



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

**Seguimiento de objetos mediante una cámara TOF utilizando  
intensidad y profundidad**

**Tesis**

Que como parte de los requisitos para obtener el Grado de  
Maestro en Ciencias de la Computación

Presenta:

Luis Kayrumet Pérez Buigas

Dirigido por:

Dr. Hugo Jiménez Hernández

Dr. Hugo Jiménez Hernández  
Presidente

Dra. Ana Marcela Herrera Navarro  
Secretaria

Dra. Diana Margarita Cordoba Esparza  
Vocal

Dra. Sandra Luz Canchola Magdaleno  
Suplente

Mtra. Erika del Río Magaña  
Suplente

Centro Universitario, Querétaro, Qro.  
Octubre 2020  
México

La detección de movimiento es una tarea complicada porque la cantidad de factores que interviene (variación de luz, sombras, ruido, reflejos etc.) complica detectar los objetos en movimiento, y por consecuencia extraer información útil para caracterizarlos. En este documento, se propone un método de sustracción de fondo basado en los modelos de mezcla gaussiana utilizando información de cámaras RGB. Para determinar si la segmentación es correcta se analizó el histograma de la región segmentada donde se detectó el movimiento haciendo una prueba de correlación de Pearson. En particular, el método se enfoca en resolver problemas de camuflaje de color y eliminar falsas detecciones antes fuertes cambios de iluminación. Para evaluar el método, se creó un nuevo conjunto de datos que contiene situaciones normales, de camuflaje de color y fuertes cambios de iluminación. El método propuesto muestra resultados de relevancia al realizar el análisis de histogramas permitiendo eliminar falsas detecciones. Por lo tanto, esta técnica ayudará a detectar de manera robusta regiones de interés como preprocesamiento en etapas de procesamiento de imágenes.

Background subtraction and movement segmentation algorithms can be improved by merging the depth and color entries. In this paper, it is proposed a background subtraction method based on the Gaussian mixture models using color and depth information. To combine color and distance information, it's used a probabilistic model based on the histogram of the area where movement was detected. In particular, the method focuses on solving color camouflage problem and depth denoising. For evaluating our method, we built a new dataset containing normal, color camouflage and depth camouflage situations. The dataset files consist of color, depth and ground truth image sequences. The proposed method shows greater resistance to changes in lighting, shadows, reflections and camouflage. Thus, this technique will help to robustly detect regions of interest as pre-processing in high-level image processing stages.

## DEDICATORIA

---

Esta tesis se la dedico a:

A mi esposa que siempre me dio el impulso para terminar la maestría.

A mis padres por estar siempre ahí para mí.

A mi hermana que siempre me ha ayudado y comprendido.

Dirección General de Bibliotecas UAQ

## AGRADECIMIENTOS

---

La culminación de esta tesis es un trabajo conjunto multidisciplinario, entre personas que saben lo que hacen y que sin su conocimiento y apoyo no hubiera sido posible lo anterior. Agradezco a mi asesor el Dr. Hugo Jiménez Hernández que ha sabido ser paciente y se ha ganado el respeto y admiración de mi persona, entre muchas cosas por su capacidad intelectual, paciencia y disponibilidad a atender inquietudes que se fueron suscitando a lo largo de este proceso que fue el desarrollo y culminación del trabajo arriba descrito. A CONACYT, quien me ofreció su beca de maestría para mi manutención a lo largo del proyecto de investigación. A mi esposa por su apoyo y comprensión incondicional en todo momento.

## Índice general

1. Introducción.....	1
1.1. Objetivo General.....	2
1.2. Objetivos específicos.....	3
1.3. Hipótesis.....	3
1.4. Justificación.....	3
1.5. Estructura del documento.....	4
2. Antecedentes.....	5
2.1. Trabajos en detección de objetos en movimiento.....	5
3. Fundamentos teóricos.....	9
3.1. Detección de movimiento.....	9
3.2. Modelo de mezcla gaussiana.....	11
3.3. Modelo del fondo.....	12
3.3.1. Entendiendo el modelo de fondo de mezcla de gaussianas.....	13
3.4. Cámara RGB.....	16
3.5. Cámara TOF.....	16
3.6. Histograma.....	16
4. Análisis de la metodología.....	18
4.1. Captura de imágenes.....	18
4.2. Detección de movimiento.....	19
4.3. Filtrado de imágenes.....	19
4.4. Evaluación.....	21
4.5. Superposición.....	23

## ÍNDICE GENERAL

---

4.6. Complemento .....	24
5. Modelo experimental y resultados .....	25
5.1. Modelo Experimental .....	25
5.2. Resultados.....	26
5.3. Discusión de los resultados .....	27
6. Conclusiones.....	31
7. Referencias bibliográficas .....	32

Dirección General de Bibliotecas UAQ

## Índice de figuras

Figura 1 Metodología utilizada para la segmentación de movimiento .....	18
Figura 2 Cámara TOF DS 525 .....	19
Figura 3 Filtros aplicados a la imagen después de aplicar el método de Zivkovic... 20	
Figura 4 Filtros aplicados a la imagen después de aplicar el método de Zivkovic a la imagen de entrada del sensor de profundidad. ....	21
Figura 5 Histogramas guardados.....	23
Figura 6 Proceso de obtener área donde se detectó movimiento y cálculo de su histograma.....	23
Figura 7 Resultados obtenidos de procesar algunas imágenes con el método propuesto .....	27
Figura 8 Valores de Medida F(F) obtenidos a lo largo de la secuencia. ....	28
Figura 9 Valores de precisión (P) obtenidos a lo largo de la secuencia. ....	29
Figura 10 Valores de recuerdo (R) obtenidos a lo largo de la secuencia. ....	30

### Índice de tablas

Tabla 1 Resultados al procesar las imágenes de la Figura 7 .....	27
--	----

Dirección General de Bibliotecas UAQ

## LISTA DE ACRÓNIMOS

---

PDF - Probability Density Function

BGS - Sustracción de Fondo

TOF - Time of Flight

RGB - Red, Green, Blue

Dirección General de Bibliotecas UAQ

## Capítulo 1

### 1. Introducción

La detección de objetos en movimiento es uno de los temas más investigados en el área de Visión por Computadoras, ya que es un proceso esencial en la vigilancia visual y el monitoreo de actividades. (Del Pizzo, Foggia, Greco, Percannella, & Vento, 2016) Entre las diversas técnicas para la detección de objetos en movimiento, convencionalmente, se han usado muchos métodos de sustracción de fondo (BGS, por sus siglas en inglés) (Del Pizzo, Foggia, Greco, Percannella, & Vento, 2016). En la literatura, se han presentado varios métodos de seguimiento de objetos que utilizan cámara cenital. Las principales tendencias que han inspirado a los investigadores en el diseño de los métodos de seguimiento de objetos comparten una arquitectura común que implica los siguientes pasos (Fernandez-Sanchez, Rubio, Diaz, & Ros, 2013): (i) realizar la detección de primer plano para separar píxeles móviles de la parte estática de la imagen (modelos de fondo); (ii) detectar cada objeto en la escena de la máscara de primer plano (modelos basados en plantillas); (iii) rastrear a cada objeto detectada usando la trayectoria para actualizar el seguimiento cuando la trayectoria pasa la línea virtual (estimación de trayectoria y seguimiento). Las principales diferencias entre las distintas propuestas radican en el enfoque utilizado para resolver los problemas en cada una de las tres etapas anteriores y en el tipo de sensor adoptado (cámara RGB tradicional o sensor de profundidad). Sin embargo, aunque en la literatura existen estas aproximaciones, cada una de ellas está limitada a condiciones acotadas, debido a la complejidad de la variabilidad de las condiciones ambientales en escenarios reales. Uno de estos problemas es la compensación dinámica de las perturbaciones lumínicas. Esto causa que los sensores se saturen o sean afectados por ruido; consecuencia directa es que no se tengan mediciones adecuadas sobre el escenario y el objeto a segmentar.

A pesar de que los algoritmos de vanguardia actuales son capaces de hacer frente a problemas clásicos (como cambios de iluminación repentinos y graduales, objetos de fondo en movimiento, movimientos repetitivos, etc.), la robustez es un requisito crítico para la analítica de video. Por esa razón, muchos autores han propuesto la fusión de diferentes tipos de características, incluyendo la intensidad, los bordes y la información de textura (Zhang & Ngan, 2010) (Zhang, Zhong, & Cao, 2011) Sin embargo, estas características son capturadas por el mismo tipo de sensor de cámara, por lo que se ven afectadas por los mismos problemas. Además, debido a la complejidad de estos métodos, requieren procesadores potentes para ejecutarse en tiempo real. Esto los hace no adecuados para sistemas integrados como cámaras inteligentes y redes de cámaras descentralizadas.

Para manejar los problemas, se han propuesto otros enfoques BGS (Camplani & Salgado, 2014) (Del Pizzo, Foggia, Greco, Percannella, & Vento, 2016) (Fernandez-Sanchez, Rubio, Diaz, & Ros, 2013) utilizando un nuevo tipo de información además de la información de color. Especialmente, la información de profundidad de cámaras estéreo, Microsoft Kinect o sensores de tiempo de vuelo (ToF), se han utilizado con información de color. A pesar de la utilización de los sensores de profundidad, existen escenarios donde la cámara de profundidad en ciertas condiciones no puede detectar las distancias, ya sea porque el objeto está muy cerca, está muy lejos del sensor o existen alguna superficie que no permite la correcta detección de las distancias.

## **1.1. Objetivo General**

Desarrollar un método robusto que ante cambios graduales lumínicos permita la detección y seguimiento de objetos mediante cámaras TOF y convencional en posición cenital en condiciones variables lumínicas, utilizando la información de color y de distancia para usarse en sistemas de vigilancia en escenarios exteriores.

## 1.2. Objetivos específicos

- Realizar un artículo de las distintas aproximaciones que combinen la información de los canales de color y de profundidad en cámaras TOF para seguir objetos
- Realizar una comparativa de los métodos existentes para el seguimiento de objetos utilizando cámaras TOF.
- Combinar en un solo método la utilización de técnicas de seguimiento de objetos, uno que utilice la información de los canales de color (Zivkovic Z. , 2004) y otros que utilice la información del canal de la profundidad (Zivkovic Z. , 2004).
- Realizar el diseño de un modelo experimental donde se demuestre la eficiencia del método propuesto.

## 1.3. Hipótesis

Si se combina la información de profundidad y la información de intensidad de luz en una escena, es posible desarrollar un método robusto en la detección de objetos en movimiento.

## 1.4. Justificación

El estudio de nuevas aproximaciones redundantes tiene por beneficio algoritmos más robustos ante un número mayor de situaciones. Un nuevo método de detección y seguimiento de objetos a partir de conocimiento existente del área investigada, usando cámaras TOF que son más económicas, desde una posición cenital podrá resolver problemas existentes con iluminación y ruido cuando varios objetos pasan juntos por el foco de la cámara. También la complejidad de cómputo necesaria es menor para el procesamiento de la salida de la información de este tipo de la cámara.

La detección de objetos ha sido investigada e implementada por varios autores. Es por ello que la novedad que se presenta en esta investigación es incluir, un algoritmo de evaluación que permita en conocer en cada cuadro que

método de detección es más eficiente para cada objeto, si RGB, profundidad o la combinación de ambos. Esto hace que el método de detección y seguimiento sea más robusto, que sea aplicable a un mayor número de escenarios. Además, abrirá nuevos caminos para estudios sustantivos que presenten situaciones similares a las que aquí se plantean, donde se le puedan incorporar otros tipos de sensores como infrarrojos, sirviendo como marco de referencia a estas.

## **1.5. Estructura del documento**

En primer lugar, se presenta en la introducción la problemática, los objetivos, y la hipótesis que sustentan esta investigación. En el capítulo 2. Se hace un estudio del estado del arte de los métodos de detección y seguimiento de objetos con cámaras RGB y de distancias, el estudio de las aplicaciones existentes de reconocimiento y seguimiento de objetos confirmaran la pertinencia de la investigación con la búsqueda de que tan robustas son las aplicaciones y en que escenarios funcionan. En el capítulo 3 se describen los fundamentos teóricos de la detección y seguimiento de objetos donde se describen los términos más importantes asociados al surgimiento de este tipo de elementos de monitoreo, específicamente el seguimiento de objetos con cámaras TOF. En el capítulo 4, se presenta la metodología utilizada para implementar un sistema de detección de objetos más robusto ante cambios lumínicos. En el Capítulo 5 se presentan los resultados, estos se analizan y discuten con respecto a los resultados obtenidos por otros investigadores, y finalmente, se presentan las conclusiones y trabajos futuros.

## Capítulo 2

### 2. Antecedentes

En la literatura, se han presentado varios métodos de seguimiento de objetos que utilizan cámara cenital. Las principales tendencias que han inspirado a los investigadores en el diseño de los métodos de seguimiento de objetos comparten una arquitectura común que implica los siguientes pasos: (i) realizar la detección de primer plano para separar píxeles móviles de la parte estática de la imagen (modelos de fondo); (ii) detectar cada objeto en la escena de la máscara de primer plano (modelos basados en plantillas); (iii) rastrear a cada objeto detectada usando la trayectoria para actualizar el seguimiento cuando la trayectoria pasa la línea virtual (estimación de trayectoria y seguimiento).

#### 2.1. Trabajos en detección de objetos en movimiento.

Para el análisis del estado del arte se consideró que los trabajos tengan el mismo objetivo de detectar objetos en movimiento, que tengan la misma técnica de detección de movimiento y que utilicen al menos dos sensores para la captura de las imágenes.

- **Counting people by RGB or depth overhead cameras.** En este artículo se utilizó un método basado en la visión para contar el número de personas que cruzan una línea virtual. El método analiza el flujo de video adquirido por una cámara montada en una posición cenital con respecto a la línea de conteo, lo que permite determinar el número de personas que cruzan la línea virtual y proporciona la dirección de cruce para cada persona. El enfoque propuesto ha sido diseñado específicamente para lograr alta precisión y eficiencia computacional, para permitir su adopción en escenarios reales. Se ha llevado a cabo una evaluación exhaustiva del método teniendo en cuenta los principales factores que pueden afectar el rendimiento del recuento y, en particular, la tecnología de adquisición (cámara RGB tradicional y

sensor de profundidad), el escenario de instalación (interior y exterior), la densidad del flujo de personas (personas aisladas y grupos de personas), la velocidad de fotogramas de adquisición y la resolución de la imagen. También se analizó la combinación de los resultados obtenidos del sensor RGB y los sensores de profundidad como una forma de mejorar el rendimiento del recuento. Los resultados experimentales confirman la efectividad del método propuesto, especialmente cuando se combinan RGB y la información de profundidad, y las pruebas en tres arquitecturas de CPU diferentes demuestran la posibilidad de implementar el método tanto en servidores de alta gama para procesar en paralelo una gran cantidad de transmisiones de video y en CPU de bajo consumo como las integradas en cámaras inteligentes comerciales (Del Pizzo, Foggia, Greco, Percannella, & Vento, 2016).

- **Individual pig object detection algorithm based on Gaussian mixture model.** Los modelos de fondo son crucialmente importantes para la extracción de objetos para la detección de objetos en movimiento en un video. El modelo de mezcla de Gaussianas (GMM) es uno de los métodos más populares de los modelos de fondo. El modelo de mezcla gaussiano que se aplicó a la detección de objetivos porcinos tiene algunas deficiencias. Este estudio propuso un algoritmo mejorado basado en el modelo adaptativo de mezcla gaussiana, para superar las deficiencias del modelo de mezcla Gaussiano tradicional en la detección de objetos porcinos. Basado en el modelo de fondo de la mezcla gaussiana, este documento introdujo dos nuevos parámetros de marcos de video  $m$  y  $T_0$ . La distribución gaussiana se escaneó una vez cada  $m$  fotogramas, se eliminó la distribución Gaussiana excesiva para mejorar la velocidad de convergencia del modelo. Mientras tanto, utilizando diferentes ritmos de aprendizaje para suprimir fantasmas, se adoptó una tasa de aprendizaje decreciente para acelerar el modelado de fondo antes de  $T_0$ , el modelo de fondo se estabilizaría a medida que

continuara el tiempo y se pudiera usar una tasa de aprendizaje más pequeña. Para mantener un fondo estable y reducir la interferencia de ruido, se utilizó una tasa de aprendizaje fija después de T0. Los resultados de los experimentos indicaron que este algoritmo podría construir rápidamente el modelo de fondo inicial, detectar los cerdos en movimiento y extraer los contornos completos de los cerdos. El algoritmo se caracteriza por una buena solidez y adaptabilidad (Yiyang, Longqing, Yuanbing, & Yue, 2017).

- **Background Subtraction Based on Color and Depth Using Active Sensors.** En este artículo, se describe un método de fusión para combinar el color y la profundidad en base a un algoritmo avanzado basado en el color. Esta técnica ha sido evaluada por medio de un conjunto de datos completo registrado con Microsoft Kinect, que permite la comparación con el método original. El método propuesto supera a los demás en casi todas las pruebas, mostrando una mayor solidez a los cambios de iluminación, sombras, reflejos y camuflaje (Fernandez-Sanchez, Rubio, Diaz, & Ros, 2013).
- **Background subtraction based on Gaussian mixture models using color and depth information.** En este artículo, proponen un método de sustracción de fondo (BGS) basado en los modelos de la mezcla de Gaussianas utilizando información de color y profundidad. Para combinar información de color y profundidad, se utiliza el modelo probabilístico basado en la distribución gaussiana. En particular, nos enfocamos en resolver el problema del camuflaje del color y la eliminación de ruido de profundidad. Para evaluar el método, construyeron un nuevo conjunto de datos que contiene situaciones normales de camuflaje de color y profundidad de camuflaje. Los archivos del conjunto de datos consisten en secuencias de imágenes de colores, profundidad y suelo. Con estos archivos, compararon el algoritmo propuesto con las técnicas de BGS basadas en colores convencionales en términos de precisión, recuperación y F-medida.

Como resultado, el método mostró mejor rendimiento. Por lo tanto, esta técnica ayuda a detectar con solidez las regiones de interés como preprocesamiento en etapas de procesamiento de imágenes de alto nivel (Song, Noh, Yu, Park, & Lee, Background subtraction based on Gaussian mixture models using color and depth information, 2015).

- **The infrared moving object detection and security detection related algorithms based on W4 and frame difference.** Este artículo presenta la detección de objetos en movimiento por infrarrojos y algoritmos de detección relacionados con los videos de video vigilancia basado en el clásico W4 y en el algoritmo clásico de diferencia de tiempo. El algoritmo W4 clásico es uno de los poderosos algoritmos de resta de fondo que se aplican a las imágenes infrarrojas que pueden detectar objetos en movimiento de manera precisa, integral y rápida. Sin embargo, el algoritmo W4 clásico solo puede superar la deficiencia en el ligero movimiento de fondo. El error se hará cada vez más grande para el sistema de vigilancia a largo plazo, ya que el modelo de fondo no se modifica una vez establecido. En este artículo, se presenta un algoritmo de detección basado en el W4 clásico y la diferencia de trama. No solo puede superar la deficiencia de la detección falsa debido a las mutaciones de estado del fondo, sino que también elimina los agujeros causados por la diferencia de trama (Yin, Liu, Li, & Liu, 2016).

## Capítulo 3

### 3. Fundamentos teóricos

#### 3.1. Detección de movimiento

La labor de identificar el movimiento físico de un objeto dentro de una determinada región o área se llama detección de movimiento. Los enfoques tradicionales para realizar esta tarea se pueden organizar en tres grupos (Nayagam & Ramar, 2015):

- **Sustracción de fondo.** El proceso de extracción de primer plano o de objetos en movimiento como también se denomina, consiste en una serie de métodos que permiten distinguir entre zonas de fondo o estáticas y zonas en movimiento que se corresponden con el primer plano. Para lograr extraer el fondo se guarda una representación de la escena utilizando el modelado digital del fondo. Esta representación debe almacenar la información necesaria para en cada instante establecer, por cada nueva imagen de entrada, qué áreas de la misma se corresponden efectivamente con el fondo y cuáles no. Dado que el fondo puede presentar cambios en el tiempo, es preciso efectuar un modelado dinámico del fondo para adaptarlo a las variaciones que puedan ocurrir en el mismo. Al inicio, cualquier cambio significativo en un área de la imagen respecto del modelo de fondo será tomado como zona en movimiento. En general, cada componente del modelo se actualiza por cada cierto intervalo de tiempo, siempre que este haya perdido significancia.
- **Diferenciación de cuadros.** Considera la diferencia entre dos cuadros consecutivos para detectar el objeto en movimiento. Generalmente, este enfoque no puede obtener todo el contorno del objeto por lo que se realizan otras operaciones antes de tener un resultado usable.
- **Flujo óptico.** Utiliza vectores para detectar regiones en movimiento aun cuando la cámara no está fija. Con este enfoque es posible obtener un conocimiento completo sobre el movimiento del objeto. No obstante,

tiene una elevada complejidad computacional y es muy sensible al ruido.

Algunos de los métodos más comunes que utilizan el enfoque de sustracción de fondo son los siguientes (Piccardi, 2004):

- Promedio gaussiano continuo. Pretende modelar cada píxel del fondo de manera independiente utilizando una función de densidad de probabilidad (PDF, por sus siglas en inglés). Con cada nuevo cuadro se calcula el promedio y la desviación estándar de la PDF utilizando una media ponderada entre los valores de la imagen actual y de la PDF anterior. El movimiento se detecta calculando la diferencia entre el promedio de la PDF de cada píxel y el valor actual del píxel.
- Filtro mediano temporal. Se diferencia del promedio gaussiano continuo ya que, en lugar de usar una media ponderada, utiliza una mediana de cada píxel de las imágenes más recientes para calcular las nuevas PDFs. Su desventaja es que requiere mantener en memoria las imágenes anteriores más recientes.
- Mezcla de gaussianos (MOG). Este método permite la detección en situaciones donde es posible que diferentes objetos formen parte del fondo en diferentes instantes, por ejemplo, nieve, lluvia, olas o árboles cuando hay vientos fuertes. Calcula la probabilidad de observar cierto valor en un píxel utilizando entre 3 y 5 distribuciones gaussianas. Cada una de las distribuciones representa a un objeto del fondo o en movimiento. Para distinguir a los objetos, se supone que los objetos del fondo son los que presentan las distribuciones más altas y compactas.
- Estimación de densidad de kernel (KDE). Se basa en el método no paramétrico para calcular PDFs, el cual tiene el mismo nombre. Calcula la PDF del fondo utilizando los fondos detectados más recientemente, generalmente se usan alrededor de 100. Cada uno genera un kernel gaussiano y se asume que todos tienen la misma varianza. Una de sus

desventajas es que requiere almacenar en memoria la cantidad de fondos a utilizar.

### 3.2. Modelo de mezcla gaussiana

Un modelo de mezcla gaussiana (GMM) es una función de densidad de probabilidad paramétrica representada como una suma ponderada de las densidades de componentes gaussianas. Los GMM se usan comúnmente como un modelo paramétrico de la distribución de probabilidad de mediciones continuas o características en un sistema biométrico, como las características espectrales relacionadas con el tracto vocal en un sistema de reconocimiento de hablante. Los parámetros de GMM se estiman a partir de los datos de entrenamiento utilizando el algoritmo iterativo de expectativa-maximización (EM) o la estimación de máximo a posteriori (MAP) de un modelo previo bien entrenado.

Un modelo de mezcla de Gauss es una suma ponderada de las densidades gaussianas de los  $M$  componentes que se da por la ecuación (Stauffer & Grimson, 1999)

$$p(x|\lambda) = \sum_{i=1}^M w_i g(x|\mu_i, \Sigma_i), \quad (3.1)$$

donde  $x$  es un vector de datos de valor continuo  $D$ -dimensional (es decir, medición o características),  $w_i$ ,  $i = 1, \dots, M$ , son los pesos de la mezcla, y  $g(x|\mu_i, \Sigma_i)$ ,  $i = 1, \dots, M$ , son las densidades gaussianas componentes. Cada densidad de componente es una función gaussiana  $D$ -variable de la forma, (Stauffer & Grimson, 1999)

$$g(x|\mu_i, \Sigma_i) = \frac{1}{(2\pi)^{D/2} |\Sigma_i|^{1/2}} \exp \left\{ -\frac{1}{2} (x - \mu_i)' \Sigma_i^{-1} (x - \mu_i) \right\} \quad (3.2)$$

con el vector medio  $\mu_i$  y la matriz de covarianza  $\Sigma_i$ . Los pesos de la mezcla satisfacen la restricción de que  $\sum_{i=1}^M w_i = 1$ . El modelo completo de la mezcla gaussiana está parametrizado por los vectores medios, matrices de covarianza y la mezcla de los pesos de todas las densidades de los componentes. Estos

parámetros están representados colectivamente por la notación, (Stauffer & Grimson, 1999)

$$\lambda = \{w_i, \mu_i, \Sigma_i\} \quad i = 1, \dots, M. \quad (3.3)$$

Hay varias variantes en el GMM que se muestran en la ecuación (3). Las matrices de covarianza,  $\Sigma_i$ , pueden ser de rango completo o restringidas para ser diagonales. Además, los parámetros se pueden compartir o vincular entre los componentes gaussianos, como tener una matriz de covarianza común para todos los componentes. La elección de la configuración del modelo (número de componentes, matrices de covarianza diagonales o completas y vinculación de parámetros) a menudo está determinada por cantidad de datos disponibles para estimar los parámetros GMM y cómo se usa el GMM en una aplicación biométrica particular.

También es importante tener en cuenta que, dado que el componente gaussiano está actuando conjuntamente para modelar la densidad de características general, las matrices de covarianza completas no son necesarias, incluso si las características no son estadísticamente independientes. La combinación lineal de diagonal covarianza base Gaussianas es capaz de modelar las correlaciones entre los elementos del vector de características. El efecto de utilizar un conjunto de  $M$  matrices de covarianza completa de gaussianos se puede obtener igualmente utilizando un conjunto mayor de gaussianos de covarianza diagonal.

### 3.3. Modelo del fondo

En visión por computadoras la modelación del fondo se refiere a estimar una imagen o estimar estadísticos del fondo de una escena que aparece en una secuencia de imágenes o video. La forma más sencilla de modelar el fondo es tomar una imagen de la escena sin objetos en movimiento y entonces utilizar esta imagen como modelo de fondo. Los objetos que no se encuentren en la imagen tomada no pertenecen al fondo y pueden ser

determinados por medio de la diferencia absoluta de cuadros, esto consiste en comparar cada píxel de la secuencia actual con la imagen estimada del fondo.

Esa solución debería de ser suficiente en ambientes controlados, pero en condiciones no controladas como escenas al aire libre, las condiciones de iluminación varían con el tiempo, además que podemos estar en presencia de un fondo cambiante como árboles en movimiento por el viento, caída de lluvia o nieve. Por lo que se hace necesario utilizar un modelo de fondo que se adapte a la escena. El primer modelo de estimación de fondos fue propuesto por (Stauffer & Grimson, 1999).

El método para la modelación del fondo utilizado en este trabajo fue introducido por [Zivkovic, 2004], que extiende del trabajo propuesto por (Stauffer & Grimson, 1999).

### 3.3.1. Entendiendo el modelo de fondo de mezcla de gaussianas

Considérese los valores de un píxel en particular a lo largo del tiempo como un "proceso de píxeles". El "proceso de píxeles" es una serie temporal de valores de píxeles, ejemplo, escalares para valores grises o vectores para imágenes en color. En cualquier momento,  $t$ , lo que se sabe sobre un píxel en particular,  $\{x_0, y_0\}$ , es su historia.

$$\{X_1, \dots, X_t\} = \{I(x_0, y_0, i) : 1 \leq i \leq t\} \quad (3.4)$$

En la práctica, la iluminación en la escena puede cambiar gradualmente (durante el día o las condiciones climáticas en una escena al aire libre) o repentinamente (cambiar la luz en una escena interior). Un nuevo objeto podría entrar en la escena o un objeto presente ser eliminado de la escena. Para adaptarse a los cambios, se puede actualizar el conjunto de entrenamiento agregando nuevas muestras y descartando las antiguas. Se elige un período de tiempo razonable  $T$  y en el momento  $t$  tenemos  $X_T = \{x^{(t)}, \dots, x^{(x-T)}\}$ . Para cada nueva muestra, se actualiza el conjunto de datos de entrenamiento  $X_T$  y se reestima el modelo de fondo  $\hat{p}(\vec{x}|X_T, BG)$ . Sin

embargo, entre las muestras de la historia reciente podría haber algunos valores que pertenecen a los objetos en primer plano y deberíamos denotar esta estimación como  $\hat{p}(\vec{x}|X_T, BG + FG)$ . Se utiliza GMM con  $M$  componentes

$$\hat{p}(\vec{x}|X_T, BG + FG) = \sum_{m=1}^M \hat{\pi}_m \mathcal{N}(\vec{x}; \hat{\mu}_m, \hat{\sigma}_m^2 I) \quad (3.5)$$

donde  $\hat{\mu}_1, \dots, \hat{\mu}_m$  son las estimaciones de las medias y  $\hat{\sigma}_1^2, \dots, \hat{\sigma}_m^2$  son las estimaciones de las varianzas que describen los componentes gaussianos. Se asume que las matrices de covarianza son diagonales y la matriz de identidad  $I$  tiene las dimensiones adecuadas. Los pesos de mezcla indicados por  $\hat{\pi}_m$  no son negativos. Dada una nueva muestra de datos  $\vec{x}^{(t)}$  en el momento  $t$ , las ecuaciones de actualización recursiva son:

$$\hat{\pi}_m \leftarrow \hat{\pi}_m + \alpha(o_m^{(t)} - \hat{\pi}_m) \quad (3.6)$$

$$\hat{\mu}_m \leftarrow \hat{\mu}_m + o_m^{(t)} (\alpha / \hat{\pi}_m) \vec{\delta}_m \quad (3.7)$$

$$\hat{\sigma}_m^2 \leftarrow \hat{\sigma}_m^2 + o_m^{(t)} (\alpha / \hat{\pi}_m) (\vec{\delta}_m^T \vec{\delta}_m - \hat{\sigma}_m^2) \quad (3.8)$$

donde  $\vec{\delta}_m = \vec{x}^{(t)} - \hat{\mu}_m$ . En lugar del intervalo de tiempo  $T$  que se mencionó anteriormente, aquí la constante  $\alpha$  describe una cubierta de caída exponencial que se usa para limitar la influencia de los datos antiguos. Mantenemos la misma notación teniendo en cuenta que aproximadamente  $\alpha = 1/T$ . Para una nueva muestra la propiedad  $o_m^{(t)}$  se establece en 1 para el componente "cercano" con la mayor  $\hat{\pi}_m$  y los demás se establecen en cero. De hecho, la muestra está "cercana" a un componente si la distancia de Mahalanobis (Rivero Perez, Ribeiro, & Morell Perez, 2016) del componente es, por ejemplo, menos de tres desviaciones estándar. La distancias de Mahalanobis se calcula utilizando la función  $d_m(\vec{x}, \vec{y}) = \sqrt{(\vec{x} - \vec{y})^T + \Sigma^{-1}(\vec{x} - \vec{y})}$ . La distancia al cuadrado desde el componente  $m$ -th se calcula como:  $D_m^2(\vec{x}^{(t)}) = \vec{\delta}_m^T \vec{\delta}_m / \hat{\sigma}_m^2$ . Si no hay componentes "cercaños" un nuevo componente es generado con  $\hat{\pi}_{M+1} = \alpha, \hat{\mu}_{M+1} = \vec{x}^{(t)}$  y  $\hat{\sigma}_{M+1} = \sigma_0$  donde  $\sigma_0$  es una varianza inicial apropiada.

Si el máximo número de componentes es alcanzado se eliminar el componente con el menor  $\hat{\pi}_m$ .

El algoritmo presenta un algoritmo de agrupamiento en línea. Por lo general, los objetos intrusos de primer plano se representarán mediante algunos grupos adicionales con pesos pequeños  $\hat{\pi}_m$ . Por lo tanto, podemos aproximar el modelo de fondo por los primeros grupos B más grandes:

$$\hat{p}(\vec{x}|X_T, BG) = \sum_{m=1}^B \hat{\pi}_m \mathcal{N}(\vec{x}; \hat{\mu}_m, \hat{\sigma}_m^2 I) \quad (3.9)$$

Si los componentes están ordenados para tener pesos descendentes  $\hat{\pi}_m$  tenemos:

$$B = \arg \min_b (\sum_{m=1}^b \hat{\pi}_m > (1 - c_f)) \quad (3.10)$$

donde  $c_f$  es una medida de la porción máxima de los datos que pueden pertenecer a objetos de primer plano sin influir en el modelo de fondo. Por ejemplo, si un nuevo objeto entra en una escena y permanece estático durante algún tiempo, probablemente generará un clúster estable adicional. Dado que el fondo anterior está ocluido, el peso  $\pi_{B+1}$  del nuevo grupo aumentará constantemente. Si el objeto permanece estático el tiempo suficiente, su peso se hace más grande que  $c_f$  y se puede considerar como parte del fondo. Si observamos (6), podemos concluir que el objeto debería estar estático por aproximadamente  $\log(1 - c_f)/\log(1 - \alpha)$  cuadros. Por ejemplo  $c_f = 0.1$  y  $\alpha = 0.001$  obtenemos 105 cuadros.

### 3.3.1.1. Seleccionando el número de componentes

El peso  $\hat{\pi}_m$  describe la cantidad de datos que pertenecen al componente m-ésimo de la GMM. Puede considerarse como la probabilidad de que una muestra provenga del componente m-ésimo y de esta manera los  $\hat{\pi}_m$ -s definen una distribución multinomial subyacente. Supongamos que tenemos  $t$  muestras de datos y cada uno de ellos pertenece a uno de los componentes de la GMM. Supongamos también que el número de muestras que pertenecen al componente m-ésimo es  $n_m = \sum_{i=1}^t o_m^{(i)}$  donde  $o_m^{(i)}$ -s se definen en la sección anterior.

### 3.4. Cámara RGB

Una cámara equipada con un sensor a través del cual se adquieren imágenes coloreadas de personas y objetos. La adquisición de imágenes, generalmente se expresa en megapíxeles (por ejemplo, 12MP, 16MP) que definen la cantidad de píxeles (es decir, longitud x altura) que componen una foto.

### 3.5. Cámara TOF

El **tiempo de vuelo** o **TOF** (del inglés *time-of-flight*) es una técnica utilizada para estimar distancias de cuerpos calculando el tiempo transcurrido entre la emisión y la recepción de un haz de luz infrarrojo. Este impulso está modulado en 20 MHz para que sea posible distinguirlo de la luz de fondo (S, Hakan, & Cyrus, 2004).

Se utiliza, hoy en día, en aparatos de adquisición de imágenes, como escáneres o cámaras. De esta forma, es posible adquirir imágenes y secuencias en 3D. Es una técnica bastante sencilla de aplicar y, por lo tanto, económicamente asequible a la hora de fabricar aparatos de obtención de profundidad de escenarios o cuerpos.

### 3.6. Histograma

Un histograma es una representación precisa de la distribución de datos numéricos. Es una estimación de la distribución de probabilidad de una variable continua (variable cuantitativa) y fue presentada por primera vez por Karl Pearson. (Pearson, 2011) Es un tipo de gráfico de barras. Para construir un histograma, el primer paso es "separar" el rango de valores, es decir, dividir todo el rango de valores en una serie de intervalos, y luego contar cuántos valores caen en cada intervalo. Los contenedores generalmente se especifican como intervalos consecutivos sin superposición de una variable. Los contenedores (intervalos) deben ser adyacentes, y suelen ser (pero no es obligatorio) del mismo tamaño.

En visión por computadoras el histograma de imagen es un tipo de histograma que actúa como una representación gráfica de la distribución de color en una imagen digital. Traza el número de píxeles para cada valor de color. Al mirar el histograma de una imagen específica, un espectador podrá juzgar toda la distribución tonal de un vistazo (Mikhail, Konstantinos, & Andrew, 2008).

Dirección General de Bibliotecas UAQ

## Capítulo 4

### 4. Metodología

En este capítulo se muestra la metodología de esta investigación. La cual consiste en encontrar un método que detecte los objetos en movimiento antes cambios. El ciclo de la metodología se describe a continuación.

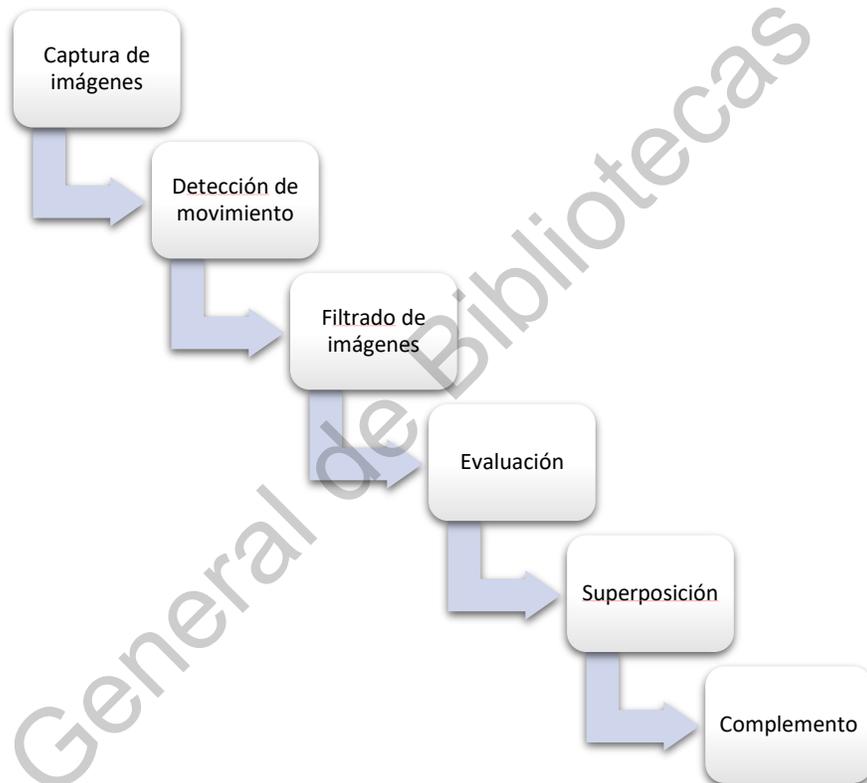


Figura 1 Metodología utilizada para la segmentación de movimiento

#### 4.1. Captura de imágenes

La adquisición de las imágenes es una etapa muy importante en cualquier proceso de tratamiento de imágenes, dependiendo de muchos factores como iluminación, cámara utilizada para la captura, amplitud del sensor, distancia y superficie sobre la que se graba la escena. Para la adquisición de las imágenes se utilizó una cámara TOF DS525 (SoftKinetics), la cual se muestra en la Figura 2. Es una cámara para interacción cercana en 3D que se puede utilizar hasta distancias de 3 m con una resolución de

320x240 de entre 25-30 fps, combinada con un sensor RGB de alta definición que permite hasta 720p.



*Figura 2 Cámara TOF DS 525*

La captura de las imágenes se realizó en un ambiente controlado, con iluminación artificial, a una altura aproximada de 250cm desde el suelo en posición cenital, obteniéndose imágenes simultáneas del sensor RGB y de profundidad; durante la captura se realizaron varios cambios de iluminación para tener diferentes escenarios y ver el comportamiento de la propuesta de solución antes cambios lumínicos. Para la obtención de las imágenes se combinó OpenCV en su versión 4.5.1 con la DepthSenseSDK en la versión 1.9.0 para Windows 64 compilado con Visual Studio 2013 (Sony DepthSensing Solutions, n.d.).

### **4.2. Detección de movimiento**

La detección de movimiento se obtuvo utilizando el algoritmo propuesto por (Zivkovic Z. , 2004) descrito en el capítulo anterior, utilizando 500 cuadros como historia, 25 como umbral para segmentar el primer plano del fondo para las imágenes capturadas con el sensor RGB y para las imágenes obtenidas con el sensor de profundidad se utilizó la misma cantidad de cuadros para la historia, pero un umbral de 40. La salida de este proceso es una imagen binaria donde los píxeles con valores de 255 representan movimiento y los píxeles con valores de 0 representan el fondo.

### **4.3. Filtrado de imágenes**

El filtrado de las imágenes es el conjunto de técnicas que se aplican a las imágenes digitales con el objetivo de mejorar la calidad o facilitar la

búsqueda de información. Es una parte muy importante en cualquier método de detección de objetos en movimiento, ya que permite una mejor segmentación y/o eliminar ruido existente después de realizar la segmentación.

En orden de obtener una segmentación más definida luego de realizar la detección de las áreas con movimiento se realiza dos operaciones morfológicas. Una operación morfológica normalmente se realiza en imágenes binarias. Necesita dos entradas, una es nuestra imagen original, la segunda se llama elemento de estructuración o *kernel* que decide la naturaleza de la operación. Primero se realiza una apertura morfológica con elemento estructurante en forma de elipse con un tamaño de 4 píxeles de ancho y de alto para imágenes obtenidas luego de realizar la detección de movimiento en las imágenes RGB de entrada. La apertura se encarga, como se observa en la Figura 3, de eliminar pequeños grupos de píxeles con valores de 255 que puedan existir en la imagen de entrada. A la imagen de salida de este paso se le realiza una cerradura morfológica con un elemento estructurante en forma de elipse con 14 píxeles de ancho por 10 de alto, esta técnica se utiliza para eliminar grupos de píxeles con valor igual a 0 que puedan estar dentro de las áreas detectadas como movimiento.

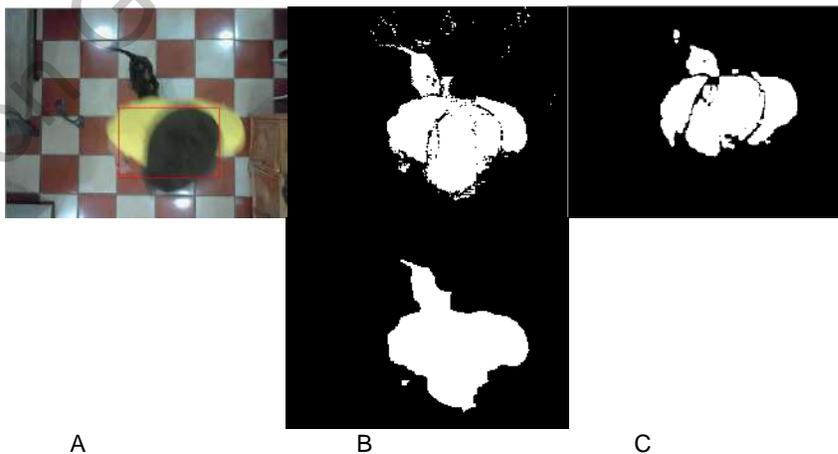


Figura 3 Filtros aplicados a la imagen después de aplicar el método de Zivkovic.

A imagen de entrada, B salida del método de detección de movimiento, C salida de aplicar apertura, D salida de aplicar cerradura morfológica.

A las imágenes de salida de la detección de movimiento de con las imágenes de entrada del sensor de profundidad también se le aplicaron filtros morfológicos, ya que también existe ruido la salida de este proceso. Se aplicaron los mismos filtros y en el mismo orden que se aplicó en RGB, pero con la diferencia de que en el caso de la cerradura se utilizó el mismo elemento estructurante, pero con tamaño diferente, 6 píxeles de ancho y 6 de alto (ver Figura 4).

Luego de realizado este proceso de filtrado se puede proceder al siguiente paso que consiste en evaluar cada una de las áreas donde se segmento el movimiento.



Figura 4 Filtros aplicados a la imagen después de aplicar el método de Zivkovic a la imagen de entrada del sensor de profundidad.

A imagen de entrada, B salida del método de detección de movimiento, C salida de aplicar apertura, D salida de aplicar cerradura morfológica.

#### 4.4. Evaluación

Esta fase consiste en evaluar cada una de las áreas donde se detectó movimiento en el en las imágenes de entrada del sensor RGB con el objetivo de conocer si estamos en presencia de una correcta detección de movimiento, para esto, primero se encontró el contorno de cada una de las formas que se encontraban en la salida del paso anterior. Los contornos se pueden explicar simplemente como una curva que une todos los puntos continuos (a lo largo del límite), que tienen el mismo color o intensidad. Los contornos son una herramienta útil para el análisis de formas y la detección y reconocimiento de objetos. Después de obtenido los contornos se tomaron aquellos objetos que

estuvieran formados por de 50 pixeles, con el objetivo de solo evaluar objetos de mayor importancia y eliminar posibles ruidos existentes aun en la imagen.

A cada área de movimiento mayor a 50 pixeles se le calcula su histograma, este histograma se calcula en la imagen de entrada del sensor RGB, el histograma obtenido se compara con dos histogramas almacenados previamente, estos histogramas guardados pertenecen a imágenes capturadas por el sensor saturado como se muestra en la Figura 5, donde normalmente es difícil hacer la detección de movimiento, ya sea porque existe mucha iluminación o hay muy poca iluminación. Con referencia a este punto, el uso de histogramas se refiere a la situación de tal manera que el promedio de energía percibida por la cámara se mueve hacia posiciones altas o bajas del rango, es decir, las condiciones de luz se convierten en oscuridad o mucha luz para ello. La comparación se realiza utilizando la correlación de Pearson (Pearson, 2011).

$$r = \frac{\Sigma(x-m_x)(y-m_y)}{\sqrt{\Sigma(x-m_x)^2\Sigma(y-m_y)^2}} \quad (4.1)$$

Donde  $m_x$  y  $m_y$  son las medias de las variables  $x$  y  $y$ , el rango del coeficiente de correlación de Pearson esta dado de -1 a 1, donde 1 es la correlación lineal positiva total, 0 no hay correlación lineal, y -1 es la correlación lineal negativa total.

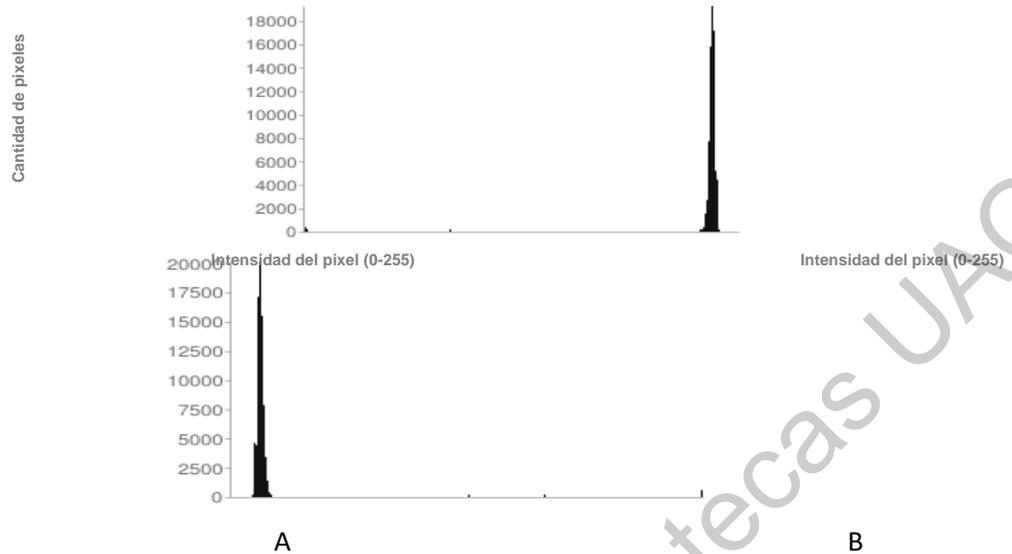


Figura 5 Histogramas guardados.  
A imagen con mucha iluminación, B imagen muy oscura

Para determinar si el movimiento detectado en RGB no se debió a un cambio repentino en la iluminación, la correlación entre el histograma calculado y alguno de los guardados debe ser menos que 0.7. Las áreas con menos de 0.7 correlación se determinarán como si la detección de movimiento en esa área en ese instante de tiempo del tiempo fuera correcta.

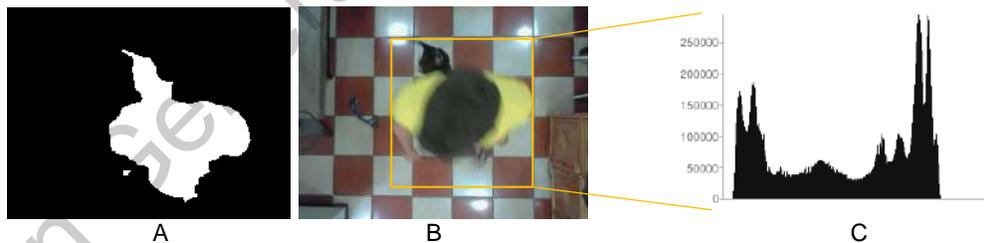


Figura 6 Proceso de obtener área donde se detectó movimiento y cálculo de su histograma.  
(A) imagen con movimiento detectado, (B) segmentación del área donde se detectó el movimiento, (C) histograma del área segmentada

## 4.5. Superposición

En presencia de una buena segmentación en el canal RGB, el canal de profundidad se segmenta en las mismas posiciones donde se detectó el movimiento en el canal RGB y se analiza su información para determinar si también se detectó movimiento en el canal de profundidad, ya que debido a que el sensor está en posición cenital puede que se genere más ruido debido a la reflexión de la iluminación sobre la superficie de fondo. Por lo que lo

anterior se realiza con la meta de obtener una segmentación más robusta en el sensor RGB ante cambios lumínicos, eliminando sombras y falsas detecciones de movimiento

Después, se calcula el centro de cada objeto en movimiento detectado que haya pasado la prueba de correlación para almacenarlo en una matriz denotada por  $F$ .

#### **4.6. Complemento**

Una vez que se encontraron los objetos en movimiento en el canal RGB, se precedió para analizar los objetos en movimiento detectados en el canal de profundidad, a fin de determinar los objetos no detectados en el sensor RGB y agregarlos a la segmentación. Se realizó una prueba de correlación, pero en este caso, se utilizó 0.2 como umbral para descartar posibles detecciones falsas. Una vez que se determinaron los objetos en movimiento, se segmentaron y se realizó una superposición de cada objeto en la matriz  $F$ , para determinar qué objetos ya se habían detectado en el canal RGB y solo agregar los que no se habían detectado previamente.

## Capítulo 5

### 5. Modelo experimental y resultados

#### 5.1. Modelo Experimental

A continuación se presentan los resultados experimentales obtenidos tras el análisis del algoritmo presentado en la sección anterior, se valida la propuesta analizando los resultados de salida de varios métodos de segmentación de movimiento.

Se identificaron las siguientes desventajas de utilizar experimentación capturando video en vivo:

- Después de cada prueba se debe realizar la selección imagen por imagen de cuales pixeles corresponden a movimiento.
- En cada prueba la entrada es diferente, dificultando encontrar errores en la implementación.

Con el fin de simplificar el proceso de evaluación de resultados, se optó por utilizar secuencias de imágenes grabadas en un escenario controlado que tuviera las diferentes escenas necesarias para validar la solución, El conjunto de datos contiene 3 categorías: 1) situación normal, 2) camuflaje de color y 3) camuflaje de profundidad. Cada escena consta de secuencias de imágenes de color y profundidad como se muestra en la figura 7.

1. Situación normal: no hay ninguna situación de color o profundidad de camuflaje.
2. Camuflaje de color: existen regiones de camuflaje de color deseadas, donde un objeto de interés tiene el mismo color que el fondo.
3. Profundidad camuflaje: existen regiones de camuflaje de profundidad, cuando un objeto de interés tiene la misma profundidad (es decir, la distancia desde el sensor de profundidad) con el fondo.

El diseño experimental consistió en procesar simultáneamente dos secuencias de 430 imágenes, una en escala de grises para la información obtenida del

## MODELO EXPERIMENTAL Y RESULTADOS

sensor de profundidad y la otra en RGB, con un tamaño de 320 por 240 pixeles cada imagen de cada secuencia, para la cual se cuenta con la referencia. El enfoque usado para la evaluación es a nivel pixel mediante la precisión, la sensibilidad y la medida F (David, 2011). Primero se obtienen los verdaderos positivos (TP), falsos positivos (FP), verdaderos negativos (TN) y falsos negativos (FN). Después la precisión P se obtiene mediante:

$$P = \frac{TP}{TP + FP} \quad (5.1)$$

y la sensibilidad R se calcula de la siguiente manera:

$$R = \frac{TP}{TP + FN} \quad (5.2)$$

por último, una métrica de precisión, F, que combina precisión y la sensibilidad para evaluar la calidad de la segmentación. La medida F se define de la siguiente manera:

$$F = \frac{2P \cdot R}{P + R} \quad (5.3)$$

### 5.2. Resultados

Una muestra de las imágenes obtenidas se observa en la Figura 7, Los valores de precisión (P) y sensibilidad (R) calculados con estas muestras de imágenes se exponen en la Tabla 1, En esta tabla también se presentan los valores promedio de precisión, sensibilidad y la medida F en la secuencia. En estas tablas también se demuestran los valores promedio de precisión y sensibilidad en la secuencia. Al calcular estos valores promedio, se excluyeron las imágenes que tenían poco o nada de movimiento ya que resultaban en valores extremos. La distribución de los valores de medida F, precisión y sensibilidad en la secuencia se puede observar en los graficas presentadas en las figuras 8,9 y 10, respectivamente. También se muestra los valores obtenidos de la medida F, precisión y recuerdo al ejecutar el algoritmo solamente con el sensor de distancias y al ejecutarlos solamente con el sensor RGB.

#	Entrada RGB	Entrada de profundidad	Referencia	Detección de Movimiento en RGB	Detección de Movimiento en profundidad	Salida del algoritmo
---	-------------	------------------------	------------	--------------------------------	--	----------------------

## MODELO EXPERIMENTAL Y RESULTADOS

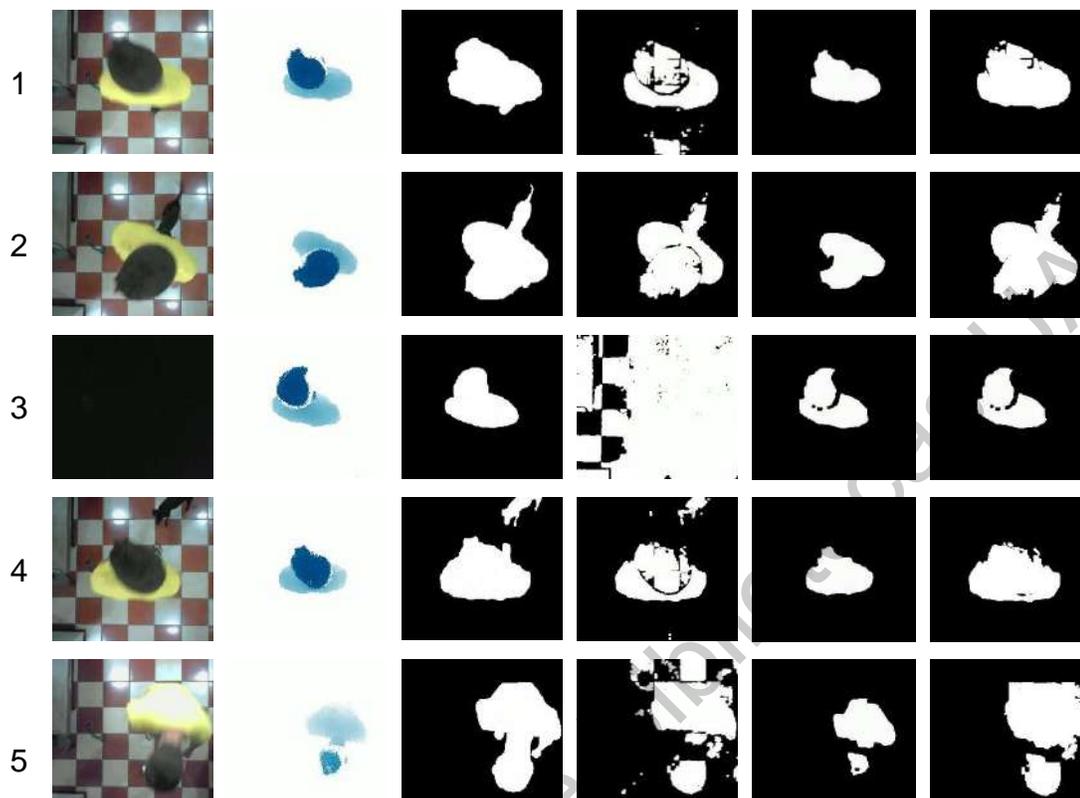


Figura 7 Resultados obtenidos de procesar algunas imágenes con el método propuesto

Tabla 1 Resultados al procesar las imágenes de la Figura 7

#	TP	FP	FN	TN	Precisión	Sensibilidad	Medida F
1	11885	578	1521	62816	0.95	0.88	0.91
2	14999	1492	733	59576	0.90	0.95	0.93
3	8250	113	510	67927	0.98	0.94	0.96
4	11108	131	3738	61823	0.98	0.74	0.85
5	14966	2757	1857	57220	0.84	0.88	0.86
<b>Promedio de la secuencia</b>					<b>0.93</b>	<b>0.87</b>	<b>0.90</b>

### 5.3. Discusión de los resultados

Con la metodología desarrollada, se procede a validar el planteamiento de la hipótesis, la cual propone que se pueden obtener una mejor segmentación

## MODELO EXPERIMENTAL Y RESULTADOS

ante cambios lumínicos graduales permitiendo hacer un mejor seguimiento de objetos en movimiento.

Examinando los resultados de las figuras 8, 9 y 10 se observa que los valores de la medida F, precisión y sensibilidad para la secuencia en general los valores en algunos instantes son muy bajos, esto se pueden atribuir a que la estimación de fondo retiene poca información sobre el movimiento anterior y detecta más movimiento del que realmente está sucediendo debido a los cambios lumínicos y las sombras. Además, en los instantes finales no detecta parte del movimiento debido a ser una superficie donde el sensor de distancias no rebota de forma correcta y nos es capaz de determinar la diferencia de distancias. En las gráficas el eje de  $x$  representa el número de cuadro y el eje  $y$  representa el valor de la variable se esté representando (medida F, precisión o sensibilidad) y sus valores están dados de 0 a 1.

En la Figura 10 se puede percibir que el método propuesto muestra los mejores resultados, dando por correcta la combinación de la información obtenida de los sensores de distancia y RGB. Esto permitió contrarrestar el ruido existente en cada sensor, especialmente el sensor RGB, ya que los cambios lumínicos y las sombras lo saturan con facilidad.

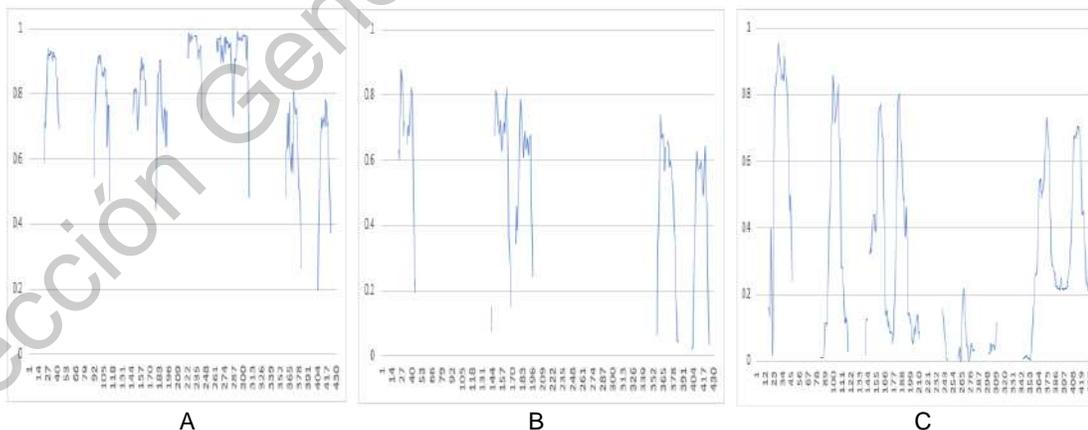


Figura 8 Valores de Medida F (F) obtenidos a lo largo de la secuencia.

A salida del método propuesto, B método propuesto utilizando profundidad, C método propuesto utilizando RGB.

Examinando los resultados de la Figura 9 se observa que los valores de precisión para la secuencia que la utilización de los dos sensores de forma

## MODELO EXPERIMENTAL Y RESULTADOS

simultánea en el método arroja mejores resultados que la utilización de cualquiera de los dos por separado. A pesar de que en algunos instantes el sensor de profundidad por si solo muestra mejores resultados, en otros instantes no detecta el movimiento.

Al analizar los resultados obtenidos en los valores de recuerdo de la Figura 10 se puede distinguir que la gráfica (a), muestra los mejores resultados, en cada instante en que ocurre el movimiento donde los valores mayormente están en un rango de 0.6 a .0.95, mientras que en las gráficas (b) y (c), los promedios son mucho más bajos.

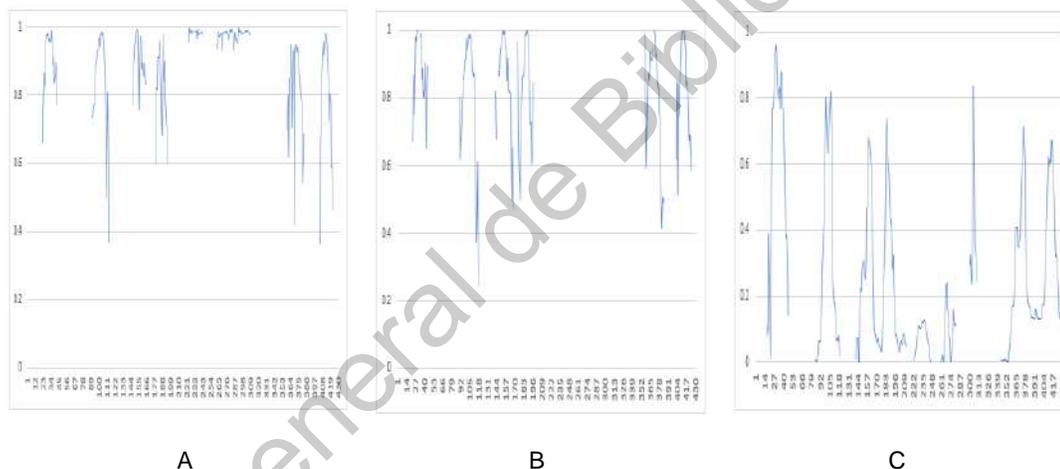


Figura 9 Valores de precisión (P) obtenidos a lo largo de la secuencia.  
A salida del método propuesto, B método propuesto utilizando profundidad, C método propuesto utilizando RGB.

# MODELO EXPERIMENTAL Y RESULTADOS

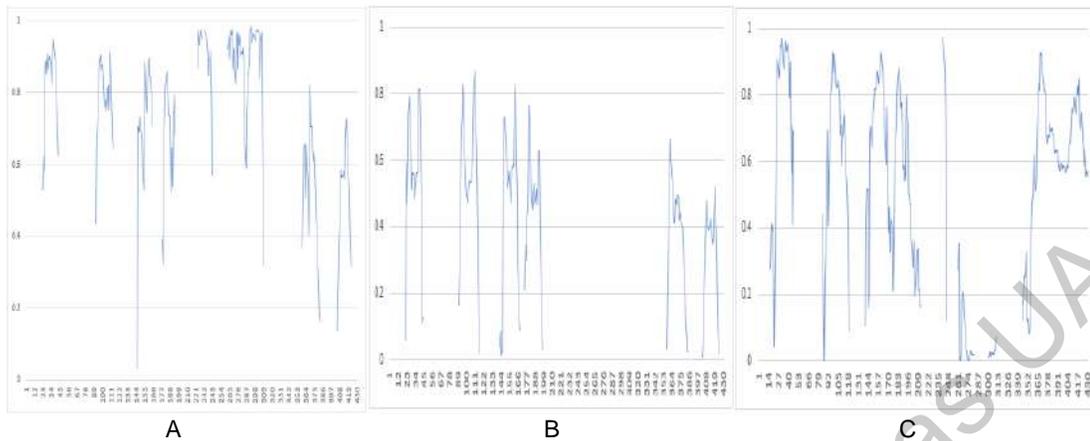


Figura 10 Valores de recuerdo (R) obtenidos a lo largo de la secuencia.  
A salida del método propuesto, B método propuesto utilizando profundidad, C método propuesto utilizando RGB.

Dirección General de Bibliotecas UAQ

### 6. Conclusiones

Esta tesis presentó un método de segmentación de movimiento desarrollado e implementado para trabajar en condiciones de cambios lumínicos. Demostrando a través de las pruebas realizadas que se obtienen buenos resultados en escenarios donde exista camuflaje de color debidos a cambios de iluminación o donde existan objetos en movimiento con colores similares al fondo, donde a otros métodos (Camplani & Salgado, 2014) (Cucchiara, Piccardi, & Prati, October 2003) (Del Pizzo, Foggia, Greco, Percannella, & Vento, 2016) (Zivkovic Z. , 2004) de segmentación se les dificulta su segmentación.

Como conclusiones se tiene que a través de los resultados obtenidos en el capítulo anterior se valida la hipótesis planteada, ya que se demuestra que el método propuesto es capaz de obtener una mejor segmentación de los objetos en movimiento permitiendo que se pueda hacer un mejor seguimiento.

Trabajos futuros

- Mejorar el método mediante la incorporación de otros sensores.
- Incorporar *Deep Learning* para que el método sea capaz de aprender y hacer mejor detección.

### 7. Referencias bibliográficas

- Camplani, M., & Salgado, L. (2014). Background foreground segmentation with RGB-D Kinect data: An efficient combination of classifiers. *Journal of Visual Communication and Image Representation*.
- Cucchiara, R., Piccardi, M., & Prati, A. (October 2003). Detecting moving objects, ghosts, and shadows in video streams. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24, 1337-1342.
- David, M. (2011). Evaluation: from precision, recall and f-measure to ROC, informedness, markedness and correlation. *International Journal of Machine Learning Technology*, 2, 37.
- Del Pizzo, L., Foggia, P., Greco, A., Percannella, G., & Vento, M. (2016). Counting people by RGB or depth overhead cameras. *Pattern Recognition Letters*, 81, 41-52.
- Fernandez-Sanchez, E., Rubio, L., Diaz, J., & Ros, E. (2013). Background subtraction model based on color and depth cues. *Machine Vision and Applications*, 1211-1225.
- Harville, M., Gordon, G., & Woodfill, J. (2001). Foreground segmentation using adaptive mixture models in color and depth. *Proceedings of the IEEE Workshop on Detection and Recognition of Events in Video*, 311.
- Kumar Chauhan, A., & Krishan, P. (2013). Moving Object Tracking using Gaussian Mixture Model and Optical Flow. *International Journal of Advanced Research in Computer Science and Software Engineering*, 3, 243-246.
- Kumar, D., & Singh, K. (2016). A combined approach of Kullback–Leibler divergence and background subtraction for moving object detection in thermal video. *Infrared Physics & Technology*, 76, 21-31.
- Maddalena, L., & Petrosino, A. (March 2010). A fuzzy spatial coherence-based approach to background/foreground separation for moving object detection. *Neural Computing and Applications*, 19, 179-186.
- Maddalena, L., & Petrosino, A. (July 2008). A self-organizing approach to background subtraction for visual surveillance applications. *IEEE Transactions on Image Processing*, 17, 1168-1177.
- Mikhail, S., Konstantinos, D., & Andrew, H. (2008). Histogram-based search: A comparative study. *2008 IEEE Conference on Computer Vision and Pattern Recognition*.

## REFERENCIAS BIBLIOGRÁFICAS

---

- Nayagam, M., & Ramar, K. (2015). A survey on Real time Object Detection and Tracking Algorithms. *International Journal of Applied Engineering Research*.
- Noh, S., & Jeon, M. (2012). A new framework for background subtraction using multiple cues. *Asian Conference on Computer Vision*.
- Pearson, K. (2011). Contributions to the Mathematical Theory of Evolution. II. Skew Variation in Homogeneous Material. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 186, 343–414.
- Piccardi, M. (2004). Background subtraction techniques: a review. *IEEE International Conference on Systems, Man and Cybernetics*, 4, pp. 3099-3104.
- Rivero Perez, J., Ribeiro, B., & Morell Perez, C. (2016). Mahalanobis distance metric learning algorithm for instance-based data stream classification. *2016 International Joint Conference on Neural Networks (IJCNN)*.
- S, B., Hakan, Y., & Cyrus, B. (2004). A Time-Of-Flight Depth Sensor – System Description, Issues and Solutions. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*.
- SoftKinetics. (n.d.). Retrieved from [https://www.sony-depthsensing.com/Portals/0/Download/WEB\\_20120907\\_SK\\_DS325\\_Datasheet\\_V2.1.pdf](https://www.sony-depthsensing.com/Portals/0/Download/WEB_20120907_SK_DS325_Datasheet_V2.1.pdf)
- Song, Y.-m., Noh, S., Yu, J., Park, C.-w., & Lee, B.-g. (2015). Background subtraction based on Gaussian mixture models using color and depth information. *The International Conference on Control, Automation and Information Sciences*.
- Song, Y.-m., Noh, S., Yu, J., Park, C.-w., & Lee, B.-g. (2015). Background subtraction based on Gaussian mixture models using color and depth information. *ICCAIS*.
- Sony DepthSensing Solutions. (n.d.). Retrieved from <https://www.sony-depthsensing.com/Default.aspx?tabid=457&error=An%20unexpected%20error%20has%20occurred&content=0>
- Stauffer, C., & Grimson, W. (1999). Adaptive background mixture models for real-time tracking. *Proc IEEE Conf on Comp Vision and Pattern*, (pp. 246-252).
- Yao, J., & Odobez, J.-M. (June 2007). Multi-Layer Background Subtraction Based on Color and Texture. *Proc IEEE Conf on Comp Vision and Pattern Recognition*.

## REFERENCIAS BIBLIOGRÁFICAS

---

- Yin, J., Liu, L., Li, H., & Liu, Q. (2016). The infrared moving object detection and security detection related algorithms based on W4 and frame difference. *Infrared Physics & Technology*, 77, 302-315.
- Yiyang, L., Longqing, S., Yuanbing, Z., & Yue, L. (2017). Individual pig object detection algorithm based on Gaussian mixture model. *International Journal of Agricultural and Biological Engineering*, 10, 186–193.
- Zhang, B., Zhong, B., & Cao, Y. (2011). Complex background modeling based on Texture Pattern Flow with adaptive threshold propagation. *J. Visual Commun. Image Represent*, 516–521.
- Zhang, Q., & Ngan, K. (2010). Multi-view video based multiple objects segmentation using graph cut and spatiotemporal projections. *J. Visual Commun. Image Represent*, 453-461.
- Zivkovic, Z. (2004). Improved Adaptive Gaussian Mixture Model for Background Subtraction. *ICPR*, 3.
- Zivkovic, Z., & van der Heijden, F. (2006). Efficient adaptive density estimation per image pixel. *Pattern Recognition Letters*, 27, 773-780.