)$$

El conjunto de elementos característicos que describe a cada objeto en la escena, debe ser organizado en estructuras, donde el nivel de abstracción nos permita explicar la información. Por lo que se requiere de un método de agrupación que asocie los objetos que comparten características homogéneas.

## 2.4 Clasificadores

La clasificación es un procedimiento que permite asociar un conjunto de píxeles a categorías específicas; de acuerdo a su valor de color (Aggarwal y Reddy, 2018). Existen distintos métodos para discriminar objetos y clasificarlos, a continuación, se mencionarán algunos.

### 2.4.1 Clasificador $K - means$

Mediante este método se determinan las medias de  $n$  clases aleatorias de manera iterativa, se establecen  $k$  centroides aleatorios y se calcula la mínima distancia entre los datos evaluados y los centroides de las clases existentes más cercanas  $C = C_1, C_2, \dots, C_k$  de acuerdo a la Ecuación (2.21).

$$c = \min \sum_{i=1}^k \sum_{x_j \in C_i} \|x_j - \mu_i\|^2 \quad (2.21)$$

Por cada iteración del algoritmo, se recalculan nuevas medias para reclasificar la información y calcular un nuevo centroe de la clase, hasta alcanzar la convergencia de la información (Sinaga y M.-S. Yang, 2020)

### 2.4.2 Por máxima probabilidad

Este clasificador se basa en el cálculo de la probabilidad de distintas clases en la escena que expresan alguna incertidumbre (probabilidad a priori) antes de tomar en



cuenta la probabilidad de determinada clase. Estos clasificadores de máxima probabilidad derivan en el término de Bayes (Wade y Ghahramani, 2018)

$$p(\omega_i|x) = \frac{p(x|\omega_i)p(\omega_i)}{p(x)} \quad (2.22)$$

### 2.4.3 Clasificador de árbol de decisión

Un árbol de decisión, comúnmente es utilizado para asociar objetos a conjuntos de clases predefinidas, acordes a sus atributos, lo cuales pertenecen a estructuras de semi-orden. Principalmente consta de:

- Nodos, los cuales forman un árbol dirigido con un solo borde entrante y que etiquetan atributos.
- Ramificaciones, son conectores que establecen las relaciones entre los nodos y se etiquetan con el posible atributo.
- Hojas, son nodos terminales los cuales son asignados a una clase que representa su valor más apropiado.

Algunos algoritmos más importantes de árboles de decisión se listan a continuación:

ID3: Este algoritmo aprende usando la diferencia de información de los datos para analizar y amplificar la información obtenida, la cual usa como criterio de división. El algoritmo converge cuando la mejor ganancia de información no es mayor que cero. ID3 no aplica ninguna poda o procesamiento para atributos numéricos o valores faltantes. Su principal ventaja es la simplicidad, pero no se garantiza la solución perfecta, es posible

que converja a un óptimo local, para evitarlo se puede utilizar retro-propagación durante la búsqueda. Aunque este algoritmo produce árboles pequeños, no siempre produce el árbol más pequeño posible (Gupta y col., 2019).

C4.5: Una evolución del algoritmo ID3, es el C4.5, a diferencia del ID3, este algoritmo usa la tasa de ganancia como criterio del cociente. La convergencia se alcanzará cuando el número de instancias baje de cierto umbral. La poda de este algoritmo se lleva a cabo en la etapa de crecimiento, la cual maneja propiedades digitales y son llevadas a cabo para eliminar ramas que no contribuyen a la fidelidad y son reemplazadas con nudos de hojas. Ocupa rasgos continuos dividiendo el rango de valores en dos subconjuntos, buscando el mejor umbral para maximizar el criterio de ganancia. Los valores que se encuentren más grandes que el umbral forman el primer subconjunto y los valores por debajo del umbral forman el otro subconjunto de datos (Gupta y col., 2019).

CART: es un algoritmo con base en árboles de clasificación y regresión. En este tipo de árbol, cada nodo cuenta con dos aristas salientes, y son usados los criterios de Twoing, este tipo de árbol proporciona una distribución de probabilidad previa. Como característica importante, los árboles de regresión predicen un número real y no una clase, por lo que se busca minimizar el error cuadrático de predicción. (Zacharis, 2018)

## 2.5 Distancias estadísticas

Las distancias estadísticas permiten representar geoméricamente a un objeto como puntos en un espacio métrico adecuado, estas se clasifican de acuerdo a sus variables cuantitativas, binarias, categóricas o mixtas. En el cuadro sinóptico mostrado en la [Figura 2.1](#) se muestra la clasificación y los principales métodos desarrollados en la literatura.



[Figura 2.1](#). Clasificación de métodos de distancias estadísticas. Fuente: Elaboración propia

La distancia estadística en un conjunto de elementos  $D$ , es definida como la distancia mínima de longitud  $L$  entre pixeles  $p$  y  $q$ , tales que  $p$  y  $q$  son subconjunto de

$P = (p_1, p_2, \dots, p_l)$  en  $D$ . Por lo que se define como:

$$\delta_{p,q} = \min\{L(P) | p_1 = p, p_l = q \wedge P \subseteq\} \quad (2.23)$$

### 2.5.0.1 Propiedades de las distancias estadísticas

Gil Pascual (2020) define una distancia como una aplicación en un conjunto  $\in$ , el cual cuantifica la relación entre variables. Cada tupla  $(i, j)$  corresponde con un número real  $\delta(i, j) = \delta_{i,j}$  cumpliendo con las siguientes propiedades definidas por Cuadras:

- $\delta_{ij} \geq 0$
- $\delta_{ij} = 0$
- $\delta_{ij} = \delta_{ji}$
- $\delta_{ij} \leq \delta_{ik} \cdot \delta_{jk}$
- $\delta_{ij} = 0$  si y solo si  $i = j$
- $\delta_{ij} = \max\{\delta_{ik}, \delta_{jk}\}$  (desigualdad ultra métrica)
- $\delta_{ij} + \delta_{kl} = \max\{\delta_{ik} + \delta_{jl}, \delta_{jl} + \delta_{ik}\}$  (desigualdad aditiva)
- $\delta_{ij}$  es euclídeana
- $\delta_{ij}$  es riemanniana
- $\delta_{ij}$  es una divergencia

## Capítulo 3 Metodología

La metodología para este proyecto se basa en la propuesta por Grauman y Leibe (2011), las etapas de la metodología son las siguientes:

### 3.1 Segmentación

Considere la Ecuación (3.1) (1), como un conjunto de imágenes en  $R^2$ .

$$S^k = s_1 \times s_2 \times \dots \times s_k \quad (3.1)$$

Donde  $S = \{x | x \in M_{(m \times n)}^k\}$ , y cada imagen representa un instante en el tiempo  $t$ , por lo que si  $I \in S$ , entonces:

$$I(x) = \{(x, y) | (x, y) \in N\}, I(x) \in M_{m \times n} \quad (3.2)$$

Ahora bien, se genera una transformación  $I_G$  sobre  $I$ , definida como  $M_{m \times n} \times M_{m \times n} \times M_{m \times n} \rightarrow M_{m \times n}$  de manera que  $(A, B, C) \rightarrow D$  donde:

$$d_{ij} = 0,2989 \cdot a_{ij} + 0,5870 \cdot b_{ij} + 0,1140 \cdot c_{ij} \quad (3.3)$$

Se define la operación  $\cdot$  como  $k \times M_{m \times n} \rightarrow M_{m \times n}$  de manera que  $(k, A) \rightarrow B$

donde:  $b_{ij} = k \cdot a_{ij}$  para toda  $A \in M_{m \times n}$  y  $k \in N$ .

El modelo de fondo es generado partir de la caracterización de textura de los  $t$  fotogramas, modelando la intensidad del pixel mediante la [Ecuación \(3.4\)](#).

$$m_r = \sum_{i=1}^N (x_i - \bar{x})^r \cdot P_b(I_i) \quad (3.4)$$

El segundo momento estadístico es utilizado ya que proporciona una media de suavidad de la región. El fondo en el instante de tiempo  $t$  para un cuadro  $x$  en particular (indicado como  $x_i$ ) se calcula de la siguiente manera  $F = X - \bar{X}$ ,  $\bar{X}$  es el promedio de las  $n$  distribuciones  $M_i$  y  $P_i$  (ver [Ecuaciones \(2.14\)](#) y [\(2.15\)](#)) y  $X$  es la distribución más reciente del mapa de textura. Para discernir entre objetos estacionarios o en movimiento, se establece un criterio de umbralización  $M_B$ :

$$M_B = \begin{cases} B_k & \text{si } m_r \leq \pm k\sigma \\ O_k & \text{en otro caso} \end{cases} \quad (3.5)$$

en donde  $M_B = \{x | x \in M^k, x = 0, 1\}$  para el conjunto de imágenes  $M^k = M_1 \times M_2 \times M_k$ . Considere que  $\cap(A : A \in M_B) = \emptyset$ . Entonces, para encontrar el borde de la imagen se aplica la [Ecuación \(3.6\)](#)

$$I_B = \delta_B - A \quad (3.6)$$

Donde  $\delta_B = \{x | (B)_x \cap A \neq \emptyset\}$ . Se define la operación  $-$  como  $M_{m \times n}^k \times M_{m \times n}^k \rightarrow$

$M_{m \times n}^k$  de manera que  $(A, B) \rightarrow C$ , donde  $c_{ij} = a_{ij} - b_{ij}$  para toda  $A \in M_{m \times n}^k$  y  $B \in M_{m \times n}^k$ .

## 3.2 Extracción de características

Considere el resultado de la [Ecuación \(3.6\)](#) como un conjunto  $r$  de puntos en el espacio tal que  $r \in I_B$ ,  $r$  es la imagen obtenida de la aplicación topológica de un segmento abierto de la recta  $[a, b]$ . Y sea  $a < t < b$  el segmento obtenido de la aplicación de una función, debido a que las coordenadas  $x$  e  $y$  de la posición de cada punto en el conjunto  $r$  dependen del punto  $t$  del segmento, por lo que se definen las funciones  $\gamma_1$  y  $\gamma_2$  del parámetro  $t$ , tal que  $x = \gamma_1(t)$  e  $y = \gamma_2(t)$ . Entonces:

$$r(t) = \gamma_1(t)i + \gamma_2(t)j \quad (3.7)$$

La [Ecuación \(3.7\)](#) representa una curva regular cerrada  $r$  en  $R^2$ , es decir  $k$  veces diferenciable.

Para la extracción de características se calculó el ángulo con respecto al centroide, el ángulo tangencial y la curvatura de los objetos.

### 3.2.1 Ángulo con respecto al centroide

Sea  $C$  el centro de la curva  $r$  y sean  $a$  y  $b$  dos puntos cuales quiera, donde  $a = \{x|x \in r\}$  y  $b = \{x|x \in r\}$ , el segmento de recta  $(\bar{C}a)$  se usa para determinar los  $n$  ángulos  $\theta$  con respecto al segmento de recta  $\bar{C}b$ , los cuales corresponden al contorno

de la forma. El ángulo se obtiene de acuerdo con el producto punto entre vectores de acuerdo con la Ecuación (3.8).

$$\bar{C}b \cdot \bar{C}a = \|\bar{C}b\| \|\bar{C}a\| \cos \theta \quad (3.8)$$

Por lo que al despejar el ángulo  $\theta$ , se obtiene:

$$\theta = \left( \frac{\bar{C}b \cdot \bar{C}a}{\|\bar{C}b\| \|\bar{C}a\|} \right) \quad (3.9)$$

### 3.2.2 Ángulo tangencial

Sea  $r$  la curva regular de la Ecuación (3.7),  $P$  un punto de la misma, el parámetro  $t$  el valor del punto  $P$  y  $u$  una recta tangente que pasa por el punto  $P$ . Suponga que se toma un punto  $Q$  de la curva, del cual es calculada su distancia  $d$  y  $h$  hasta el punto  $P$  y la recta  $u$ . Como la recta  $u$  es tangente a la curva  $r$  en el punto  $P$ , entonces:

$$\frac{h}{d} \rightarrow 0 \text{ cuando } Q \rightarrow P \quad (3.10)$$

Por lo cual, si la curva  $r$  tiene tangente en el punto  $P$ , la recta  $\bar{P}Q$  tiende a la tangente cuando  $Q \rightarrow P$ . Por lo que, la tangente en el punto  $P$  corresponde al valor  $t$  del parámetro y tiene la dirección del vector  $r'(t)$ . Entonces el vector unitario tangente



a la curva es definido como:

$$\vec{T} = \frac{r'(t)}{\|r'(t)\|} \quad (3.11)$$

El vector tangente marcará la dirección en la que varía la curva  $r$  en cualquier punto  $P$ . Ahora bien, considere los vectores  $T_i$  y  $T_{i+1}$  no nulos de la curva cerrada  $r$ , para encontrar un máximo o mínimo local que represente una característica de la curvatura, se calcula el ángulo  $\theta$  entre ambos vectores con la [Ecuación](#) (3.12).

$$\theta = \arccos \left( \frac{\vec{T}_i \cdot \vec{T}_{i+1}}{\|\vec{T}_i\| \cdot \|\vec{T}_{i+1}\|} \right) \quad (3.12)$$

### 3.2.3 Curvatura

Sea  $r$  una curva cerrada en  $R^2$ , en la que  $r(t) = x(t)i + y(t)j$ . Sea  $a = x|(x, y) \in r$ ,  $b = x|(x, y) \in r$  y  $c = x|(x, y) \in r$  puntos tangentes a la forma del objeto, con los cuales se construye una circunferencia osculatriz. Considerando que una circunferencia con de radio  $r$  y centro en  $(h, k)$ , se define la circunferencia mediante la [Ecuación](#) (3.13).

$$(x - h)^2 + (y - k)^2 = r_c^2 \quad (3.13)$$

Al expandir el binomio se obtiene

$$x^2 + y^2 - 2hx - 2hy + h^2 + k^2 = r_c^2 \quad (3.14)$$

Entonces la ecuación general de la circunferencia osculatriz es el sistema de la Ecuación (3.15).

$$x^2 + y^2 + Ax + By + C = 0 \quad (3.15)$$

Donde:

$$\begin{aligned} A &= -2h \\ B &= -2k \\ C &= h^2 + k^2 - r_c^2 \end{aligned} \quad (3.16)$$

Dado que se conocen los puntos  $a, b$  y  $c$ , estos son sustituidos en la Ecuación (3.15) para resolver el sistema y encontrar las variables  $A, B$  y  $C$ . Con las cuales se determina el centro  $(h, k)$  y el radio  $r_c$  de la circunferencia osculatriz. Para determinar un candidato a característica, se evalúan los puntos  $a, b$  y  $c$ . Mientras más cercanos sean estos puntos a  $P$  el radio de curvatura ( $r_c$ ) será más pequeño, y la curvatura será cada vez más grande, este procedimiento se realiza hasta alcanzar la convergencia de  $C$  en  $P$ , es decir el punto máximo de la curvatura.

De acuerdo con Pogolerov (1994), la curvatura de una curva plana es definida por la Ecuación (3.17)

$$K_1^2 = \frac{r_2''}{(1 + r_2'^2)^2} \quad (3.17)$$

Como el radio de curvatura es inversamente proporcional al valor absoluto de la

curvatura, entonces:

$$\left(\frac{1}{r_c}\right)^2 = \frac{r_2''}{(1 + r'^2)^2} \quad (3.18)$$

Finalmente, el radio de curvatura se obtiene despejando  $r_c$  de la [Ecuación \(3.19\)](#).

$$r_c = \sqrt{\frac{(1 + r'^2)^{\frac{3}{2}}}{r''}} \quad (3.19)$$

Por lo que  $r_c$  será:

$$r_c = \frac{(1 + r'^2)^{\frac{3}{2}}}{r''} \quad (3.20)$$

Debido a que algunos contornos de los bordes pueden producir variaciones con respecto a la forma real, se emplea un umbral de ángulo  $\alpha$ . Si el ángulo estimado  $\angle(aT_1 - aT_2) = \beta$  es mayor que el umbral de ángulo  $\alpha$ , la característica candidata se elimina del conjunto de características. Para el cálculo del ángulo  $\beta$ , se realiza usando dos tangentes  $aT_1$  y  $aT_2$  en ambos lados de la característica  $P$ . Si los tres puntos son colineales, el radio de curvatura es grande por lo que esa característica será descartada.

Para fines de esta tesis, se ha explorado principalmente la medida del ángulo tangencial, pero no se descarta el uso de las medidas de ángulo con respecto al centroide o curvatura ya que sus resultados son tan prometedores como la medida usada.

Todos los valores obtenidos de la [Ecuación \(3.20\)](#) son asociados al conjunto  $M$  para calcular una matriz de disimilitud, con la cual se busca encontrar la mayor diferencia

entre las características obtenidas, mientras mayor sea el ángulo, mayor será el valor de la medida de disimilitud. Esta relación está dada por la [Ecuación \(3.21\)](#).

$$D_{n \times n} = \{x | x \in R^2, x = \theta_k - \theta_k^T\} \quad (3.21)$$

donde  $D$  es la medida de disimilitud para cada  $r \subset I_B$ .

Ahora bien, observe que  $D$  es una matriz simétrica de  $n \times n$  y puede ser factorizada en:

$$D_{n \times n} = U \times \Sigma \times V \text{ con } U \in R^{m \times m}, \Sigma \in R^{m \times n} \text{ y } V \in R^{n \times n} \quad (3.22)$$

Donde:

$$\Sigma = \{\sigma_1, \sigma_2, \dots, \sigma_n | n \in N, \sigma_n = |DD^T - \sigma I|\} \quad (3.23)$$

$$U = DD^T - \sigma_i X \quad (3.24)$$

$$DV = U\Sigma \quad (3.25)$$

Por consiguiente, la [Ecuación \(3.22\)](#) podría descomponerse en  $k$  términos como se

expresa en la [Ecuación \(3.26\)](#).

$$A_k = U_1 \Sigma_1 V_1 + U_2 \Sigma_2 V_2 + \cdots + U_k \Sigma_k V_k \quad (3.26)$$

Mediante esta factorización es posible exponer mejor las diversas relaciones entre los elementos de datos que presentan las mayores variaciones de los cambios de la forma.

### 3.3 Análisis de objetos

Finalmente, se determinan  $n$  clases aleatorias mediante la [Ecuación \(24\)](#), las cuales representan las formas de los objetos de acuerdo con los vectores propios de la [Ecuación \(3.27\)](#).

$$c = \min \sum_{i=1}^k \sum_{x_j \in C_i} \|x_j - \mu_i\|^2 \quad (3.27)$$

El algoritmo se describe en los siguientes pasos:

- Inicializar  $k$  grupos, estableciendo  $k$  centroides aleatorios.
- Asignar los vectores propios  $U\Sigma$  al grupo con el centroide más cercano.
- Calcular el nuevo centroide de cada grupo, tomando como nuevo centroide la posición de la media de los datos pertenecientes a los  $k$  grupos.
- Repetir los pasos de asignación de datos a grupos y cálculo de nuevos centroides hasta que se alcance la convergencia.

# Capítulo 4 Resultados y Discusión

## 4.1 Resultados de la segmentación

En esta etapa se procesaron imágenes de distintas bases de datos, para validar el algoritmo de modelo de fondo, los retos que presentan se describen a continuación.

**Majecka dataset:** consiste en un conjunto objetos identificados como personas que caminan por el patio de un edificio en la Universidad de Edimburgo. su escena consiste en una cámara fija que puede ver la planta baja. De acuerdo con Majecka (2009), la característica más destacable es la iluminación en constante cambio debido a la posición del sol, las fuentes de luz artificial y las sombras proyectadas por algunos edificios.



Figura 4.1. Escenarios de la Universidad de Edimburgo . jMajecka2009

**PETS dataset:** fue usada para un análisis más detallado del algoritmo, esta escena incluye personas dispersas que caminan al azar con chaquetas brillantes y oscuras con texturas uniformes y desiguales, las imágenes son de  $720 \times 560$  píxeles (Ferryman y Shahrokni, 2009).

**CDNET dataset:** Las escenas de esta base de datos involucran a personas en una



Figura 4.2. Escenarios PETS . Fuente: Ferryman y Shahrokni (2009)

estación de tren, con efectos de sombras y reflejos debido a las condiciones de iluminación; escenas de tráfico vehicular con árboles que generan fondos dinámicos debido al movimiento de las hojas y cambios de perspectiva de los vehículos; en otra escena se ve a un hombre que entra a una oficina y lee un libro, esta escena está sujeta a cambios de luz y la intensidad del color de la ropa en relación con el fondo son similares; finalmente, la última escena se compone de algunas personas caminando o en bicicleta en un parque. El desafío de esta escena se debe a la iluminación excesiva y el muestreo insuficiente de la secuencia (Y. Wang y col., 2014).



Figura 4.3. Escenarios CDNET Fuente: Y. Wang y col. (2014).

Para caracterizar los cambios lumínicos y cuantificar la textura de los objetos a distintos niveles de profundidad de los *dataset*, es usado el modelo de fondo Gabor-Moments Background Subtraction Model (GMBSM) y con ello segmentar los objetos que se encuentran en movimiento en la escena. Mediante este modelo se asume que cada

región de la escena tiene un nivel de intensidad o de iluminación, esta variabilidad de la intensidad es considerada como textura. Por lo que los cambios de luz de una secuencia de imágenes  $I_t = I_1, I_2, I_3, \dots, I_n$ , son caracterizados y cuantificados de acuerdo con la Ecuación (2.13) en mapas de niveles de textura (ver Figura 4.4), que representa a cada imagen  $I_i(x)$ .

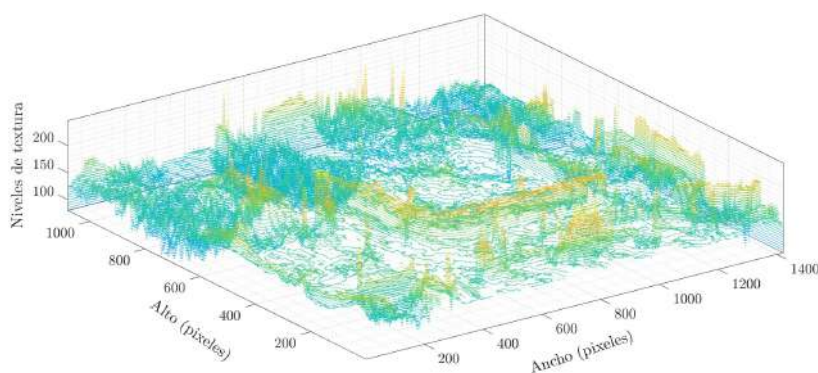


Figura 4.4. Niveles de Textura de la escena Fuente: Elaboración propia.

Al obtener los mapas de niveles de textura, estos se cuantifican debido a que las texturas del escenario no son completamente uniformes, para ello se emplean los momentos de orden  $r$  de manera que es posible describir los niveles de textura en términos de un modelo estadístico. En esta etapa, el segundo momento estadístico es utilizado ya que proporciona una media de suavidad de la región. Sea  $I_i$  la imagen dada en la Figura 4.4 y  $V_i(x)$  un vecindario de la imagen en la posición  $x$  dada, entonces el modelo de fondo  $\xi(x) \in (0, 1)^{m \times n}$ , es obtenido de la siguiente manera:

$$\xi(x) = F(x)^2 \cdot P_i(x) \quad (4.1)$$

donde  $F = X - \bar{X}$ ,  $\bar{X}$  es el promedio de las  $n$  distribuciones  $M_i$  y  $P_i$  del mapa de



textura,  $X$  es la distribución más reciente del mapa de textura.

La superficie resultante al aplicar la Ecuación (4.1) se muestra en la Figura 4.5 en donde se reflejan las distribuciones de los momentos en la escena. Mientras que la distribución del escenario se muestra casi homogéneo, la distribución del objeto muestra una mayor dispersión en su vecindario, por lo que ahora es posible comparar la variabilidad de los datos y de acuerdo a estas distribuciones de los bordes, el cambio puede ser detectable.

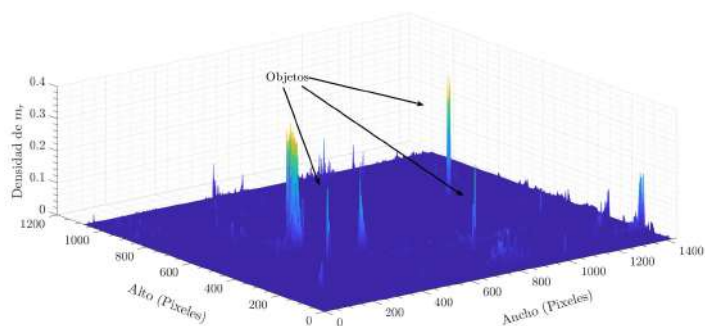


Figura 4.5. Cuantificación de la textura de la escena Fuente: Elaboración propia.

Finalmente, para segmentar entre los objetos estacionarios (fondo de la escena) y los no estacionarios (objetos en movimiento) es seleccionado un umbral. Dado que ahora el escenario presenta la distribución de la Figura 4.6, los objetos que se encuentran en movimiento pueden ser localizados a partir de  $\pm k\sigma$ , en este sentido,  $k\sigma$  es el umbral de confianza que representa a los objetos estacionarios, el cual se encuentra entre  $(-k\sigma, +k\sigma)$  y los objetos en movimiento pueden ser determinados entre  $(\pm k\sigma, \pm \text{inf})$ .

La distribución de los momentos en la escena tiene una desviación típica de  $\sigma$ , el cuál es tomado como umbral de segmentación. Las partículas restantes del resultado de segmentación de los objetos en movimiento son removidas del fondo de la escena,

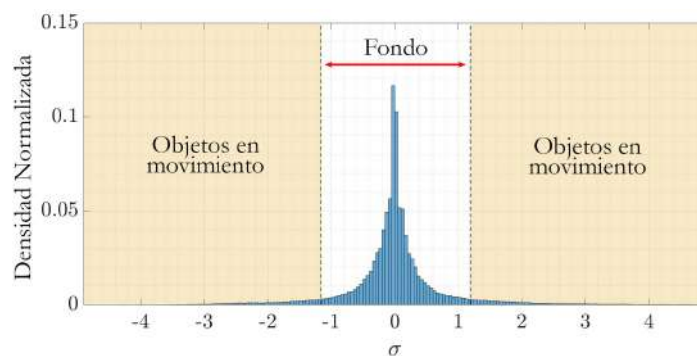


Figura 4.6. Desviación típica de la textura caracterizada en la escena Fuente: Elaboración propia.

conservando aquellas formas que se mantengan después de  $n$  erosiones binarias. Esta segmentación se muestra en la Figura 4.7. Para este caso ha sido usado el valor de  $n = 20$ .



(a) Cuadro original      (b) Cuadro segmentado

Figura 4.7. Segmentación de objetos en movimiento

A continuación se presentan los resultados experimentales que se han llevado a cabo para validar los algoritmos desarrollados, la relación de experimentos realizados se muestra en la Cuadro 4.1.

Experimento	Dataset usado
Segmentación a distintos niveles de textura	Majecka y PETS dataset
Sustracción de fondo	CDNET dataset

Cuadro 4.1: Relación de experimentos en bases de datos. Fuente: Elaboración propia

Los siguientes resultados experimentales consisten en ajustar los parámetros del

filtro para caracterizar los objetos ante perturbaciones como cambios de iluminación y texturas, en la cual se busca que la información segmentada sea relevante para el análisis, es decir profundizar en niveles de intensidad en la escena para minimizar la cantidad de detalles que se usarán para el análisis del objeto.

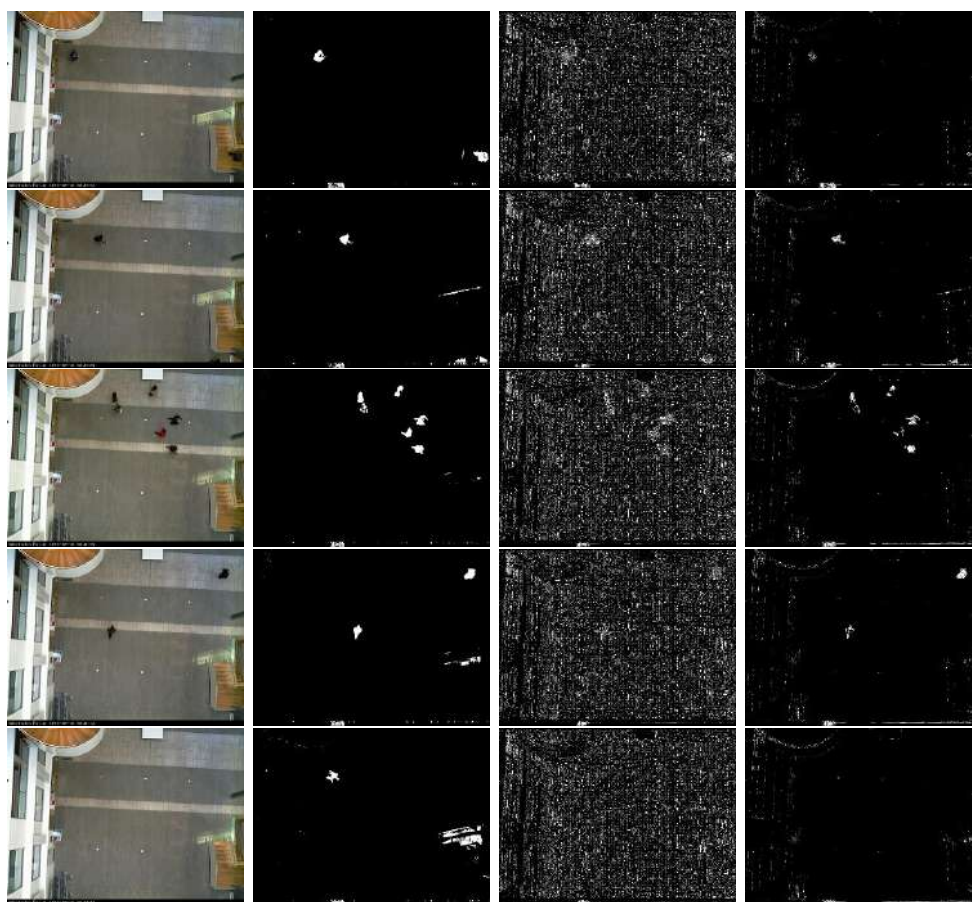


Figura 4.8. Ajuste del valor  $\lambda$  para caracterización de la escena

La Figura 4.8 muestra los resultados del análisis de textura de nivel de la escena en la que se ha ajustado el valor de la frecuencia, para resaltar la textura de las personas, estructura del piso y bordes de edificios.

Los parámetros usados para el modelo de fondo son:  $\lambda = 0,25$  para la textura de las personas,  $\lambda = 0,6$  para la estructura del piso y  $\lambda = 0,4$  para los bordes de edificios.

Además,  $\sigma_x = 3,35$ ,  $\sigma_y = 1,675$ ,  $\phi = 0$  y 24 orientaciones con un desplazamiento angular de 15 grados. Esta parametrización permite observar los objetos de la escena tanto estáticos como dinámicos a distintas profundidades.

En la siguiente etapa de pruebas se han ajustado los parámetros del filtro para para resaltar la textura, los bordes y la estructura de las personas (ver [Figura 4.9.](#))



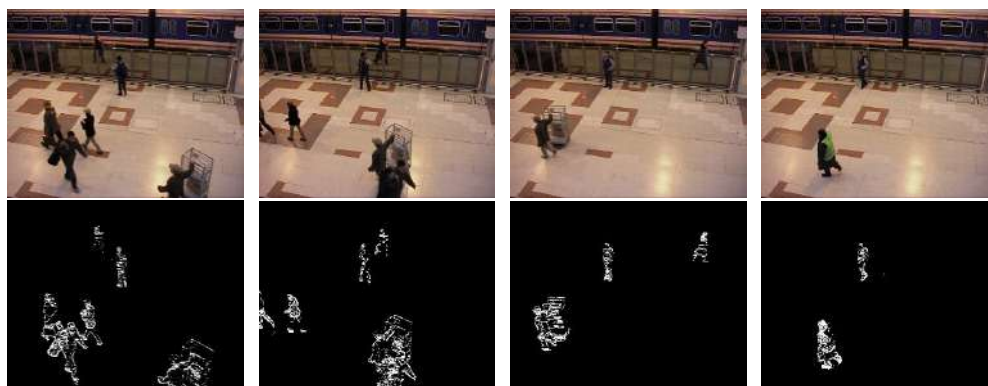
[Figura 4.9.](#) Ajuste del valor  $\lambda$  para caracterización de las personas en la escena

Los parámetros usados para el modelo de fondo son:  $\lambda = 0,35$  para la textura de las personas,  $\lambda = 0,4$  para los bordes de las personas y  $\lambda = 1,05$  para la estructura

de las personas. Además,  $\sigma_x = 3,35$ ,  $\sigma_y = 1,675$ ,  $\phi = 0$  y 24 orientaciones con un desplazamiento angular de 15 grados. Esta parametrización permite observar los objetos de la escena tanto estáticos como dinámicos a distintas profundidades.

Para analizar el funcionamiento de modelo de fondo, se analizaron secuencias de imágenes de distintos escenarios, en los cuales se han reducido los efectos producidos por las sombras conservando la mayor parte de la estructura.

En la [Figura 4.10](#), se ajustan los parámetros para reducir los efectos de sombras y reflejos debido a las condiciones de iluminación en una estación de tren, para este escenario los valores de los parámetros del modelo son los siguientes  $\sigma = 2,5$ ,  $\phi = 0$ ,  $\lambda = 1,5$ ,  $\sigma_x = 3,25$  y  $\sigma_y = 1,25$ .



[Figura 4.10](#). Modelo de fondo en escenario A

En el escenario de la [figura 4.11](#) se tratan problemas de movimiento dinámico, provocado por el movimiento en las hojas de los árboles, además de segmentar elementos con perspectiva, para este escenario los valores de los parámetros del modelo son los siguientes  $\sigma = 1,5$ ,  $\phi = 0$ ,  $\lambda = 0,29$ ,  $\sigma_x = 3,25$  y  $\sigma_y = 1,25$

En la [Figura 4.12](#) la escena está sujeta a cambios de luz generados por la bombilla

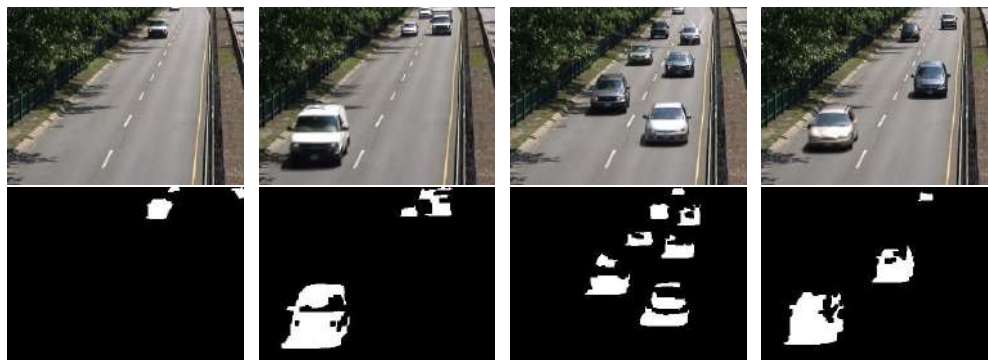


Figura 4.11. Modelo de fondo en escenario B

de la habitación, para este escenario los valores de los parámetros del modelo son los siguientes  $\sigma = 1,5$ ,  $\phi = 0$ ,  $\lambda = 0,35$ ,  $\sigma_x = 4,25$  y  $\sigma_y = 1,5$ .

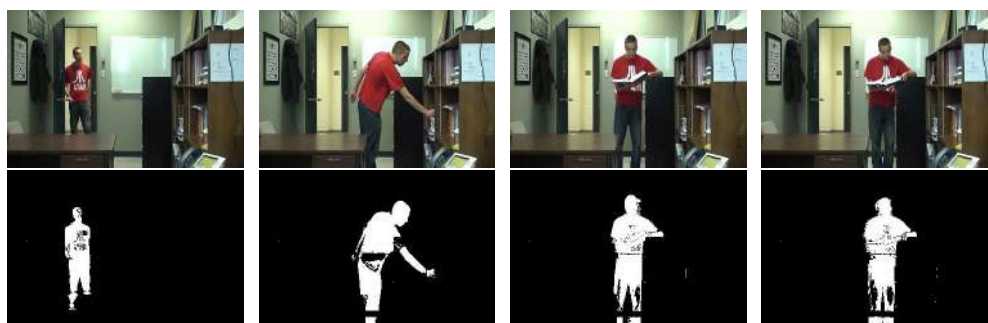


Figura 4.12. Modelo de fondo en escenario C

Finalmente, la escena de la Figura 4.13 contiene iluminación excesiva y la adquisición de las imágenes presenta muestreo insuficiente de la secuencia, en ella se pueden observar algunas personas caminando o en bicicleta en un parque. Los parámetros usados fueron los siguientes  $\sigma = 1,5$ ,  $\phi = 0$ ,  $\lambda = 0,3$ ,  $\sigma_x = 3,25$  y  $\sigma_y = 1,25$





Figura 4.13. Modelo de fondo en escenario D

## 4.2 Resultados de la Extracción de características

Para evaluar la invarianza del descriptor ante escalamiento y rotación, se utilizó el conjunto de datos Kimia MPEG7 CE-Shape-1 (Paramarthalingam y T. T, 2021). Esta base de datos contiene 1.400 imágenes, las cuales se dividen en 70 clases con 20 formas por clase. En la Figura 4.14 se muestran ejemplos de formas de objetos del conjunto de datos MPEG7 Parte B.

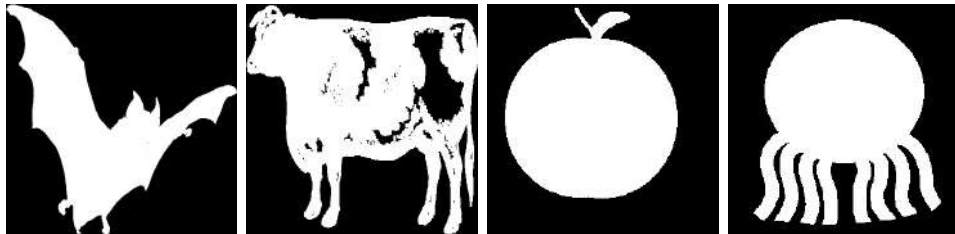


Figura 4.14. Escenarios MPEG-7 Fuente: Paramarthalingam y T. T (2021).

Los resultados experimentales mostraron que el descriptor propuesto tiene una fuerte discriminabilidad, en el Cuadro 4.2, se muestran los resultados de evaluar el descriptor ante distintos objetos de una misma clase (*apple* – 1), en la cual se incluye la varianza y el error típico que presentan los datos.

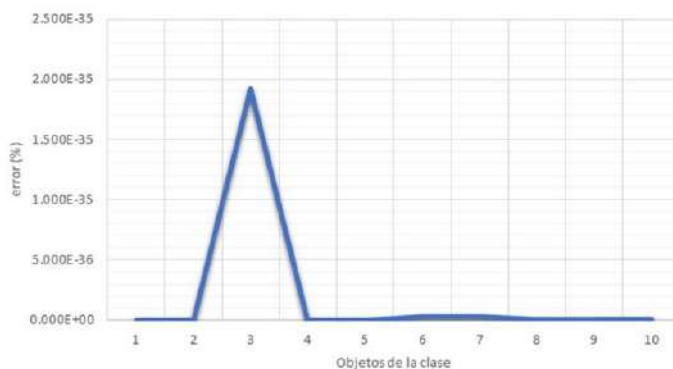
Todas las columnas  $s_n$  representan las formas evaluadas de las clases, mientras que

## 4.2 Resultados de la Extracción de características

	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	Varianza	Error típico
$e_1$	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	3.994E-31	6.320E-16
$e_2$	0.2642	0.2768	0.2661	0.2841	0.2644	0.2872	0.2703	0.2938	0.2720	0.2703	9.491E-05	9.742E-03
$e_3$	0.1201	0.1159	0.1214	0.1045	0.1217	0.1023	0.1164	0.0969	0.1139	0.1154	6.691E-05	8.180E-03
$e_4$	0.0288	0.0258	0.0274	0.0288	0.0281	0.0293	0.0289	0.0286	0.0277	0.0275	9.433E-07	9.712E-04
$e_5$	0.0220	0.0198	0.0213	0.0204	0.0214	0.0183	0.0199	0.0186	0.0226	0.0222	2.019E-06	1.421E-03
$e_6$	0.0116	0.0114	0.0117	0.0119	0.0118	0.0118	0.0121	0.0116	0.0115	0.0116	3.850E-08	1.962E-04
$e_7$	0.0082	0.0077	0.0079	0.0077	0.0081	0.0083	0.0082	0.0080	0.0081	0.0085	5.684E-08	2.384E-04
$e_8$	0.0063	0.0059	0.0063	0.0059	0.0061	0.0055	0.0057	0.0054	0.0059	0.0061	8.020E-08	2.832E-04
$e_9$	0.0045	0.0043	0.0042	0.0046	0.0045	0.0048	0.0046	0.0048	0.0044	0.0045	3.458E-08	1.860E-04
$e_{10}$	0.0037	0.0034	0.0037	0.0033	0.0036	0.0032	0.0035	0.0032	0.0037	0.0037	4.079E-08	2.020E-04

**Cuadro 4.2:** Variabilidad de los valores singulares entre formas de una misma clase  
Fuente: Elaboración propia

en se refiere a los valores singulares obtenidos para cada forma. En la [Figura 4.15](#) se muestra el error porcentual de la clase *apple* – 1, donde se muestran las variaciones en los valores singulares de los objetos de cada imagen de la clase.



**Figura 4.15.** Error porcentual de formas de la clase *apple* – 1 Fuente: Elaboración propia

Por otra parte, en el [Cuadro 4.3](#) se muestran los resultados de invariancia a rotación y traslación, cada rotación fue de  $10^\circ$ .

En la [Figura 4.16](#), se muestra el error porcentual, dado por las variaciones en los valores singulares de los objetos de la clase *apple* – 1, aplicando rotaciones cada  $10^\circ$ .

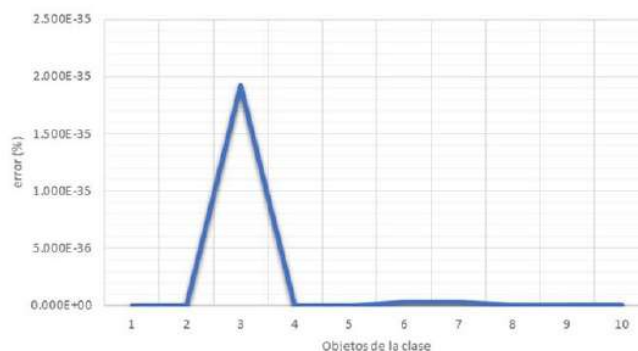
Los resultados tras escalar los elementos de una misma clase aplicando escala cada 10% del valor real de la imagen, se muestran en el [Cuadro 4.4](#).



## 4.2 Resultados de la Extracción de características

	0°	10°	20°	30°	40°	50°	60°	70°	80°	90°	Varianza	Error típico
$e_1$	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.0000	0.0000
$e_2$	0.3037	0.3002	0.2969	0.2926	0.2898	0.2850	0.2813	0.2784	0.2749	0.2735	0.0001	0.0107
$e_3$	0.0802	0.0806	0.0833	0.0884	0.0915	0.0976	0.1022	0.1072	0.1113	0.1134	0.0001	0.0126
$e_4$	0.0301	0.0346	0.0370	0.0372	0.0370	0.0354	0.0344	0.0321	0.0299	0.0284	0.0000	0.0033
$e_5$	0.0217	0.0205	0.0188	0.0178	0.0176	0.0173	0.0173	0.0184	0.0201	0.0211	0.0000	0.0017
$e_6$	0.0116	0.0118	0.0115	0.0112	0.0113	0.0118	0.0117	0.0116	0.0119	0.0117	0.0000	0.0002
$e_7$	0.0083	0.0084	0.0082	0.0081	0.0082	0.0084	0.0086	0.0083	0.0081	0.0079	0.0000	0.0002
$e_8$	0.0062	0.0059	0.0058	0.0060	0.0063	0.0063	0.0060	0.0057	0.0058	0.0061	0.0000	0.0002
$e_9$	0.0045	0.0043	0.0049	0.0049	0.0046	0.0043	0.0046	0.0049	0.0048	0.0044	0.0000	0.0002
$e_{10}$	0.0034	0.0037	0.0037	0.0035	0.0034	0.0037	0.0038	0.0035	0.0033	0.0036	0.0000	0.0002

**Cuadro 4.3:** Variabilidad de los valores singulares entre formas de una misma clase aplicando rotaciones



**Figura 4.16.** Error porcentual ante rotaciones de formas de la clase *apple* – 1 Fuente: Elaboración propia

	100 %	90 %	80 %	70 %	60 %	50 %	40 %	30 %	20 %	10 %	Varianza	Error típico
$e_1$	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.000E+00	0.000E+00
$e_2$	0.3037	0.3037	0.3037	0.3037	0.3037	0.3037	0.3037	0.3037	0.3037	0.3037	0.000E+00	0.000E+00
$e_3$	0.0802	0.0802	0.0802	0.0802	0.0802	0.0802	0.0802	0.0802	0.0802	0.0802	1.926E-34	1.463E-17
$e_4$	0.0301	0.0301	0.0301	0.0301	0.0301	0.0301	0.0301	0.0301	0.0301	0.0301	0.000E+00	0.000E+00
$e_5$	0.0217	0.0217	0.0217	0.0217	0.0217	0.0217	0.0217	0.0217	0.0217	0.0217	0.000E+00	0.000E+00
$e_6$	0.0116	0.0116	0.0116	0.0116	0.0116	0.0116	0.0116	0.0116	0.0116	0.0116	3.009E-36	1.829E-18
$e_7$	0.0083	0.0083	0.0083	0.0083	0.0083	0.0083	0.0083	0.0083	0.0083	0.0083	3.009E-36	1.829E-18
$e_8$	0.0062	0.0062	0.0062	0.0062	0.0062	0.0062	0.0062	0.0062	0.0062	0.0062	7.523E-37	9.143E-19
$e_9$	0.0045	0.0045	0.0045	0.0045	0.0045	0.0045	0.0045	0.0045	0.0045	0.0045	7.523E-37	9.143E-19
$e_{10}$	0.0034	0.0034	0.0034	0.0034	0.0034	0.0034	0.0034	0.0034	0.0034	0.0034	1.881E-37	4.571E-19

**Cuadro 4.4:** Variabilidad de los valores singulares entre formas de una misma clase con escalado gradual

En la [Figura 4.17](#) se muestra el error porcentual, dado por las variaciones en los valores singulares de los objetos de una misma clase con escalado gradual.

Los resultados tras comparar los elementos de una misma clase, se muestran en el [Cuadro 4.5](#).

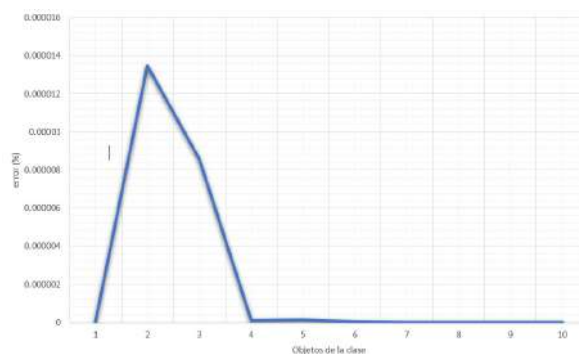


Figura 4.17. Error porcentual entre objetos de la clase *apple* – 1 aplicando escalado gradual Fuente: Elaboración propia

En la Figura 4.18 se muestra el error porcentual, dado por las variaciones en los valores singulares de los objetos de una misma clase.

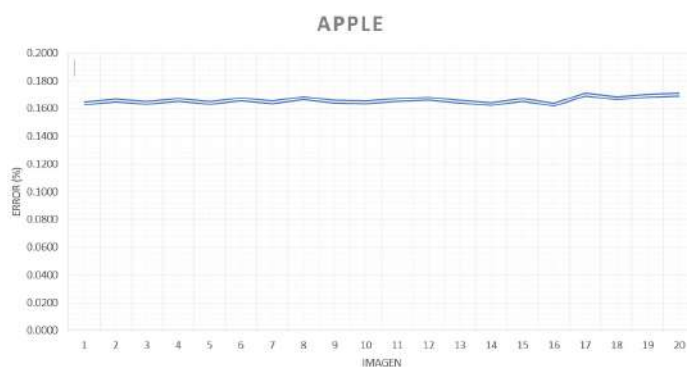


Figura 4.18. Error porcentual entre objetos la clase *apple* – 1 Fuente: Elaboración propia

De acuerdo con las pruebas realizadas, se puede apreciar que el descriptor propuesto presenta invarianza a transformaciones como: traslación, rotación y escalado. Para la siguiente etapa de pruebas, se analiza de forma cualitativa los resultados del algoritmo, usando los videos de la base de datos de *StanfordDataSet*, en la Figura 4.19 se aprecian distintos objetos como peatones, ciclistas, patinadores, automóviles, autobuses y carritos de golf (Robicquet y col., 2016). En estas escenas es aplicado el modelo de fondo explicado en la sección de segmentación de la metodología, se deter-

## 4.2 Resultados de la Extracción de características

Figura	$e_1$	$e_2$	$e_3$	$e_4$	$e_5$	$e_6$	$e_7$	$e_8$	$e_9$	$e_{10}$	$e_{11}$	$e_{12}$
in000001	0.5000	0.2642	0.1201	0.0288	0.0220	0.0116	0.0082	0.0063	0.0045	0.0037	0.0269	0.1639
in000002	0.5000	0.2768	0.1159	0.0258	0.0198	0.0114	0.0077	0.0059	0.0043	0.0034	0.0275	0.1657
in000003	0.5000	0.2661	0.1214	0.0274	0.0213	0.0117	0.0079	0.0063	0.0042	0.0037	0.0270	0.1643
in000004	0.5000	0.2841	0.1045	0.0288	0.0204	0.0119	0.0077	0.0059	0.0046	0.0033	0.0276	0.1663
in000005	0.5000	0.2644	0.1217	0.0281	0.0214	0.0118	0.0081	0.0061	0.0045	0.0036	0.0269	0.1641
in000006	0.5000	0.2872	0.1023	0.0293	0.0183	0.0118	0.0083	0.0055	0.0048	0.0032	0.0278	0.1667
in000007	0.5000	0.2703	0.1164	0.0289	0.0199	0.0121	0.0082	0.0057	0.0046	0.0035	0.0271	0.1647
in000008	0.5000	0.2938	0.0969	0.0286	0.0186	0.0116	0.0080	0.0054	0.0048	0.0032	0.0281	0.1676
in000009	0.5000	0.2720	0.1139	0.0277	0.0226	0.0115	0.0081	0.0059	0.0044	0.0037	0.0272	0.1648
in000010	0.5000	0.2703	0.1154	0.0275	0.0222	0.0116	0.0085	0.0061	0.0045	0.0037	0.0271	0.1646
in000011	0.5000	0.2835	0.1025	0.0304	0.0211	0.0113	0.0078	0.0056	0.0046	0.0034	0.0276	0.1661
in000012	0.5000	0.2907	0.0936	0.0296	0.0227	0.0115	0.0080	0.0058	0.0044	0.0036	0.0279	0.1670
in000013	0.5000	0.2734	0.1115	0.0286	0.0224	0.0122	0.0075	0.0060	0.0043	0.0036	0.0272	0.1649
in000014	0.5000	0.2576	0.1249	0.0313	0.0192	0.0133	0.0083	0.0058	0.0049	0.0036	0.0266	0.1632
in000015	0.5000	0.2812	0.1152	0.0248	0.0186	0.0111	0.0079	0.0054	0.0041	0.0036	0.0277	0.1664
in000016	0.5000	0.2517	0.1331	0.0294	0.0194	0.0134	0.0078	0.0059	0.0047	0.0036	0.0265	0.1629
in000017	0.5000	0.3101	0.0865	0.0290	0.0181	0.0103	0.0071	0.0053	0.0040	0.0031	0.0289	0.1701
in000018	0.5000	0.2914	0.1006	0.0285	0.0189	0.0114	0.0076	0.0057	0.0041	0.0033	0.0280	0.1674
in000019	0.5000	0.3033	0.0910	0.0290	0.0191	0.0096	0.0076	0.0056	0.0040	0.0034	0.0286	0.1690
in000020	0.5000	0.3093	0.0896	0.0277	0.0177	0.0090	0.0073	0.0057	0.0040	0.0034	0.0289	0.1700

**Cuadro 4.5:** Error porcentual entre objetos de una misma clase *apple* – 1. Fuente: Elaboración propia

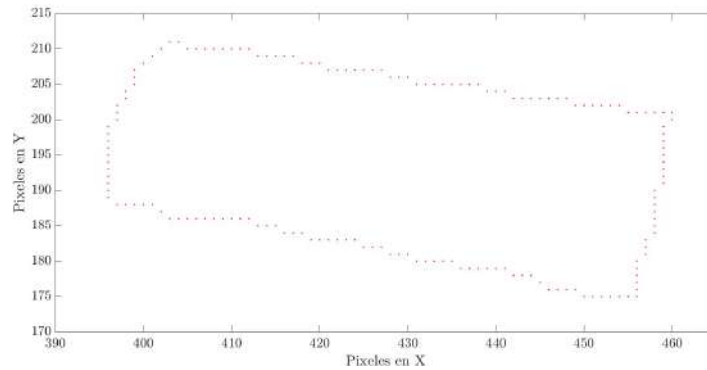
minan los vectores característicos de las formas y se clasificación mediante el algoritmo de k-means, la dificultad de estos escenarios son la distancia de los objetos con respecto a la cámara, las condiciones de iluminación y sombras generadas por la perspectiva del sol.



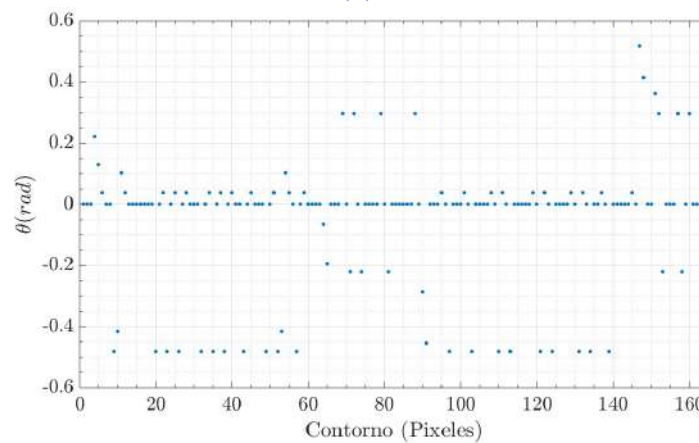
**Figura 4.19.** Escenarios Stanford Drone Fuente: Robicquet y col. (2016).

Como ya se ha mencionado en la sección metodología de este trabajo, se ha usado el ángulo tangencial para extraer las características de los objetos, considere la figura **Figura 4.7** en la cual se se aprecia la segmentación. Mientras que en la **Figura 4.20.b** se muestran los datos de una de las columnas de D (ver **Ecuación (3.21)**), estos datos

representan las variaciones que definen la forma [Figura 4.20.a](#), las variaciones más altas representan curvaturas no tan suaves como esquinas o elementos que hacen distintivos de la forma del objeto.



(a)



(b)

Figura 4.20. Variaciones en la forma del objeto

Cada termino  $k$  de la [Ecuación \(3.26\)](#) es una aproximación de la matriz  $D$ , representada por los valores singulares de la matriz  $D$ , los valores singulares contenidos en la matriz  $\Sigma$  representan los ejes de las variaciones, esta representación se muestra en la [Figura 4.21](#).

Mientras que las columnas de  $V$  son vectores singulares derechos y son una pre-imagen de los ejes principales, esta matriz deja los ejes orientados según las direcciones

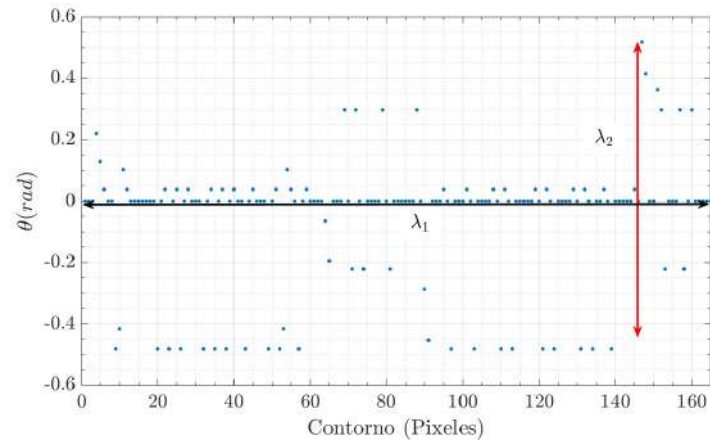


Figura 4.21. Representación de los valores singulares de los vectores de D

principales, en la descomposición forman un conjunto ortogonal sobre D. Tanto los vectores singulares derechos como los izquierdos, permiten identificar los ejes en los que son encontradas las variaciones, con lo cual se puede identificar los ejes principales de la geometría del objeto como se aprecia en la Figura 4.22.

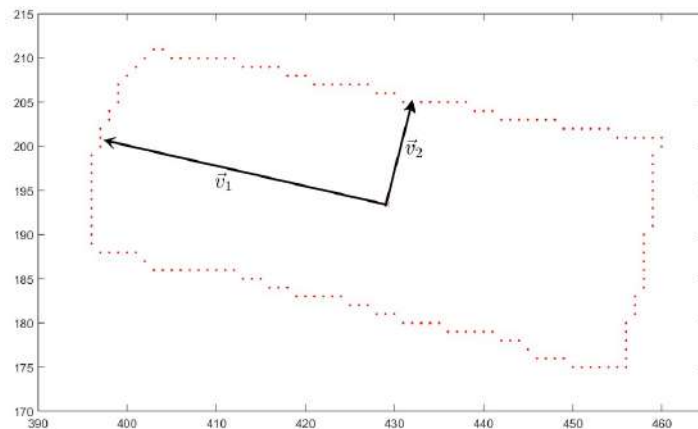


Figura 4.22. Vectores singulares de los ejes principales del objeto

En la Figura 4.23, se muestran los resultados de obtener los valores singulares de personas, en la cual se observa que a pesar que los primeros dos valores singulares son los de mayor peso. Sin embargo, se aprecian otros valores singulares que son significativos para la forma del objeto.

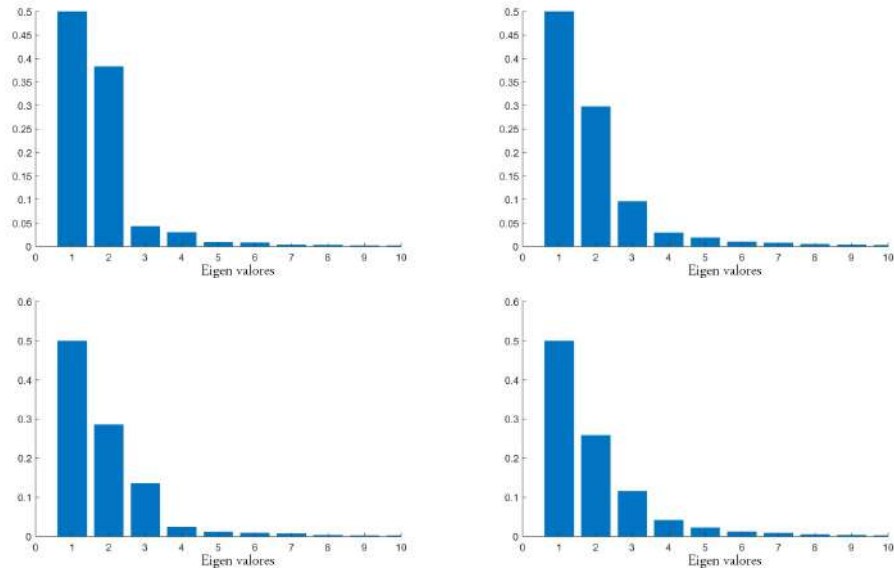


Figura 4.23. Influencia de los valores característicos de objetos tipo personas

A diferencia de los valores singulares para objetos de tipo personas, para automóviles solo se presentan dos con mayor ponderación, estos resultados se muestran en la Figura 4.24.

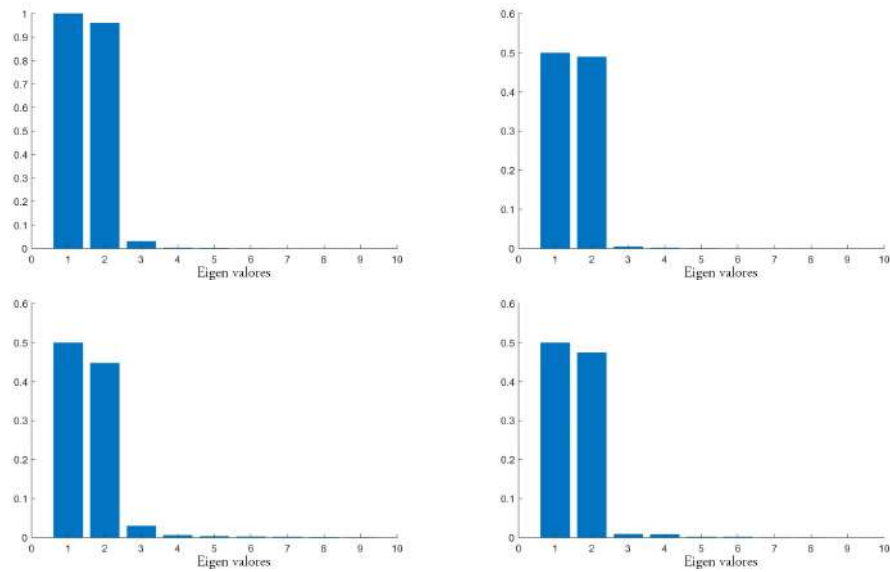


Figura 4.24. Influencia de los valores característicos de objetos tipo automóviles

En la siguiente etapa de pruebas, se analiza de forma cualitativa los resultados del algoritmo, usando los videos de la base de datos de Stanford, en la cual se aprecian distintos objetos como peatones, ciclistas, patinadores, automóviles, autobuses y carritos de golf, en estas escenas es aplicado el modelo de fondo explicado en la sección de segmentación de la metodología, se determinan los vectores característicos de las formas y se clasificaron mediante el algoritmo de k-means, la dificultad de estos escenarios son la distancia de los objetos con respecto a la cámara, las condiciones de iluminación y sombras generadas por la perspectiva del sol (ver [Figura 4.25](#)).



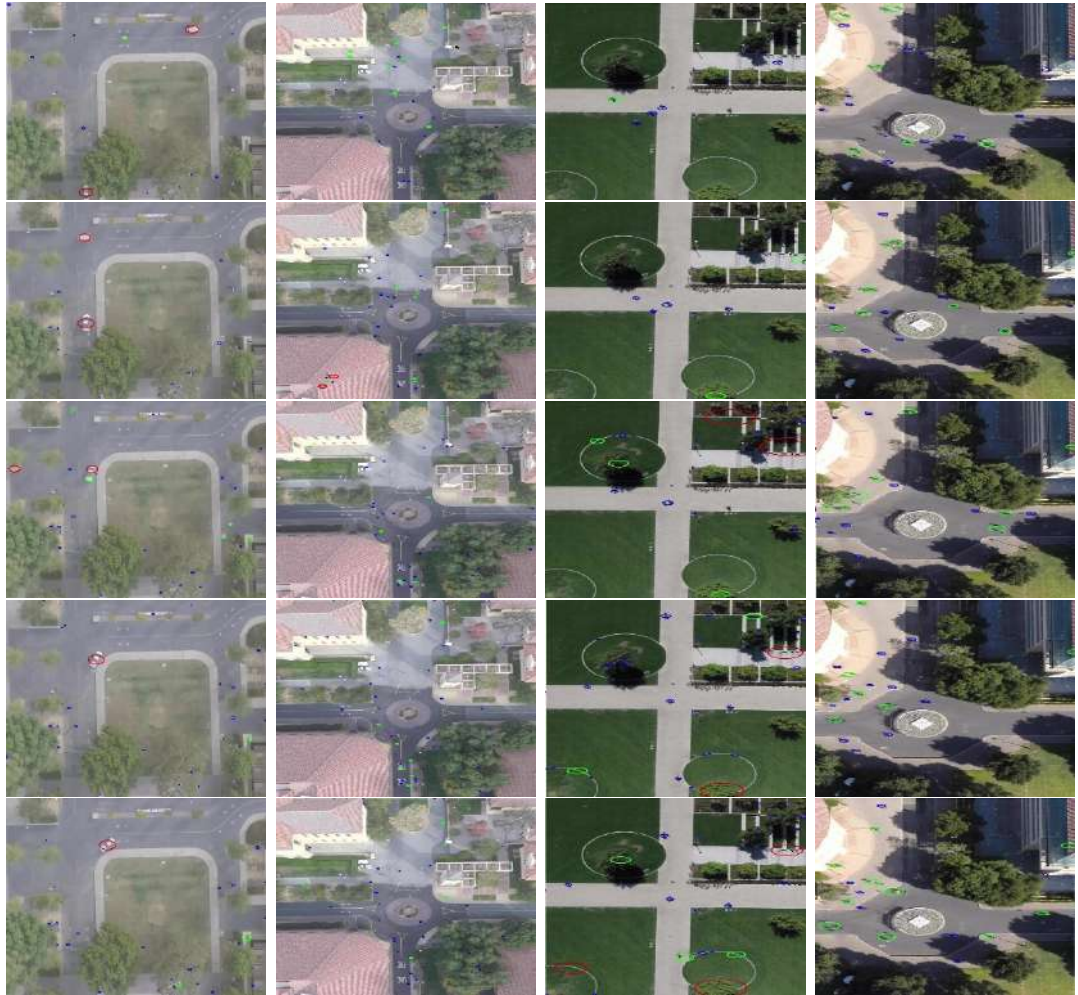


Figura 4.25. Clasificación de objetos en movimiento



## Capítulo 5 Discusión

### 5.1 Discusión

Para poder obtener una segmentación adecuada, es necesario parametrizar el algoritmo, obsérvese la [Figura 5.1](#), en la cual podrá notar que la función de Gabor depende de los parámetros  $\lambda$ ,  $\sigma_x$ ,  $\sigma_y$ ,  $\phi$  y  $\theta$  de acuerdo con la [Ecuación \(2.13\)](#), los cuales generan distintos efectos en el filtrado de la imagen.

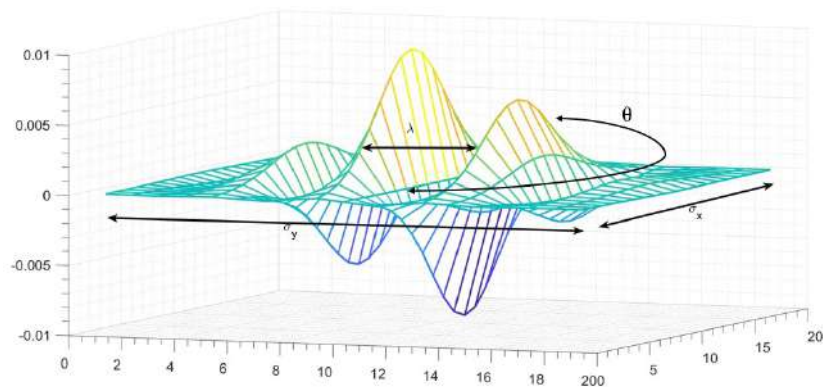
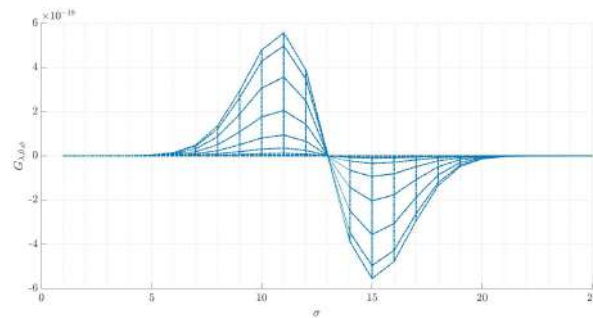


Figura 5.1. Núcleo de la función de gabor.

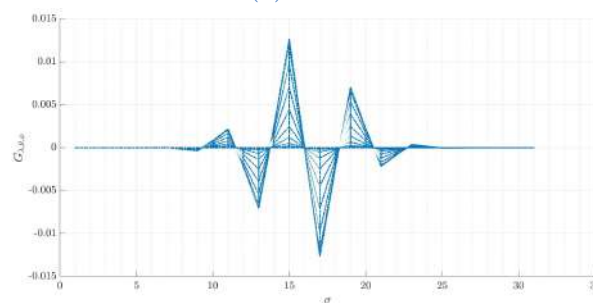
#### 5.1.1 Parámetro $\lambda$

La función portadora y la envolvente están en función de este parámetro, este parámetro incrementa o disminuye la frecuencia de la función, lo que significa que, cuando  $\lambda$  es un valor grande la frecuencia de la envolvente es menor. Sin embargo, si  $\lambda$  es pequeño, tendrá una frecuencia más alta. En la [Figura 5.2](#) se muestra el filtro con

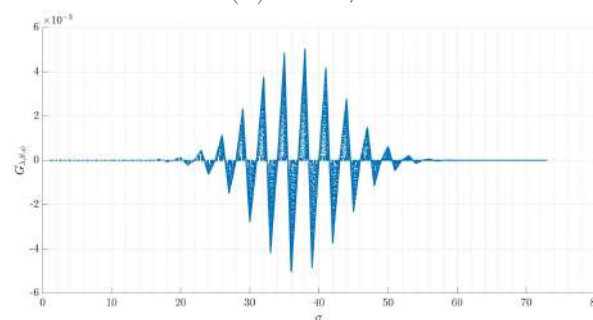
relación a algunos valores de  $\lambda$ .



(a)  $\lambda = 1$



(b)  $\lambda = 0,76$



(c)  $\lambda = 0,3$

Figura 5.2. Influencia del parámetro  $\lambda$  en el filtro de Gabor. Fuente: Elaboración propia

En términos de modelado, con un valor  $\lambda$  grande, el filtro atenuará los objetos con bordes delgados. Mientras que, con un valor  $\lambda$  pequeño el filtro atenuará los bordes gruesos, lo cual esto provocará que se visualicen más detalles, pero con una mayor susceptibilidad al ruido.

5.1.2 Parámetro  $\sigma$ 

Tanto  $\sigma_x$  como  $\sigma_y$  hacen que el término gaussiano sea grande o pequeño en algunos de sus ejes, lo que significa que, si se incrementa el valor de  $\sigma$  la función se extiende más en el eje  $x$  que en el eje  $y$ , o viceversa, por lo que el ruido y los bordes en ese eje se atenuarán más. Sin embargo, si el término gaussiano es pequeño, la suavidad de la imagen será baja y ruidosa.

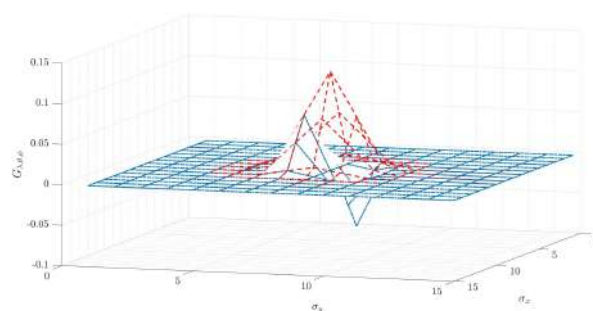
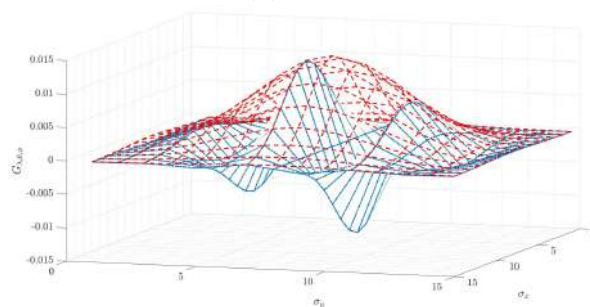
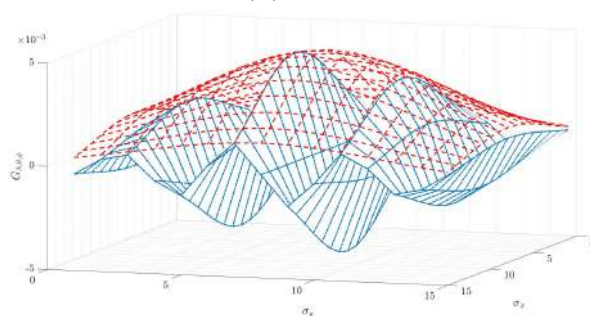
(a)  $\sigma = 1$ (b)  $\sigma = 3$ (c)  $\sigma = 4$ 

Figura 5.3. Selección adecuada del valor de  $\sigma$ . Fuente: Elaboración propia

La Figura 5.3 muestra la selección del parámetro  $\sigma$  para los ejes x-y, en la Figura 5.3.a, el valor de  $\sigma$  no es lo suficientemente grande para construir de forma adecuada la envolvente, por lo que el núcleo del filtro se ve truncado el espacio de la envolvente. En la Figura 5.3.c el valor de  $\sigma$  es muy grande, por la envolvente no cabe en el espacio del núcleo del filtro y finalmente en la Figura 5.3.b se aprecia una relación adecuada entre núcleo del filtro y la envolvente. En general, con el parámetro  $\sigma$  se construye un espacio para la envolvente del filtro.

### 5.1.3 Parámetro $\theta$

Este parámetro rotará el filtro para que puedan detectarse texturas desde distintos ángulos de la imagen (ver Figura 5.4).

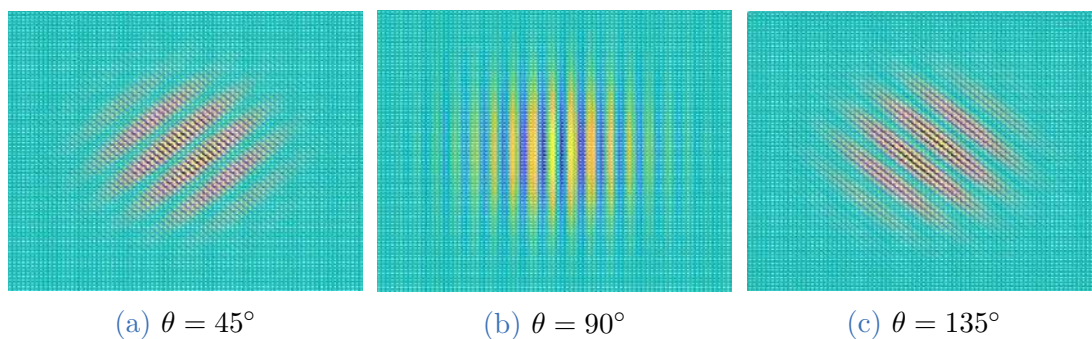


Figura 5.4. Desfase del filtro de gabor. Fuente: Elaboración propia

### 5.1.4 Parámetro $\phi$

Este parámetro desfasará el filtro sobre su eje, en la Figura 5.5, se muestran ejemplos de este desfasamiento a  $0^\circ$ ,  $90^\circ$  y  $180^\circ$ . En términos de modelado, este desfasamiento es especialmente útil para detectar parte de la textura que el filtro no detectó. Pero

la cantidad de bordes detectados es mínima por lo que no influye significativamente en el modelo de fondo.

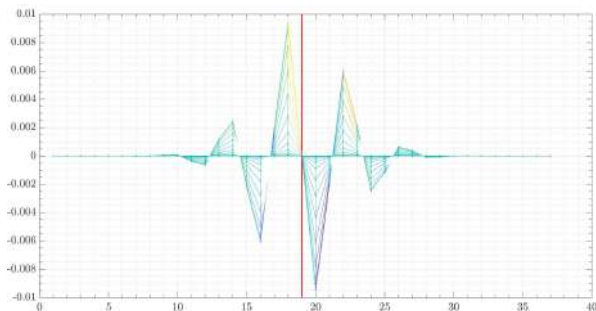
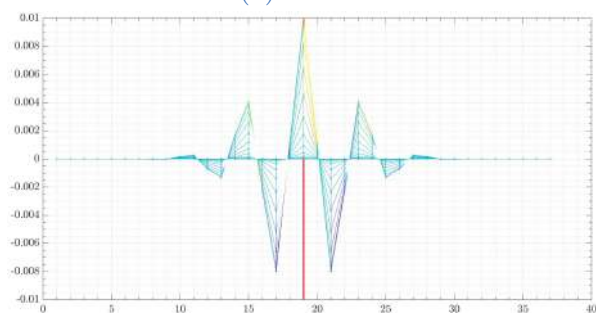
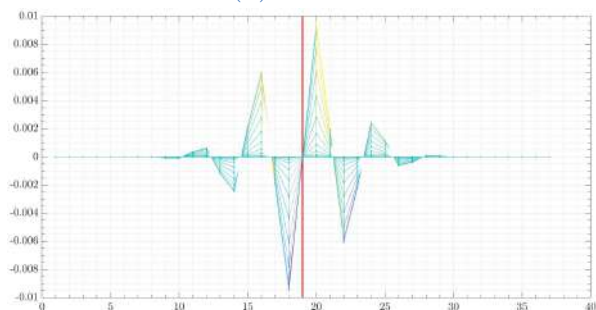
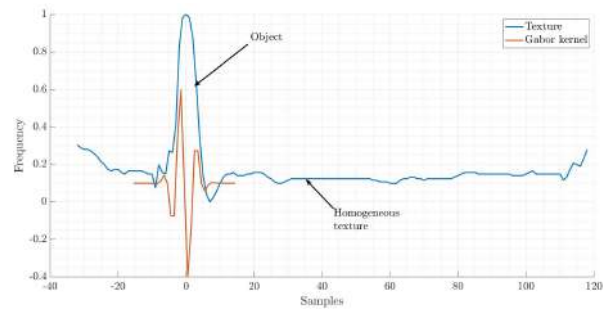
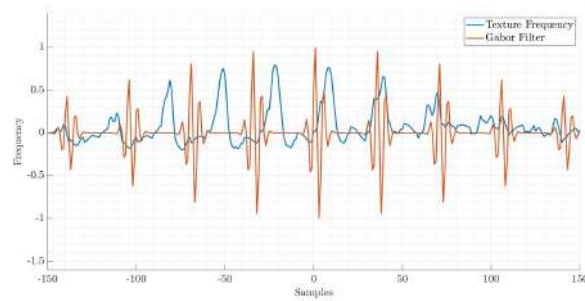
(a)  $\theta = 0^\circ$ (b)  $\theta = 90^\circ$ (c)  $\theta = 180^\circ$ 

Figura 5.5. Desfase del filtro de Gabor. Fuente: Elaboración propia

La selección adecuada de estos parámetros permite que los valores máximos, tanto positivos como negativos, representen puntos de muestreo y midan la deformación de la textura en la estructura del objeto cuando la frecuencia de la región es no uniforme o aperiódica. Sin embargo, cuándo la estructura es periódica y la frecuencia es similar al



(a)



(b)

Figura 5.6. Textura periódica y no periódica de la escena

filtro de Gabor, estas estructuras no serán reconocidas porque los cambios detectados no son tan significativos, que el filtro los atenuará (ver Figura 5.6).

## Capítulo 6 Conclusiones

En esta tesis se presenta un algoritmo completo para la descripción de formas, en el que se realizan dos aportaciones principales:

1) Algoritmo de segmentación de fondo para modelar textura. El cual analiza los niveles de textura en la escena mediante la función de Gabor, y se cuantifica la información usando momentos de orden  $r$ . Los resultados del análisis del algoritmo demuestran invarianzas a cambios lumínicos graduales, que reduce la presencia de sombras, además, debido al análisis en niveles de textura se logra la detección de objetos en entornos casi homogéneos.

Como ventajas principales, el modelo de fondo propuesto conserva la mayor parte de la superficie del objeto, reduce la presencia de sombras y mediante el ajuste adecuado de los parámetros del modelo, es capaz de analizar la escena a distintos niveles de textura.

Sus principales desventajas son que reduce su rendimiento cuando las imágenes son de tamaño pequeño, se requiere aumentar la sensibilidad del algoritmo mediante la manipulación de sus parámetros, por lo cual aumenta la detección de ruido debido a los cambios de iluminación o el movimiento dinámico hojas en los árboles.

2) Algoritmo para descripción de formas. Mediante el cual se analiza la forma de los objetos en movimiento detectados, usando el ángulo tangencial de la curva cerrada,

---

son caracterizados los cambios que se producen en la forma. Debido a que la caracterización se lleva a cabo con el contorno del objeto, el algoritmo se vuelve invariante al escalamiento, la rotación y traslación, con lo cual se logra la confiabilidad de las características extraídas. Estos rasgos característicos, son usados para obtener sus valores singulares, el uso de los valores singulares de SVD permite reducir la dimensionalidad para reconstruir la forma, cumpliendo con la propiedad de homotopía.

La ventaja de usar valores característicos es que permiten la repetibilidad de la forma otra propiedad de los descriptores.

Finalmente, mediante la combinación de estas técnicas fue posible analizar la dinámica atípica en escenarios estructurados y asociar las características de las atipicidades mediante el algoritmo de agrupamiento k-means y de esta forma clasificar los objetos que se encuentran en movimiento para posteriormente se pueda inferir la dinámica de dichos objetos.



## Bibliografía

- Aggarwal, C. & Reddy, C. (2018). Data Clustering: Algorithms and Applications. CRC Press. <https://books.google.com.mx/books?id=cH50DwAAQBAJ>
- Barnouti, N. H., Al-Mayyahi, M. H. N. & Al-Dabbagh, S. S. M. (2018). Real-Time Face Tracking and Recognition System Using Kanade-Lucas-Tomasi and Two-Dimensional Principal Component Analysis. 2018 International Conference on Advanced Science and Engineering (ICOASE), 24-29. <https://doi.org/10.1109/ICOASE.2018.8548818>
- Bataineh, B. (2018). An Iterative Thinning Algorithm for Binary Images Based on Sequential and Parallel Approaches. Pattern Recognition and Image Analysis, 28, 34-43. <https://doi.org/10.1134/S1054661818010030>
- Berjón, D., Cuevas, C., Morán, F. & García, N. (2018). Real-time nonparametric background subtraction with tracking-based foreground update. Pattern Recognition, 74, 156-170. <https://doi.org/https://doi.org/10.1016/j.patcog.2017.09.009>
- Bertrand, G. & Couprie, M. (2017). Chapter 7 - Parallel skeletonization algorithms in the cubic grid based on critical kernels. En P. K. Saha, G. Borgefors & G. Sanniti di Baja (Eds.), Skeletonization (pp. 181-209). Academic Press. <https://doi.org/https://doi.org/10.1016/B978-0-08-101291-8.00008-0>

- Boudaoud, B., Solaiman, L. & Basel Tari, A. (2018). A modified ZS thinning algorithm by a hybrid approach. *The Visual Computer*, 34, 689-706. <https://doi.org/10.1007/s00371-017-1407-4>
- Bouwmans, T., Silva, C., Marghes, C., Zitouni, M. S., Bhaskar, H. & Frelicot, C. (2018). On the role and the importance of features for background modeling and foreground detection. *Computer Science Review*, 28, 26-91. <https://doi.org/https://doi.org/10.1016/j.cosrev.2018.01.004>
- da Fontoura, L., Costa & Cesar, R. (2018). *Shape Analysis and Classification: Theory and Practice*. CRC Press. <https://doi.org/https://doi.org/10.1201/9781315222325>
- de la Puente Viedma, C. (2018). *Estadística descriptiva e inferencial y una introducción al método científico: con un apéndice al método*. IDT CB.
- Díaz, J. A. S. (2021). *Descripción de imágenes de eventos: HOG, GIST y DFT (Tesis doctoral)*. Universidad de Sevilla. Departamento de Ingeniería de Sistemas y Automática.
- Ding, R., Li, X., Nie, L., Li, J., Si, X., Chu, D., Liu, G. & Zhan, D. (2019). Empirical Study and Improvement on Deep Transfer Learning for Human Activity Recognition. *Sensors*, 19(1). <https://doi.org/10.3390/s19010057>
- Ferryman, J. & Shahrokni, A. (2009). PETS2009: Dataset and challenge. 2009 Twelfth IEEE International Workshop on Performance Evaluation of Tracking and Surveillance, 1-6. <https://doi.org/10.1109/PETS-WINTER.2009.5399556>
- Gao, X., Zhang, C. & Duan, H. (2020). An In-Car Objects Detection Algorithm Based on Improved Spatial-Temporal Entropy Image. 2020 IEEE 5th International

- Conference on Signal and Image Processing (ICSIP), 55-59. <https://doi.org/10.1109/ICSIP49896.2020.9339280>
- García-González, J., Ortiz-de-Lazcano-Lobato, J. M., Luque-Baena, R. M., Molina-Cabello, M. A. & López-Rubio, E. (2019). Foreground detection by probabilistic modeling of the features discovered by stacked denoising autoencoders in noisy video sequences. *Pattern Recognition Letters*, 125, 481-487. <https://doi.org/https://doi.org/10.1016/j.patrec.2019.06.006>
- Gil Pascual, J. (2020). APLICACIONES DE R EN ESTADÍSTICA BÁSICA Y TEXTUAL. UNED. <https://books.google.com.mx/books?id=whrdDwAAQBAJ>
- Grauman, K. & Leibe, B. (2011). *Visual Object Recognition*. Morgan & Claypool Publishers. <https://doi.org/10.2200/S00332ED1V01Y201103AIM011>
- Gupta, B., Perez, G., Agrawal, D. & Gupta, D. (2019). *Handbook of Computer Networks and Cyber Security: Principles and Paradigms*. Springer International Publishing.
- Hajian, A. & Ramli, D. A. (2018). Sharpness Enhancement of Finger-Vein Image Based on Modified Un-sharp Mask with Log-Gabor Filter [Knowledge-Based and Intelligent Information & Engineering Systems: Proceedings of the 22nd International Conference, KES-2018, Belgrade, Serbia]. *Procedia Computer Science*, 126, 431-440. <https://doi.org/https://doi.org/10.1016/j.procs.2018.07.277>
- Hu, H., Li, Z., Jin, X., Chen, M. & Shen, Y. (2020). Curve Skeleton Extraction From 3D Point Clouds Through Hybrid Feature Point Shifting and Clustering. *Computer Graphics Forum*, 39. <https://doi.org/10.1111/cgf.13906>

- Hung, K.-L. & Yen, C.-Y. (2019). Watermarking Technique Based on Harris-Laplace Feature Point Detector Capable of Resisting Geometric Attacks. 2019 14th Asia Joint Conference on Information Security (AsiaJCIS), 119-126. <https://doi.org/10.1109/AsiaJCIS.2019.00007>
- Kim, J., Um, S. & Min, D. (2018). Fast 2D Complex Gabor Filter With Kernel Decomposition. IEEE Transactions on Image Processing, 27(4), 1713-1722. <https://doi.org/10.1109/TIP.2017.2783621>
- Lu, S. & Ma, X. (2020). Adaptive random-based self-organizing background subtraction for moving detection. International Journal of Machine Learning and Cybernetics, 11, 1-10. <https://doi.org/10.1007/s13042-019-01037-x>
- Maddalena, L. & Petrosino, A. (2019). Self-organizing background subtraction using color and depth data. Multimedia Tools and Applications, 78(9), 11927-11948. <https://doi.org/10.1007/s11042-018-6741-7>
- Majecka, B. (2009). Statistical models of pedestrian behaviour in the Forum.
- Mohammed, M. J. & Hamdi, M. M. (2018). Hybrid Tow Feature Extraction Descriptor for Shape Pattern Recognition.
- Paramarthalingam, A. & T. T, M. (2021). Extraction of compact boundary normalisation based geometric descriptors for affine invariant shape retrieval. Image Processing, IET, 15, 1093-1104. <https://doi.org/10.1049/ipr2.12088>
- Ramaiah, M. & Prasad, D. K. (2021). Polygonal Approximation of Digital Planar Curve Using Novel Significant Measure. En C. Voloşencu, S. Küçük, J. Guerrero & O.

- Valero (Eds.), Automation and Control. IntechOpen. <https://doi.org/10.5772/intechopen.92145>
- Rani, R., Kumar, R. & Singh, A. P. (2016). An empirical evaluation of local descriptors in object recognition. 2016 International Conference on Advances in Computing, Communications and Informatics (ICACCI), 1517-1521. <https://doi.org/10.1109/ICACCI.2016.7732263>
- Robicquet, A., Sadeghian, A., Alahi, A. & Savarese, S. (2016). Learning Social Etiquette: Human Trajectory Understanding In Crowded Scenes. En B. Leibe, J. Matas, N. Sebe & M. Welling (Eds.), Computer Vision – ECCV 2016 (pp. 549-565). Springer International Publishing.
- Sinaga, K. P. & Yang, M.-S. (2020). Unsupervised K-Means Clustering Algorithm. IEEE Access, 8, 80716-80727. <https://doi.org/10.1109/ACCESS.2020.2988796>
- Song, S.-b. & Kim, J.-h. (2019). SFMOG : Super Fast MOG Based Background Subtraction Algorithm.
- Upadhyay, P. K. & Chandra, S. (2019). Salient Bag of Feature for Skin Lesion Recognition. International Journal of Performability Engineering, 15(4), jourarticle 1083, 1083. <https://doi.org/10.23940/ijpe.19.04.p3.10831093>
- Vázquez-Esquivel, A. V., Aguilera-Navarrete, I. & Juárez-Campos, B. (2021). Method for the simulation of deformation processes by processing metallographs. 2021 Mexican International Conference on Computer Science (ENC), 1-8. <https://doi.org/10.1109/ENC53357.2021.9534810>

- Wade, S. & Ghahramani, Z. (2018). Bayesian Cluster Analysis: Point Estimation and Credible Balls (with Discussion). *Bayesian Analysis*, 13(2), 559-626. <https://doi.org/10.1214/17-BA1073>
- Wang, C., Wang, X., Xia, Z., Ma, B. & Shi, Y.-Q. (2020). Image Description With Polar Harmonic Fourier Moments. *IEEE Transactions on Circuits and Systems for Video Technology*, 30(12), 4440-4452. <https://doi.org/10.1109/TCSVT.2019.2960507>
- Wang, Y., Jodoin, P.-M., Porikli, F., Konrad, J., Benezeth, Y. & Ishwar, P. (2014). CDnet 2014: An Expanded Change Detection Benchmark Dataset. 2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops, 393-400. <https://doi.org/10.1109/CVPRW.2014.126>
- Wu, G. & Xu, L. (2019). Shape description and recognition by implicit Chebyshev moments. *Pattern Recognition Letters*, 128, 137-145. <https://doi.org/https://doi.org/10.1016/j.patrec.2019.08.031>
- Wu, Z., Shanshan, J., Xiaolei, Z., Yuanyuan, W., Yuanyuan, Z., Zhewei, W., Lei, L. & Liu, Q. (2020). Application of image retrieval based on convolutional neural networks and Hu invariant moment algorithm in computer telecommunications. *Computer Communications*, 150, 729-738. <https://doi.org/https://doi.org/10.1016/j.comcom.2019.11.053>
- Yang, L., Oyen, D. & Wohlberg, B. (2019). A Novel Algorithm for Skeleton Extraction From Images Using Topological Graph Analysis. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*.

- Zacharis, N. (2018). Classification and Regression Trees (CART) for Predictive Modeling in Blended Learning. *International Journal of Intelligent Systems and Applications*, 10, 1-9. <https://doi.org/10.5815/ijisa.2018.03.01>
- Zeng, D. & Zhu, M. (2018). Background Subtraction Using Multiscale Fully Convolutional Network. *IEEE Access*, 6, 16010-16021. <https://doi.org/10.1109/ACCESS.2018.2817129>
- Zhou, W., Zhong, B. & Yang, J. (2019). Shape Description and Retrieval in a Fused Scale Space. En T. Gedeon, K. W. Wong & M. Lee (Eds.), *Neural Information Processing* (pp. 70-82). Springer International Publishing.
- Zhu, S., Gao, X., Wang, H., Xu, G., Xie, Q. & Yang, S. (2018). Moving Object Real-time Detection and Tracking Method Based on Improved Gaussian Mixture Model. *2018 IEEE 7th Data Driven Control and Learning Systems Conference (DDCLS)*, 654-658. <https://doi.org/10.1109/DDCLS.2018.8515905>