

 Institución Universitaria	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-27

Procesamiento de video para rastreo de personas, usando información de profundidad de una cámara Time-of-Flight posicionada en una vista aérea.

Cristián Suárez Acosta

Mateo Mejía Herrera

Ingeniería Electrónica

Ingeniería Mecatrónica

Dr. Sergio D. Cabrera

Dr. Jorge Herrera Ramírez

INSTITUTO TECNOLÓGICO METROPOLITANO

30 de agosto de 2018

	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

RESUMEN

Este trabajo se enfoca en el procesamiento de imágenes adquiridas por una cámara con tecnología Time-of-Flight incorporada en un dispositivo Kinect V2; y cuyo objetivo es recolectar información óptima y confiable para su posterior interpretación en el seguimiento de personas en una escena.

La metodología utilizada para este proyecto se desarrolla sobre el software MATLAB y se ejecuta dos etapas: la primera es el procesamiento de la imagen que conlleva las tareas de almacenamiento de la información, la substracción del fondo, la eliminación del ruido por medio de procesos como el inpainting y ACWMF; la segunda etapa es el rastreo de objetos en movimiento con un algoritmo basado en Multiple Object Tracking proporcionado por MATLAB. En este trabajo no se llega una etapa de identificación y diferenciación de seres humanos, esta parte se deja para trabajos futuros.

Los resultados del análisis y procesamiento de las imágenes han permitido realizar la detección de objetos en movimiento sobre una escena. El trabajo evidencia que es recomendable el uso de imágenes de profundidad, debido a la cantidad reducida de datos, y que permite la monitorización del tráfico de personas sin importar cambios de iluminación en el ambiente. *Palabras clave:* Palabras técnicas significativas que identifiquen el trabajo desarrollado.

Palabras clave: Procesamiento de imágenes, Kinect V2, MATLAB, imágenes de profundidad, rastreo.

	<p style="text-align: center;">INFORME FINAL DE TRABAJO DE GRADO</p>	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

RECONOCIMIENTOS

Agradecimientos a la profesora María Constanza Torres por brindarnos la maravillosa oportunidad de iniciar este proyecto con todo el apoyo y las facilidades requeridas, a María Antonia López contratista de la oficina de relaciones nacionales e internacionales por su excelente gestión en los trámites relacionados con nuestra movilidad, a nuestros padres por hacernos las personas inquietas que somos, a nuestros compañeros por brindarnos el apoyo moral y el acompañamiento personal para hacer esta experiencia única, a el Dr. Sergio D. Cabrera y a su estudiante de posgrado Basavarajaiah Shanmukhayya Totada por el conocimiento brindado y el tiempo dedicado, al profesor Jorge Herrera Ramírez por su dedicación y disposición para ayudarnos y facilitar nuestro proceso de aprendizaje e investigación, y finalmente pero no menos importante a la señora rectora de nuestra institución María Victoria Mejía y en general a toda la comunidad Itemista por el esfuerzo y por cumplir un sueño que sin ustedes no hubiese sido posible, de corazón gracias.

	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

ACRÓNIMOS

Tof Time-of-Flight

RGB Red, Green, blue

FPS Frames per second

HSV Hue, Sature, Value

HSI Hue, Sature, Intensity

ACWMF Adaptative Center-Weighted Median Filter

MP4 Media Player 4

JPEG Joint Photographic Experts Group

ROI Region of Interest

LED Light-Emitting Diode

 Institución Universitaria	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

TABLA DE CONTENIDO

1. INTRODUCCIÓN.....	6
2. MARCO TEÓRICO	9
3. METODOLOGÍA	17
4. RESULTADOS Y DISCUSIÓN.....	29
5. CONCLUSIONES, RECOMENDACIONES Y TRABAJO FUTURO.....	32
REFERENCIAS	34
APÉNDICE.....	36

	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

1. INTRODUCCIÓN

El seguimiento-detección de gestos o acciones humanas es importante para la recopilación y estudio de su comportamiento. En los últimos años se ha popularizado el uso de cámaras digitales para aplicaciones de inspección y vigilancia, esto ha creado la necesidad de estudiar problemas asociados a la información adquirida para el desarrollo de sistemas autónomos que contribuyan al monitoreo de multitudes para prevención y análisis de desastres o estudios de mercadeo y turismo.

Este proyecto hace parte de los esfuerzos que permiten la monitorización responsable del tráfico humano en ambientes como negocios, estaciones de transporte, hospitales, entre otros; para aplicaciones de edificios inteligentes, esto contribuye al desarrollo de “Smart cities” las cuales permiten la creación de información útil en el estudio del comportamiento humano.

El seguimiento/conteo de personas con el fin de analizar el flujo peatonal es objeto de estudio en diversas investigaciones y en general se puede clasificar en tres etapas: procesamiento de la imagen, rastreo de objetos en movimiento y finalmente identificación y diferenciación de seres humanos. En busca de reducir problemas de superposición entre los objetos en movimiento se ha planteado el uso de cámaras desde una vista superior (Crnojevi, 2009) (Cai, Yu, Liu, & Zhang, 2014), esto para facilitar la etapa de rastreo. Sin embargo, debido a la cantidad de información la tarea se hace más compleja en cámaras de color, por ello Breiman y Chom et. al. (Breiman, 2001) (Cho, Chow, & Leung, 1999), plantean una solución usando modelos inteligentes. Si bien estos modelos pueden dar resultado, debido a temas legales, se busca que la identidad de las personas en la escena no sea revelada. Así es que una propuesta útil que preserva la identidad de las personas es el uso, desde una vista superior, de cámaras de profundidad de tipo tiempo de vuelo o ToF

 Institución Universitaria	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

(Bevilacqua, Di Stefano, & Azzari, 2006) como el Kinect V2 (Stahlschmidt, Gavriilidis, Velten, & Kummert, 2013). Usando este tipo de tecnología, (Zhang et al., 2012) proponen el uso de mínimos locales para la detección de la cabeza, ya que éste es el punto más cercano del cuerpo respecto a la cámara y es una característica de estudio importante. Además, la cabeza es un factor determinante para la detección de personas, aunque no es lo único que se debe tener en cuenta. En (Stahlschmidt, Gavriilidis, Velten, & Kummert, 2016) se plantea la detección de personas basándose en el modelo gaussiano del sombrero mexicano para identificar las mismas, pero este enfoque no es suficiente para casos en los que las personas sostienen objetos sobre la cabeza. De este modo, en el proceso de seguimiento de personas podemos encontrar dos problemas recurrentes: la detección de personas con accesorios en la cabeza y la distinción de individuos dentro de un grupo de personas.

En este trabajo se estudian los problemas relacionados con el procesamiento de las imágenes y el rastreo de objetos en movimiento, lo que corresponde, como se había descrito, a las dos primeras etapas típicas de un proyecto en el seguimiento/conteo de personas. Por tanto, acá se tendrá en cuenta la clasificación de ruido, eliminación del fondo así como el almacenamiento y el ajuste de resolución de la información adquirida con un Kinect V2 para el rastreo de objetos en movimiento dentro de la escena. Adicionalmente, también se plantean posibles soluciones a los problemas presentados en la identificación de personas, que es la parte que corresponde a la tercera etapa de este tipo de proyectos. Así, y de acuerdo a lo anterior, se plantean los siguientes objetivos para este trabajo:

Objetivo general:

Hallar por medio de operaciones de procesamiento de imágenes digitales información óptima y confiable para el seguimiento de personas dentro de una escena usando imágenes obtenidas con una cámara Kinect v2.

	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

Objetivos específicos:

- Almacenar información adquirida con el *Kinect V2* a través del software Matlab para esta investigación y posible creación de bases de datos.
- Procesar videos usando técnicas de eliminación del ruido y del fondo, que contribuyan a la adquisición de datos útiles para el seguimiento de personas.
- Identificar objetos en movimiento dentro de la escena usando algoritmos de Matlab para conocer el comportamiento del flujo peatonal en el área de estudio.

Este trabajo presenta la aplicación de metodologías sobre información adquirida por medio de un Kinect V2 en cuanto al ajuste de su resolución y pre-procesamiento por medio de algoritmos simples en Matlab. Así, se plantea un método experimental para el ajuste de resolución de ambas cámaras del Kinect V2, con sus resultados, y también se hace la segmentación de distancias y color en HSV para la eliminación del fondo en imágenes de profundidad y color, respectivamente. Finalmente se analiza el comportamiento de los objetos en movimiento dentro de la escena, sus problemas y algunos métodos propuestos para la tercera etapa del proyecto. Los resultados dejan en evidencia los problemas causados por el ruido de las cámaras ToF, la efectividad de procesos como el blind inpainting en situaciones con ruido gaussiano y las perspectivas iniciales del rastreo de objetos en movimiento.

	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

2. MARCO TEÓRICO

Dado que el eje central de este proyecto se encuentra enfocado en el procesamiento de imágenes, se requiere de una serie de pasos para adquirir los datos y que puedan ser reconocidos por una computadora. Por ello, en este apartado se exponen algunos de los conceptos y procesos a los cuales se somete una imagen para su reconocimiento y su procesamiento.

Imágenes

Una imagen es una matriz conformada por un conjunto de unidades mínimas conocidas como píxeles, es decir, estos píxeles son cada una de las casillas de la imagen. En una imagen a color, cada uno de estos píxeles contiene los datos de representación de color para unas coordenadas espaciales específicas (x, y) . Existen diversos modelos para representar los colores que hay en una imagen como son: HSV (Hue, Saturation, Value), HSI (Hue, Saturation, Intensity) o RGB (Red, Green, Blue), entre otros, donde estas letras representan cada uno de los componentes que hay en cada píxel. Entre estos esquemas de representación del color, el RGB es el más comúnmente usado. Por esto, cada píxel de una imagen a color contiene la información separada en 3 canales o componentes, que combinados de manera adecuada dan el color específico de ese píxel. La codificación más usual de las intensidades de estos 3 canales se da en un esquema de 8 bits por canal, es decir, 24 bits en total por píxel. Así, cada componente de color RGB toma 256 valores en un rango de 0 a 255.

También se pueden encontrar imágenes en escala de grises, que son aquellas que están conformadas de un solo canal codificado regularmente en 8 bits, es decir, representan los niveles de gris en el rango de 0(negro) a 255(blanco). Entre sus utilidades está que tienen menor peso en cuanto a consumo de memoria, más adelante se hablará acerca de lo

	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

efectivas que son este tipo de imágenes. Por último, se encuentran las imágenes binarias las cuales solo se componen por valores de 0 que representan el color negro y valores de 1 representando el color blanco. La representación matemática de una escena con valores de intensidad $I(x, y)$, en una imagen digital $I(k, l)$ de $M \times N$ píxeles, sería:

$$I(x, y) \approx I(k, l) = \begin{bmatrix} I(1,1) & \cdots & I(1, N) \\ \vdots & \ddots & \vdots \\ I(M, 1) & \cdots & I(M, N) \end{bmatrix} \quad 1$$

Donde cada píxel $I(k, l)$ estaría asociado al muestreo espacial de las coordenadas (x, y) y puede representar la intensidad de un color, una escala de gris o un valor binario, dependiendo del tipo de imagen (Mendieta, 2003).

Kinect V2

El dispositivo Kinect V2 fue lanzado por Microsoft en julio de 2014, originalmente construido para su uso en consolas de video juegos, permite la interacción entre usuario-consola por medio del cuerpo y la voz. El dispositivo Kinect V2 permite obtener en tiempo real la imagen a color con una resolución de 1920x1080 píxeles, una imagen de profundidad con resolución de 512x424 y audio de cualquier escena (Bloesch, Rodriguez, Kaestner, Hutter, & Siegwart, 2015). Este dispositivo se ha convertido en una herramienta muy útil debido a que cuenta con una cámara RGB y una cámara de profundidad con sensores infrarrojos, basada en la técnica Time-of-Flight, que permite capturar imágenes en completa oscuridad, con un campo de visión de 70.6° x 60°, un rango de medición de 0.5 hasta 4.5 metros; desde su distribución en el mercado este dispositivo ha sido muy apetecido debido a las características integradas que hacen su comportamiento similar a las cámaras web, gracias a su diseño, éste permite ser instalado en cualquier sitio para realizar estudios sobre diversos ambientes (Fig. 1).

	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22



Figura 1 Diseño del Kinect V2 y ubicación de las cámaras y sensores. (Tomado de (Bloesch et al., 2015)).

Tecnología Time-of-Flight (ToF)

La tecnología Time-of-Flight (ToF) es una de las técnicas más utilizadas para la adquisición de imágenes de profundidad. En esta se estima la distancia desde la cámara hasta los objetos de una escena por medio de la emisión y reflexión de luz infrarroja. Es decir, que cada valor de los píxeles describen la distancia desde un punto correspondiente en el mundo real hasta el sensor (Stahlschmidt et al., 2016). Cada uno de los píxeles de la imagen codifica la distancia a la cual corresponde cada uno de los puntos sobre la escena, por medio de una onda de infrarrojo que se dirige hacia el objetivo, el sensor detecta el componente infrarrojo reflejado, es decir, al medir la diferencia de fase entre las dos ondas, irradiada y reflejada, se puede calcular la distancia del objeto, como se observa en la siguiente figura:

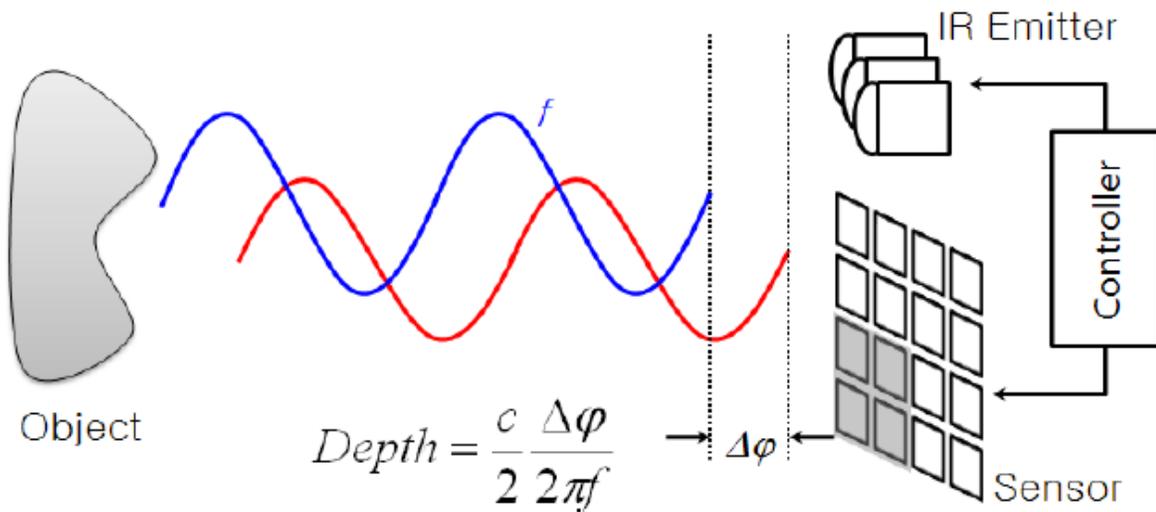


Figura 2. Comportamiento de la luz infrarroja emitida (roja) y reflejada (azul) (Tomado de (Hansard et al., 2012)).

Sin embargo, la precisión puede ser limitada debido a la potencia de la señal del infrarrojo puesto que la luz del día generada por el sol, también emiten en el espectro de luz infrarroja que interfieren con esta señal, aunque también puede influir el material del objeto debido a que puede interferir en el momento de reflexión de la luz infrarroja haciendo que la intensidad de esta luz sea baja y no pueda ser detectada por el sensor del equipo.

La iluminación creada por los sistemas que utilizan la técnica de Time-of-Flight viene de un diodo emisor de luz (LED). Esta iluminación está en el rango del infrarrojo cercano por lo que es invisible al ojo humano, pero puede ser confundida con la luz del medio ambiente que también contiene este tipo de radiación, debido a esto se genera ruido que puede alterar la lectura de los datos de la escena tomada.

El ruido es uno de los factores fundamentales con los cuales se debe tratar a la hora de adquirir imágenes a través de una cámara, especialmente en la imagen de profundidad; algunos ruidos son ocasionados por pequeñas fugas eléctricas por los circuitos que componen el dispositivo, haciendo que los píxeles generen una señal diferente y la señal resultante no sea uniforme; los ruidos se presentan de tipo impulsivo y no impulsivo, pueden ser corregidos con filtros adaptativos encargados de preservar los detalles mientras

	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

suaviza el ruido sobre imagen, trabajo que un filtro tradicional no hace (Garnett, Huegerich, Chui, & He, 2005).

Algoritmo y software

Los algoritmos son un conjunto de operaciones sistemáticas que permiten hallar la solución a diversos tipos de problemas, el procesamiento de imágenes es uno de ellos y por ende la visión por computador ha sido una de las técnicas en las que se utiliza el análisis de imágenes digitales mediante algoritmos. Así, se requiere de un software capaz de realizar este tipo de procesos, de los más utilizados en la comunidad científica e ingenieril es MATLAB. Con este software se pueden resolver algunos problemas gracias a los Toolbox de procesamiento de imágenes que posee. Este conjunto de librerías están diseñadas para cumplir con diversas tareas de procesamiento de imágenes. Este software permite acceder a cada uno de los componentes que contiene una imagen, y aplicar a éstos diferentes funciones como filtros, operaciones morfológicas, operaciones de segmentación, entre muchas otras. No obstante, es posible encontrar que para algunos casos el tiempo de procesamiento para cumplir con una tarea deseada puede ser demasiado largo, aunque esto depende de las características (Hardware) con las que cuente el equipo encargado del procesamiento (Valdemar & Jimenez, 2007).

Las operaciones morfológicas es una de las técnicas dentro del procesamiento de imágenes, acá se consta de operaciones que pueden conservar características principales de los objetos, suprimir ruidos, ayudar en la descripción de objetos o simplificar formas. Por ejemplo, la operación morfológica de dilatación se encarga de adicionar píxeles al contorno de los objetos que se encuentran en la imagen y la erosión tiene la función contraria de eliminar el contorno de los objetos. Dentro de Matlab varias de estas operaciones ya están implementadas y hacen parte de los Toolbox del programa. Otras rutinas usadas para este proyecto fueron las relacionadas con el Motion-Based Multiple Object Tracking, para la detección de movimientos sobre una escena.

	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

Los recursos con los que se contaban para realizar todas las actividades de procesamiento fueron suficientes para un óptimo desempeño y desarrollo de las técnicas de visión por computador, en nuestro caso se utilizó una laptop ASUS ROG GL552VW con un procesador Intel Core i7 6700HQ, sistema operativo Windows 10 Pro, memoria RAM DDR4 de 8GB, un HDD de 1TB, una tarjeta gráfica integrada de 4GB y una tarjeta gráfica NVIDIA GeForce GTX 960M de 6GB.

Investigaciones previas

Durante los últimos años se han desarrollado trabajos para el conteo de personas en un ambiente controlado (Del Pizzo, Foggia, Greco, Percannella, & Vento, 2016); (Luna et al., 2017); (Stahlschmidt et al., 2016); (Zhang et al., 2012). Tales sitios como: casas, edificios, centros comerciales, entre otros, han hecho que el uso de cámaras de vigilancia se convierta en una forma de seguridad en ambientes donde el flujo de personas es constante. Sin embargo, con el pasar del tiempo se está haciendo más énfasis en el uso de técnicas para mejorar las imágenes que se obtienen. El uso de cámaras de tecnología ToF ha sido fundamental para aplicaciones mucho más concretas que buscan resultados más precisos, por ejemplo el conteo de personas que están pasando por algún sitio en específico (Del Pizzo et al., 2016). En el trabajo de (Del Pizzo et al., 2016), el video es adquirido por una cámara Kinect V1 desde una vista superior obteniendo información tanto de la imagen en RGB como de profundidad. Aunque este dispositivo no cuenta con tecnología ToF, ha sido utilizado para detectar el número de personas que están cruzando a través de unos escenarios controlados. El método propuesto en ese trabajo ha sido el de instalar el Kinect V1 en un área controlada para evitar perturbaciones de la señal y obtener una imagen de referencia del fondo con el fin de hacer una substracción de éste con otra imagen en la cual se encuentra una persona. La toma de las imágenes se realiza a 30 FPS con una resolución de 640 x 480 píxeles de la cámara de color, con un ajuste de las resoluciones hasta 320 x 240 y 160 x 120 píxeles para comparar entre imágenes obtenidas con la cámara RGB, con la

	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

cámara de profundidad y una combinación de las dos. Con un análisis de sus datos llegaron a la conclusión de que las imágenes de profundidad eran las adecuadas para su posterior procesamiento. Usando un filtro Gaussiano que permite suavizar la imagen para reducir el ruido, obtuvieron buenos resultados en la detección de personas en una escena.

En el trabajo de (Stahlschmidt et al., 2016), se usan imágenes de profundidad de baja resolución desde una vista superior sobre una región de interés (ROI). A toda esta información obtenida mediante una cámara con tecnología Time-of-Flight integrada en el dispositivo Kinect V2, es necesario realizarle un pre-procesamiento para reducir el ruido. Además, en el trabajo de detección, se separa la multitud de personas en objetos individuales para facilitar la tarea del seguimiento utilizando el filtro Kalman, dicho filtro se enfoca en predecir estados futuros, basado en un espacio de estados de un sistema para estimar el estado futuro y la salida futura realizando un óptimo filtrado a la señal de salida. Este filtro es una serie de ecuaciones matemáticas que implementan un tipo predictor-corrector óptimo en el sentido que minimiza el error estimado de la covariancia cuando algunas condiciones son dadas (Bravo, Arias, & Cardenas, 2013). Independientemente de la cantidad de personas que se encuentre en la escena, se busca hacer una detección de cada persona y realizar un seguimiento a su recorrido y así, los autores comprueban lo eficiente que puede ser el filtro Kalman para esta tarea.

Por otra parte, (Zhang et al., 2012) usan el Kinect V1 también desde una vista superior para adquirir imágenes tanto en RGB como imágenes de profundidad. En este trabajo, los autores usan una técnica de filtrado llamada GMM (Gaussian Mixture Model) para remover la influencia del fondo. El fondo se representa como un modelo probabilístico dado por un conjunto de mezclas gaussianas, es decir, que todos los datos generados a partir de la imagen son una mezcla de distribuciones gaussianas. Esta mezcla modela los datos especialmente cuando provienen de distribuciones aleatorias cuyos valores se desconocen. Además, el algoritmo propuesto para reconocer las personas sobre la escena hace uso de

 Institución Universitaria	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

los mínimos locales, éstos se encuentran sobre la imagen para simbolizar los puntos más cercanos a la cámara y así identificar cada una de las personas, esta técnica permite que este algoritmo sea efectivo y robusto.

 Institución Universitaria	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

3. METODOLOGÍA

Adquisición de Imágenes y Video.

El primer paso para el desarrollo del presente proyecto es la adquisición de información de la cámara de profundidad, en este caso el *Kinect v2*. Para el uso de éste se requiere la instalación de los respectivos drivers presentes en la página del fabricante (Microsoft) para las especificaciones computacionales ya mencionadas. Haciendo uso de la herramienta de Matlab *Image Acquisition Toolbox Support Package for Kinect for Windows Sensor versión 17.1.1* se visualiza y adquiere la información otorgada por las cámaras de profundidad y de color, así es posible verificar las resoluciones de las imágenes entregadas y la calidad de éstas.

Es necesario considerar el problema computacional que representa adquirir la información en ambas cámaras durante el mismo periodo de tiempo, esto para establecer un marco comparativo entre ambas cámaras y obtener los mismos datos por medio de un ajuste de resolución del que se tratará más adelante. Para el almacenamiento de dicha información fue desarrollado un algoritmo inicial que permite conocer el tiempo de procesamiento del almacenamiento de una cantidad X de cuadros, posteriormente este es modificado para almacenar el mismo tiempo de video como se muestra en la Fig. 3.



Figura 3 Ajuste del tiempo de captura

	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

Se obtuvo que una cantidad de 600 frames con los recursos computacionales usados, se da en un tiempo alrededor de 61 segundos de video, los que se traducen en aproximadamente 9.84 FPS. Este video se almacena en un formato MP4 que permite visualizar fácilmente la información. El tiempo de procesamiento puede variar en función de la cantidad de procesos en segundo plano que puedan estar en ejecución durante el tiempo de adquisición del video.

Los videos de profundidad y color inicialmente adquiridos, los cuales serán mencionados en adelante como OrigDepth y OrigColor, se obtienen con una resolución de 512x424 y 1920x1080, respectivamente. La Fig. 4 muestra un par de frames de este tipo.

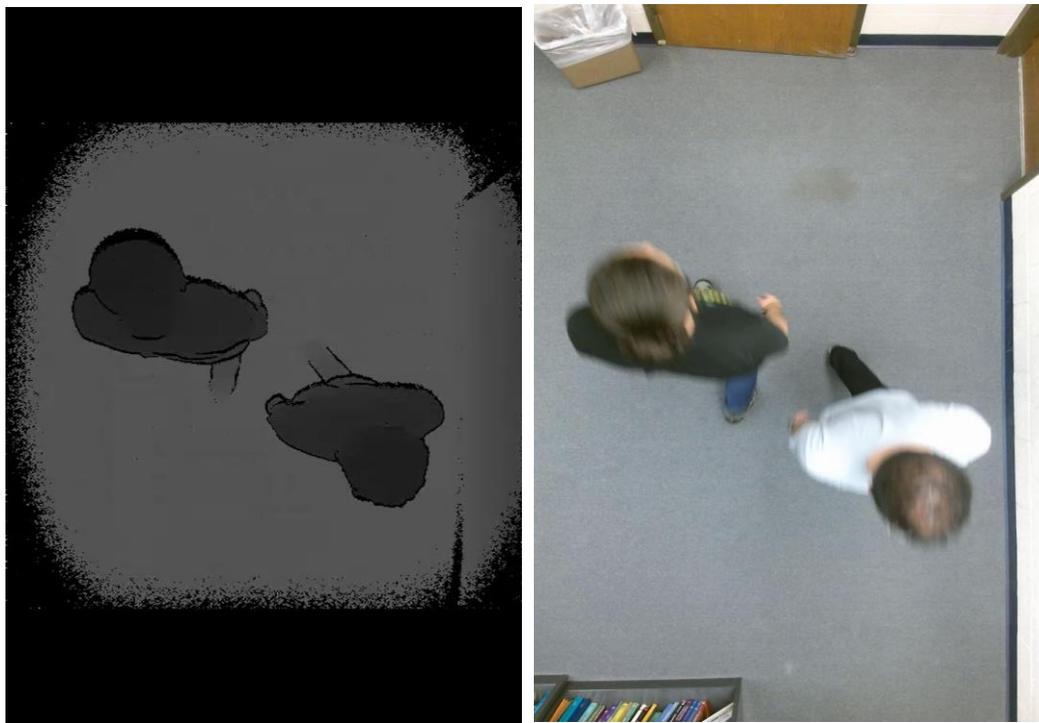


Figura 4 Imágenes originales Profundidad (izquierda) y Color (derecha).

	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

Ajuste de Resolución

Una vez almacenado el video de ambas cámaras se realiza un ajuste de la resolución, para ello se plantea un montaje experimental que permite adquirir aproximadamente la misma información de las cámaras de color y profundidad integradas en el Kinect. El montaje se diseñó para percibir puntos de interés de la matriz correspondiente a las imágenes. Esto se logró posicionando dos esferas de cuatro centímetros de diámetro en la imagen de profundidad a las distancias límites del rango de detección horizontal. La Fig. 5 muestra este montaje. En la imagen de color, las esferas se ubicaron en los límites verticales, como se puede observar en la Fig. 6.



Figura 5 Montaje para ajuste de resolución



Figura 6 Ubicación de puntos de referencia dentro de las imágenes de color y profundidad

Para la detección efectiva de los puntos de interés es necesario eliminar el fondo de ambas imágenes, esto se logra al conservar la cámara en la posición inicial para la captura de imágenes del fondo que será usada para realizar la operación de resta entre las matrices. El resultado de la resta es una matriz en la que se pueden detectar los puntos de interés, además del ruido pertinente a la tecnología ToF y sombras. Estos ruidos son eliminados por medio de morfología y binarización lo que facilita el proceso de identificación de los puntos de interés.

Usando la función *imfindcircles* del software Matlab con varios parámetros ajustados, el cual sirve para identificar formas similares a círculos en una imagen, se establece la posición de los centros de las esferas de referencia dentro de ambas imágenes (color, profundidad) (Fig. 7). Estos datos permiten establecer la zona de la escena que se encuentra común en ambas imágenes y así permite eliminar las zonas diferentes por medio de un recorte.



Figura 7 Detección de puntos de interés dentro de ambas imágenes.

	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

El paso final para el ajuste de resolución es realizar un cambio del tamaño del resultado del corte de OriginDepth. Este cambio de tamaño se logra con el comando “imresize” usando como nuevo tamaño, el tamaño de la imagen Origcolor después del recorte. Como es de esperar, la información no será exactamente la misma hasta que no se encuentre la distancia entre ambas cámaras. Sin embargo, en el actual trabajo se logra ajustar ambos videos para obtener aproximadamente la misma información con el desfase mencionado (Fig. 8).

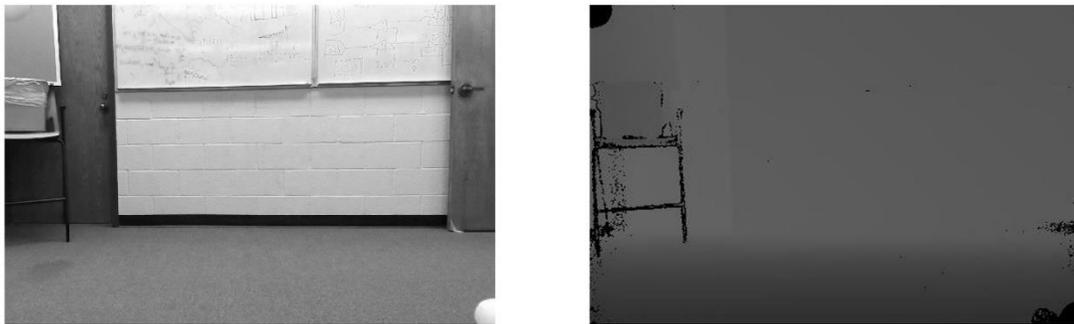


Figura 8 Imágenes tras el ajuste de resolución

Montaje físico del Kinect.

Como se plantea en un inicio de la investigación, la adquisición de la información se debe realizar desde una vista superior, no solo por la efectividad para obtener la información de posición de los objetos, sino también para conservar anónima la identidad de las personas en la escena.

La disposición del Kinect V2 fue realizada en un laboratorio con flujo peatonal controlado, si bien esto no es recomendado debido a la falta de diversidad de los datos, por motivos legales no fue posible llevar a cabo en otro ambiente.

	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

El sensor se dispuso aproximadamente a dos metros de la puerta principal del laboratorio y con una altura cercana a los 2.7 metros. La Fig.9 muestra una imagen del montaje real. Con la ayuda del soporte superior del trípode *Gear Lightweight tripod* de la marca Kodak, algunas bases metálicas y tornillería, se aseguró la calidad de las imágenes en cuanto a la estabilidad del montaje y la conservación de la direccionalidad con respecto al piso.



Figura 9 Montaje del Kinect

Pre-procesamiento

Supresión de fondo

Aunque en este proyecto se considera la segunda etapa, el procesamiento y filtrado de la información debe ser previa a la detección de objetos y en general a cualquier proceso de adquisición de información.

	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

Debido a la cantidad de datos innecesarios para la presente investigación, es de crucial importancia un proceso de supresión del fondo con el fin de determinar los objetos que aparecen en la escena. Para esto se plantea un procedimiento sencillo y usado en otras imágenes durante el ajuste de resolución. Este consiste en una operación de sustracción entre dos matrices, el cual logra establecer los elementos comunes a ambas por medio de una búsqueda de valores cercanos a ceros así:

$$X_{i,j} = \{0 : N - M < 10\} \quad 2$$

Siendo M la imagen que corresponde a la captura del fondo y N una imagen de la misma escena con objetos sobre el fondo. Se introduce un valor de variabilidad de 10 que permite cambios leves de color en la posición (i, j) y asigna valores de 0 a los pixeles ubicados en las coordenadas específicas que no se ajusten a esta premisa.

En el caso de AdjColor al realizar una resta entre los valores de referencia y los actuales, se obtiene ruido y en general no se logra sustraer bien el fondo debido a los cambios percibidos por el sensor a lo largo de la secuencia de video. Otro problema presente en este método es la detección de objetos en movimiento con los mismos valores RGB que los de M. Por esto se realiza una segmentación en el espacio de color HSV, el cual permite evaluar cambios de iluminación e intensidad, segmentando solamente las componentes H (Hue) y V (Value) con un margen de diferencia de 0.1 y 0.32, respectivamente, para evitar problemas propios de cambios de iluminación o intensidad.

	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

Aunque los resultados conservan con relativa efectividad la información perteneciente a la cabeza y los hombros de las personas también se obtiene ruido que interfiere con el proceso de detección (Fig.10), esta efectividad puede incrementar bajo ambientes controlados, ya que facilita la posibilidad de segmentar el video usando el color del fondo, para así

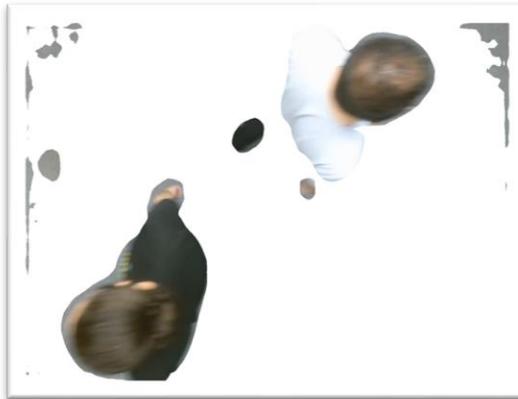


Figura 10 Eliminación del fondo usando segmentación en HSV

sustraerlo exitosamente. Incluso propuestas como (Shao-Yi Chien, Shyh-Yih Ma, & Liang-Gee Chen, 2002; Stauffer & Grimson, 1991) pueden solucionar este problema por medio de modelos probabilísticos y caracterización de píxeles, pero ese no será el enfoque del presente proyecto y se planteará para trabajos futuros.

En el caso de AdjDepth el problema se reduce, ya que la matriz que conforma la imagen es una matriz bidimensional que representa las distancias y la identificación del fondo se logra estableciendo los niveles de gris que corresponden al mismo. Estos valores varían dependiendo de la altura a la que se encuentre posicionada la cámara y el nivel de altura que se desea filtrar, al igual que un filtro pasa bajas en el algoritmo utilizado se eliminaron los valores de intensidad mayores a 86 de la imagen de profundidad esto con el fin de sustraer toda la información que se encuentre por encima de los 220 centímetros aproximadamente, esto a su vez resulta en una eliminación de algunos objetos u animales facilitando el reconocimiento de las personas en la imagen. Tras realizar esta operación se conservaba ruido en las esquinas de la imagen al igual que alrededor de la silueta de la

	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

persona, este ruido es propio de las cámaras tipo ToF y se caracteriza como ruido impulsivo. La Fig. 11 muestra el resultado del procedimiento anterior.

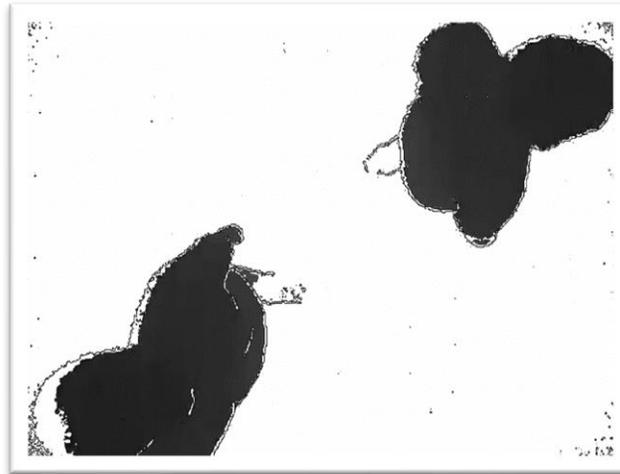


Figura 11 Eliminación del fondo imagen de profundidad.

Al comparar las imágenes de color y profundidad obtenidas tras la reducción del fondo se aprecia una mayor eficiencia en las imágenes de profundidad, ya que en algunos casos el color de la ropa u otros factores pueden alterar la reducción del fondo por medio del método utilizado en el video de color.

Eliminación del ruido

Este proceso se realiza sobre las imágenes de profundidad, debido a la efectividad del algoritmo de supresión de fondo.

Para la eliminación del ruido impulsivo previamente mencionado, se utilizaron dos tipos de filtro: el primero un filtro inpainting (Gonzalez, 2016) el cual de manera iterativa suaviza la imagen usando un filtro gaussiano de $[21,21]$ y una desviación estándar de 10 para la convolución. Finalmente reemplaza los pixeles con valores inferiores a 50 por su versión suavizada, se elige este valor debido a que todos los valores menores de 50 pueden ser considerados ruido, ya que este rango corresponde a valores muy cercanos a la cámara.

	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

Estos valores no son corresponden a la información esperada dentro de la situación planteada y son tipificados como ruido sal pimienta.

Tras obtener una versión suavizada de la imagen se procede con el método para remover el ruido impulsivo generado aleatoriamente por la cámara de profundidad. Se utiliza un filtro adaptativo de la mediana con el centro ponderado, conocido en la literatura como (ACWMF), que emplea un esquema de cambio en el cual se identifican los valores de pixeles aleatorios para después realizar el filtrado sobre estos y modificarlos sin pérdida de información como bordes o cambios de color (Fig. 13).



Figura 12 Imagen de profundidad sin ruido tras eliminación de fondo

Seguimiento de objetos en movimiento.

Una vez adquirida la imagen final sin fondo ni ruido es posible realizar un seguimiento de los diversos objetos que aparecen en movimiento dentro de los videos capturados. Dicha tarea se realiza con la ayuda del algoritmo para Matlab *Motion-Based Multiple Object Tracking*, este se encuentra como uno de los códigos adjuntos dentro del software y representa el proceso de detección de objetos en tres pasos (Mathworks, 2015):

- Sustracción del fondo de la imagen usando modelos de mezcla gaussiana o (GMM), los cuales por medio de aprendizaje de maquina detectan qué es parte del fondo y retorna una máscara binarizada.

	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

- Se realizan operaciones morfológicas sobre la máscara en busca de eliminar el ruido restante y tener los objetos dentro de la escena diferenciados de manera clara.
- Por último, se efectúa un análisis de regiones para detectar los grupos de píxeles en la imagen y conocer su posición.

El seguimiento es realizado por el algoritmo *multiObjectTracker* de Matlab el cual es responsable del seguimiento cuadro por cuadro, inicialización de las trazas para el rastreo de objetos en movimiento, actualización de estas, creación de nuevas y eliminación de las que ya no se encuentran en el cuadro. El movimiento de cada traza se predice gracias a un filtro Kalman, el cual determina las posibles posiciones de cada objeto en el siguiente cuadro.

El primer paso se ejecuta bajo las siguientes características, se utilizan nueve modelos gaussianos para la mezcla y un número de cuadros de entrenamiento igual a 40. Un mayor número de cuadros de entrenamiento facilita el proceso de detección del fondo, pero también hace necesario extender la duración de los videos captados para adquirir la información necesaria, esto es un problema relacionado a la baja densidad de FPS adquirida con los recursos computacionales actuales.

Previo a la identificación de objetos es necesario realizar la eliminación de ruido, esto se logra con operaciones morfológicas como se mencionó anteriormente, los procesos utilizados fueron los siguientes

- Apertura para formas rectangulares con tamaños 3x3
- Cierre para formas rectangulares con tamaños 15x15
- Relleno de agujeros dentro de la máscara.

El tamaño de los kernels utilizados en las primeras dos operaciones de morfología son producto de la experimentación, tras realizar un análisis del problema se realizan aperturas pequeñas a los píxeles con valores de intensidad blancos y grandes cierres sobre los valores oscuros, esto facilita obtención de un fondo uniforme que es clave para el reconocimiento de objetos.

	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

En pro de la identificación de personas se aplica un umbral durante el análisis de regiones dentro de la imagen, este permite discriminar objetos con áreas menores a los 15.000 píxeles y superiores a los 300.000 píxeles, esto no solo discrimina pequeños objetos y animales si no también sirve para identificar individuos dentro de grupos en movimiento. La información del seguimiento obtenida se muestra sobre el video en recuadros con la etiqueta del objeto identificado Fig.14, esto nos da una región de interés sobre la que se pueden realizar futuros procesos de diferenciación entre personas y objetos en movimientos.



Figura 13 Objetos en movimiento dentro de la escena

	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

4. RESULTADOS Y DISCUSIÓN

Se logró almacenar con fidelidad las imágenes obtenidas (fig. 4) por el sensor tanto a color como en imágenes de profundidad, eventualmente se pueden obtener imágenes con una mayor densidad de FPS haciendo uso de software de Microsoft, pero se requieren en ese caso mayores recursos computacionales para realizar el procesamiento y la adquisición de la información en archivos independientes. Ambos videos, RGB y profundidad, poseen ruido y su almacenamiento debe hacerse de forma específica para evitar fallas en el software Matlab o un reinicio inesperado del sensor, lo que resulta en pérdida de información.

El proceso de ajuste de resolución se llevó a cabo de manera efectiva. Se obtuvo el mismo tamaño en ambas imágenes al finalizar el proceso, si bien se presenta un desfase entre los videos RGB y profundidad, estos son aptos para realizar las operaciones necesarias para la identificación de movimiento. Se debe resaltar que, por problemas ambientales en el caso del video RGB, se dificulta el proceso de eliminación de fondo por medio de métodos aritméticos simples u operaciones de filtrado.

La supresión del fondo en el video de profundidad fue efectiva, si bien queda un ruido remanente, éste es un ruido simple que se puede tratar con operaciones morfológicas (Fig.15).

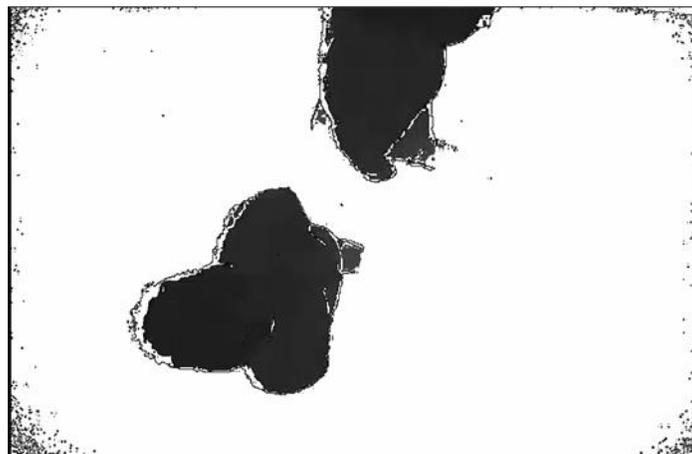


Figura 14 Ruido tras la supresión de fondo en imágenes extraídas del video de profundidad

	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

En el video a color la supresión de fondo logra conservar las características de la cabeza. Para este caso se trataron otros métodos como comparación en RGB o la creación de máscaras haciendo uso de la información de profundidad, para la segmentación de objetos en movimiento. De estas pruebas diversas, el que conserva con mayor efectividad características principales como la cabeza es el implementado por medio de HSV (Fig.16)

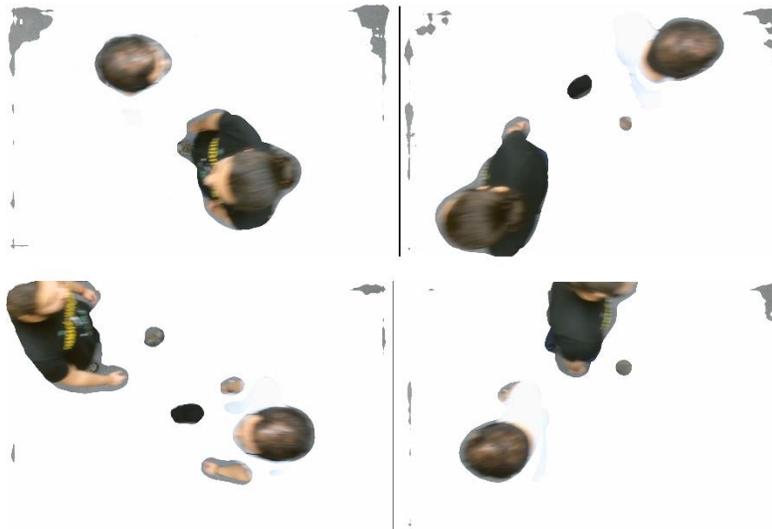


Figura 16. Resultado de supresión de fondo en imágenes RGB

Los procesos de filtrado utilizados suprimen el ruido dentro del video de profundidad en todos los cuadros en los que se presentan objetos diferentes al fondo. Sin embargo, cuando se presentan imágenes solo del fondo existe gran cantidad de ruido, este comportamiento es propio del algoritmo usado y se debe estudiar a profundidad para resolverlo. De todas maneras, en escenas donde se requiere hacer el rastreo, el proceso de inpaiting y ACWMF logran separar la información relevante. En estudios como (Zhang et al., 2012) plantean la detección de la cabeza por mínimos locales, con los resultados obtenidos todo el proceso se complica debido a que se obtiene una barrera de mínimos que rodea la mayoría de la sección en movimiento (Fig.17).

	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22



Figura 17.. Mínimos presentes en el perfil del objeto

El Algoritmo de detección de objetos en movimiento, presenta falsos positivos sobre todo en situaciones con un gran número de personas en la escena, y en ocasiones toma varios frames para identificar el movimiento de un individuo. Por tanto, se requiere estudiar más a profundidad el proceso de detección de fondo realizado durante el seguimiento y así mejorar los resultados.

Con respecto al uso del filtro Kalman durante la predicción de posiciones, los identificadores de movimiento se pierden una vez el objeto rastreado sale de la escena, incluso en algunos casos el filtro logra predecir un nuevo objeto en el cuadro como si se tratara de otro que se hubiera presentado con anterioridad.

La identificación de personas si bien no era el objeto principal de estudio en este proyecto era la finalidad última del mismo, durante el proceso se obtuvo éxito en la tarea de rastreo de personas y discriminación de algunos objetos. Sin embargo, es necesario usar enfoques como los propuestos en (Galčík & Gargalík, 2013; Vera, Monjaraz, & Salas, 2016; Zhu & Wong, 2013) para lograr un mejor acercamiento al problema y obtener una verdadera identificación de seres humanos por medio de la detección de características dentro de la región de interés.

 Institución Universitaria	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

5. CONCLUSIONES, RECOMENDACIONES Y TRABAJO FUTURO

- Se pudo corregir el desfase que existe entre la cámara RGB y profundidad para obtener la misma información de una escena. Sin embargo, sigue existiendo una separación entre estas, por lo que esta corrección más precisa se plantea como propuesta para trabajos futuros.
- Mejorando la potencia computacional es posible obtener una mayor densidad de FPS, esto facilita el procesamiento de la imagen para la reducción del background por métodos de aprendizaje de máquina, como los que se plantean en (Cho et al., 1999).
- Se puede obtener información más relevante de la estructura de cabeza-hombros al aumentar la distancia de la cámara con respecto a los transeúntes, puesto que la diferenciación de personas/objetos, se puede mejorar haciendo uso de características morfológicas como la posición de los hombros, entre otras.
- La información de la cámara de profundidad sin fondo se puede utilizar durante el proceso de supresión de ruido en el espacio de color.
- La segmentación en espacio HSV para la imagen de color, es un procedimiento que facilita la tarea de reducción de fondo, conservando características importantes para el estudio. Aunque éste no tiene confiabilidad total y en general se conserva ruido difícil de tratar por métodos tradicionales.
- Es necesario aplicar morfología y filtrado gaussiano para mejorar los resultados. El uso del inpainting complementado con ACWMF elimina efectivamente el ruido que altera el proceso de detección.
- El proceso de detección de personas se puede realizar haciendo uso de los recuadros que enmarcan los objetos en movimiento para evitar recorrer toda la imagen.

	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

- El uso de algoritmos básicos para la detección de personas no ha dado resultados concluyentes, por lo que es ideal aplicar enfoques de inteligencia artificial y aprendizaje de máquina para refinar el proceso.
- La información de RGB puede ser útil durante el proceso de detección de personas. Sin embargo, no es absolutamente necesaria, se evidencia una mayor facilidad de la ejecución del proceso sobre las imágenes de profundidad.
- El proceso de ajuste de resolución logra reducir gran cantidad del ruido presente en las imágenes del sensor Time Of Flight al recortar las esquinas de la imagen, que es donde mayor densidad de ruido hay.

 Institución Universitaria	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

REFERENCIAS

- Bevilacqua, A., Di Stefano, L., & Azzari, P. (2006). People tracking using a Time-of-Flight depth sensor. *IEEE Fifth International Conference on Advanced Video and Signal Based Surveillance*. <https://doi.org/10.1109/AVSS.2006.92>
- Bloesch, M., Rodriguez, D., Kaestner, R., Hutter, M., & Siegwart, R. (2015). Kinect v2 for Mobile Robot Navigation Evaluation and Modeling Péter Fankhauser, (August). <https://doi.org/10.1109/ICAR.2015.7251485>
- Bravo, V. A. O., Arias, M. A. Ni., & Cardenas, J. A. C. (2013). Analisis Y Aplicación Del Filtro De Kalman a Una Señal Con Ruido Aleatorio. *Scientia et Technica*, 18(1), 267–274. <https://doi.org/10.22517/23447214.8241>
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
- Cai, Z., Yu, Z. L., Liu, H., & Zhang, K. (2014). Counting people in crowded scenes by video analyzing. *Proceedings of the 2014 9th IEEE Conference on Industrial Electronics and Applications, ICIEA 2014*, (1), 1841–1845. <https://doi.org/10.1109/ICIEA.2014.6931467>
- Cho, S. Y., Chow, T. W. S., & Leung, C. T. (1999). A neural-based crowd estimation by hybrid global learning algorithm. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 29(4), 535–541. <https://doi.org/10.1109/3477.775269>
- Crnojevi, V. (2009). K-MEANS BASED SEGMENTATION FOR REAL-TIME ZENITHAL PEOPLE COUNTING. *Office*, 2565–2568.
- Del Pizzo, L., Foggia, P., Greco, A., Percannella, G., & Vento, M. (2016). Counting people by RGB or depth overhead cameras. *Pattern Recognition Letters*, 81, 41–50. <https://doi.org/10.1016/j.patrec.2016.05.033>
- Galčík, F., & Gargalík, R. (2013). Real-Time Depth Map Based People Counting. In J. Blanc-Talon, A. Kasinski, W. Philips, D. Popescu, & P. Scheunders (Eds.), *Advanced Concepts for Intelligent Vision Systems* (pp. 330–341). Cham: Springer International Publishing.
- Garnett, R., Huegerich, T., Chui, C., & He, W. (2005). A universal noise removal algorithm with an impulse detector. *IEEE Transactions on Image Processing*, 14(11), 1747–1754. <https://doi.org/10.1109/TIP.2005.857261>
- Gonzalez, R. C. (2016). *Digital Image Processing Using MATLAB. International Journal of Research* (Vol. 2).
- Hansard, M., Lee, S., Choi, O., Horaud, R., Hansard, M., Lee, S., ... Cameras, F. (2012). *Time of Flight Cameras : Principles , Methods , and Applications To cite this version : HAL Id : hal-00725654*.

 Institución Universitaria	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

- Luna, C. A., Losada-Gutierrez, C., Fuentes-Jimenez, D., Fernandez-Rincon, A., Mazo, M., & Macias-Guarasa, J. (2017). Robust people detection using depth information from an overhead Time-of-Flight camera. *Expert Systems with Applications*, 71, 240–256.
<https://doi.org/10.1016/j.eswa.2016.11.019>
- Mathworks. (2015). Multiple Object Tracking.
- Mendieta, D. (2003). Reconocimiento de Objetos Bidimensionales en Imágenes mediante la Transformada de Distancia utilizando Matlab, (c), 1–4. <https://doi.org/10.15713/ins.mmj.3>
- Shao-Yi Chien, Shyh-Yih Ma, & Liang-Gee Chen. (2002). Efficient moving object segmentation algorithm using background registration technique. *IEEE Transactions on Circuits and Systems for Video Technology*, 12(7), 577–586. <https://doi.org/10.1109/TCSVT.2002.800516>
- Stahlschmidt, C., Gavriilidis, A., Velten, J., & Kummert, A. (2013). People Detection and Tracking from a Top-View Position Using a Time-of-Flight Camera (pp. 213–223).
https://doi.org/10.1007/978-3-642-38559-9_19
- Stahlschmidt, C., Gavriilidis, A., Velten, J., & Kummert, A. (2016). Applications for a people detection and tracking algorithm using a time-of-flight camera. *Multimedia Tools and Applications*, 75(17), 10769–10786. <https://doi.org/10.1007/s11042-014-2260-3>
- Stauffer, C., & Grimson, W. E. L. (1991). Adaptive background mixture models for real-time tracking. In *Proceedings. 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No PR00149)* (Vol. 37, pp. 246–252). IEEE Comput. Soc.
<https://doi.org/10.1109/CVPR.1999.784637>
- Valdemar, E., & Jimenez, C. (2007). Visión por Computador utilizando MatLAB Y el Toolbox de Procesamiento Digital de Imágenes, 1–35. Retrieved from
<http://proton.ucting.udg.mx/tutorial/vision/cursovision.pdf>
- Vera, P., Monjaraz, S., & Salas, J. (2016). Counting pedestrians with a zenithal arrangement of depth cameras. *Machine Vision and Applications*, 27(2), 303–315.
<https://doi.org/10.1007/s00138-015-0739-1>
- Zhang, X., Yan, J., Feng, S., Lei, Z., Yi, D., & Li, S. Z. (2012). Water filling: Unsupervised people counting via vertical kinect sensor. *Proceedings - 2012 IEEE 9th International Conference on Advanced Video and Signal-Based Surveillance, AVSS 2012*, 215–220.
<https://doi.org/10.1109/AVSS.2012.82>
- Zhu, L., & Wong, K.-H. (2013). Human Tracking and Counting Using the KINECT Range Sensor Based on Adaboost and Kalman Filter. In G. Bebis, R. Boyle, B. Parvin, D. Koracin, B. Li, F. Porikli, ... D. Gotz (Eds.), *International Symposium on Visual Computing* (pp. 582–591). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-41939-3_57

	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

APÉNDICE

Códigos Fuente para implementación en Matlab de algunas de las tareas realizadas en este trabajo de grado:

Detect Moving Cars In Video

```

%
%%
% Create system objects to read file.
videoSource = vision.VideoFileReader('nnDepth4.mp4',...
    'ImageColorSpace','Intensity','VideoOutputDataType','uint8');
%%
% Setting frames to 5 because it is a short video. Set initial standard
% deviation.
detector = vision.ForegroundDetector(...
    'NumTrainingFrames', 50, ...
    'InitialVariance', 20*20);
%%
% Perform blob analysis.
blob = vision.BlobAnalysis(...
    'CentroidOutputPort', false, 'AreaOutputPort', false, ...
    'BoundingBoxOutputPort', true, ...
    'MinimumBlobAreaSource', 'Property', 'MinimumBlobArea', 20000);
%%
% Insert a border.
shapeInserter = vision.ShapeInserter('BorderColor','Black');

```

	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

%%

% Play results. Draw bounding boxes around cars.

```
videoPlayer = vision.VideoPlayer();
```

```
while ~isDone(videoSource)
```

```
    frame = step(videoSource);
```

```
    fgMask = step(detector, frame);
```

```
    bbox = step(blob, fgMask);
```

```
    out = step(shapeInserter, frame, bbox);
```

```
    step(videoPlayer, out);
```

```
end
```

%%

% Release objects.

```
release(videoPlayer);
```

```
release(videoSource);
```

	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

Background Reduction

```

Depth=VideoReader('adjDepth3.mp4');
Color=VideoReader('adjRgb3.mp4');
fondoD=(imread('fonndoD.jpg'));
fondoR=(imread('fonndoR.jpg'));

depthObject = VideoWriter('nbDepth3', 'MPEG-4');
colorObject = VideoWriter('nbRgb3', 'MPEG-4');
depthObject.FrameRate=600/61;
colorObject.FrameRate=600/61;
open(colorObject);           % Create a new mp4 file
open(depthObject);

depthFrame = (readFrame(Depth));
rgbFrame = (readFrame(Color));
[X ,Y, Z]=size(depthFrame);
fondoR = uint8(imresize( fondoR, [X, Y]));
fondoD = uint8(imresize( fondoD, [X, Y]));

kernel=fspecial('average',80);
hsvFondoR= imfilter(rgb2hsv(fondoR), kernel);
% fondoGray = rgb2gray(fondoR);
% fondoGray = medfilt2(fondoGray);
% fondoGray = imfilter(fondoGray, kernel);

inicio=0
while hasFrame(Depth) || hasFrame(Color)

```

	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

```

inicio=inicio+1
kernel=fspecial('average',70);
depthFrame = rgb2gray (readFrame(Depth));
rgbFrame = (readFrame(Color));
hsvRgb = imfilter(rgb2hsv(rgbFrame), kernel);

for i=1:X
  for j=1:Y
    if (abs(hsvFondoR(i,j,1)-hsvRgb(i,j,1))<=0.1)&&...
      (abs(hsvFondoR(i,j,3)-hsvRgb(i,j,3))<=0.32)
      rgbFrame(i,j,1) = 255;
      rgbFrame(i,j,2) = 255;
      rgbFrame(i,j,3) = 255;
    end
    if (depthFrame(i,j)>70) || (depthFrame(i,j)<10)
      depthFrame(i,j) = 255;
    end
  end
end

writeVideo(depthObject,(depthFrame));
writeVideo(colorObject,(rgbFrame));

% Algunos de los metodos empleados para hacer background reduction con
% resultados no tan buenos.

%  rgbGray = imfilter(rgb2gray(rgbFrame), kernel);

```

 Institución Universitaria	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

```

%   rgbGray = medfilt2(rgb2gray(rgbFrame));
%   imshow(rgbGray);
%   [~, threshold] = edge(rgb2gray(rgbFrame), 'sobel');
%   fudgeFactor = .5;
%   BWs = edge(rgb2gray(rgbFrame), 'sobel', threshold * fudgeFactor);
%   BWs = (edge((rgbGray), 'Canny', 0.2)); %rgb2gray(rgbFrame)
%   imshow(BWs);
%   se90 = strel('line', 30, 90);
%   se0 = strel('line', 30, 0);
%   se90 = strel('disk', 50);
%
%   BWsdil = imdilate(BWs, [se90 se0], 'full');
%   BWsdil = imclose(BWs, se90);
%   BWdfill = imfill(BWsdil);
%   imshow(BWdfill);

end

fin=0

close(depthObject);
close(colorObject);

```

	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

Depth Denoising

```

Depth=VideoReader('nbDepth3.mp4');
% Color=VideoReader('nbRgb3.mp4');

depthObject = VideoWriter('nnDepth4', 'MPEG-4');
depthObject.FrameRate=600/61;
open(depthObject); % Create a new mp4 file
se = strel('disk',5);

inicio=0

while hasFrame(Depth) || hasFrame(Color)
    inicio = inicio+1
    depthFrame = (readFrame(Depth));
    depthFrame = imdilate(depthFrame, se);
    depthFrame = imgaussfilt(depthFrame,5);
    depthFrame = basavaFilt(rgb2gray(depthFrame));
    writeVideo(depthObject,(depthFrame));
end

fin=0
close(depthObject);

```

 Institución Universitaria	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

Blind Inpainting

close all

clear all

```
I= imread('spain_prepronew.tif','tif');
```

```
figure;imshow(I);
```

```
I0= double(I);
```

```
c=double(I);
```

```
R=0;
```

```
i=0;
```

```
while 1
```

```
    c=gconv2(I,c);
```

```
    Z2(i+1)= norm(c);
```

```
    if Z2(i+1)-R<10^-3
```

```
        break;
```

```
    end
```

```
    R=Z2(i+1);
```

```
    i=i+1;
```

```
end
```

```
I2=c;
```

```
I4=uint8(I2);
```

```
figure;imshow(I4);
```

	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

```

fnameout = ['spain_prepro1', '.tif'];
imwrite(I4,fnameout);
function a=gconv2(lmg,c)

[m,n]=size(lmg);

a(:,:)=double(lmg);
b(:,:)=conv2(c,fspecial('gaussian',[21 21] ,10),'same');
%b(:,:)=conv2(c,fspecial('gaussian',[11 11] ,10),'same');

for i=1:m
    for j=1:n
        if a(i,j)<=50
            a(i,j)=b(i,j);
        end
    end
end

%%c=a;

End

```

	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

SVideoResAdj

clear all;

close all;

clc;

vid1= videoinput('kinect',1);

vid2 = videoinput('kinect',2);

srcDepth=getselectedsource(vid2);

triggerconfig(vid1, 'manual');

vid1.FramesPerTrigger = inf;

triggerconfig(vid2, 'manual');

vid2.FramesPerTrigger = inf;

start(vid2);

start(vid1);

depthObject = VideoWriter('Depth3', 'MPEG-4');

colorObject = VideoWriter('Rgb3', 'MPEG-4');

depthObject.FrameRate=600/61;

colorObject.FrameRate=600/61;

open(colorObject); % Create a new AVI file

open(depthObject);

trigger(vid1)

trigger(vid2)

tic;

 Institución Universitaria	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

```
for iFrame = 1:600 % Capture 300 frames
```

```
    iFrame
```

```
    I = getdata(vid2,1);
```

```
    writeVideo(depthObject,mat2gray(I));
```

```
    F = getdata(vid1,1);
```

```
    writeVideo(colorObject,F); % Add the frame to the AVI filee
```

```
    flushdata(vid1);
```

```
    flushdata(vid2);
```

```
end
```

```
time=toc
```

```
stop(vid2);
```

```
stop(vid1);
```

```
close(colorObject);
```

```
close(depthObject);
```


 Institución Universitaria	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

ACTA NO. _____
FECHA ENTREGA: _____