



UNIVERSIDAD TÉCNICA DEL NORTE

FACULTAD DE INGENIERÍA EN CIENCIAS APLICADAS

ESCUELA DE INGENIERÍA EN MECATRÓNICA

TRABAJO DE GRADO PREVIO A LA OBTENCIÓN DEL TÍTULO
DE INGENIERO EN MECATRÓNICA

TEMA:

“SISTEMA PARA CONTEO DE PERSONAS BASADO
EN VISIÓN POR COMPUTADOR”

“SYSTEM FOR PEOPLE COUNTING BY COMPUTER
VISION”

AUTOR: JHONNY SANTIAGO CHAMORRO RODRIGUEZ

DIRECTOR: CARLOS XAVIER ROSERO

IBARRA-ECUADOR
SEPTIEMBRE 2017



UNIVERSIDAD TÉCNICA DEL NORTE
BIBLIOTECA UNIVERSITARIA
AUTORIZACIÓN DE USO Y PUBLICACIÓN A FAVOR DE LA
UNIVERSIDAD TÉCNICA DEL NORTE
IDENTIFICACIÓN DE LA OBRA

La Universidad Técnica del Norte dentro del proyecto *Repositorio Digital Institucional*, determinó la necesidad de disponer de textos completos en formato digital con la finalidad de apoyar los procesos de investigación, docencia y extensión de la Universidad.

Por medio del presente documento dejo sentada mi voluntad de participar en este proyecto, para lo cual pongo a disposición la siguiente información:

DATOS DEL AUTOR			
CÉDULA DE IDENTIDAD	1004427207		
APELLIDOS Y NOMBRES	CHAMORRO RODRIGUEZ JHONNY SANTIAGO		
DIRECCIÓN	1003727870		
EMAIL	jschamorro@utn.edu.ec - jhonnychamorro92@gmail.com		
TELÉFONO FIJO	062615731	TELÉFONO MÓVIL	0999657323
DATOS DE LA OBRA			
TÍTULO	"SYSTEM FOR PEOPLE COUNTING BY COMPUTER VISION"		
AUTOR	"IDENTIFICACIÓN DE UN CUADRICÓPTERO"		
FECHA	JHONNY SANTIAGO CHAMORRO RODRIGUEZ		
PROGRAMA	Septiembre DE 2017		
TÍTULO POR EL QUE OPTA	PREGRADO		
DIRECTOR	INGENIERO EN MECATRÓNICA		
	CARLOS XAVIER ROSERO C.		

AUTORIZACIÓN DE USO A FAVOR DE LA UNIVERSIDAD TÉCNICA DEL NORTE

Yo, Jhonny Santiago Chamorro Rodriguez con cédula de identidad Nro. 1004427207, en calidad de autor y titular de los derechos patrimoniales de la obra o trabajo de grado descrito anteriormente, hago entrega del ejemplar respectivo en formato digital y autorizo a la Universidad Técnica del Norte, la publicación de la obra en el Repositorio Digital Institucional y uso del archivo digital en la Biblioteca de la Universidad con fines académicos, para ampliar la disponibilidad del material y como apoyo a la educación, investigación y extensión; en concordancia con la Ley de Educación Superior, Artículo 144.

CONSTANCIA

El autor manifiesta que la obra objeto de la presente autorización es original y se la desarrolló sin violar derechos de autor de terceros, por lo tanto la obra es original, y que es el titular de los derechos patrimoniales, por lo que asume la responsabilidad sobre el contenido de la misma y saldrá en defensa de la Universidad en caso de reclamación por parte de terceros.

Ibarra, Septiembre de 2017



Jhonny Santiago Chamorro Rodriguez
C.I.: 1004427207



UNIVERSIDAD TÉCNICA DEL NORTE
FACULTAD DE INGENIERÍA EN CIENCIAS APLICADAS
CESIÓN DE DERECHOS DE AUTOR DEL TRABAJO DE GRADO A
FAVOR DE LA UNIVERSIDAD TÉCNICA DEL NORTE

Yo, Jhonny Santiago Chamorro Rodriguez con cédula de identidad Nro. 1004427207, manifiesto mi voluntad de ceder a la Universidad Técnica del Norte los derechos patrimoniales consagrados en la Ley de Propiedad Intelectual del Ecuador, artículos 4, 5 y 6, en calidad de autor (es) de la obra o trabajo de grado denominado “SYSTEM FOR PEOPLE COUNTING BY COMPUTER VISION”, que ha sido desarrollado para optar por el título de Ingeniero en Mecatrónica, en la Universidad Técnica del Norte, quedando la Universidad facultada para ejercer plenamente los derechos cedidos anteriormente. En mi condición de autor me reservo los derechos morales de la obra antes citada. En concordancia suscribo este documento en el momento que hago entrega del trabajo final en formato impreso y digital a la Biblioteca de la Universidad Técnica del Norte.

Ibarra, Septiembre de 2017

Jhonny Santiago Chamorro Rodriguez
C.I.: 1004427207



UNIVERSIDAD TÉCNICA DEL NORTE
FACULTAD DE INGENIERÍA EN CIENCIAS APLICADAS
CERTIFICACIÓN

En calidad de director del trabajo de grado “SYSTEM FOR PEOPLE COUNTING BY COMPUTER VISION”, presentado por el egresado JHONNY SANTIAGO CHAMORRO RODRIGUEZ, para optar por el título de Ingeniero en Mecatrónica, certifico que el mencionado proyecto fue realizado bajo mi dirección.

Ibarra, Septiembre de 2017



Carlos Xavier Rosero
DIRECTOR DE TESIS



UNIVERSIDAD TÉCNICA DEL NORTE
FACULTAD DE INGENIERÍA EN CIENCIAS APLICADAS
DECLARACIÓN

Yo, Jhonny Santiago Chamorro Rodriguez con cédula de identidad Nro. 1004427207, declaro bajo juramento que el trabajo aquí escrito es de mi autoría; que no ha sido previamente presentado para ningún grado o calificación profesional; y, que he consultado las referencias bibliográficas que se incluyen en este documento.

A través de la presente declaración cedo mis derechos de propiedad intelectual correspondientes a este trabajo, a la Universidad Técnica del Norte - Ibarra, según lo establecido por la Ley de Propiedad Intelectual, por su Reglamento y por la normativa institucional vigente.

Ibarra, Septiembre de 2017

Jhonny Santiago Chamorro Rodriguez
C.I.: 1004427207

Agradecimiento

Agradezco a mis padres César Chamorro y Consuelo Rodríguez por apoyarme moral y económicamente para poder llegar a ser un profesional, por darme los consejos más adecuados como todos los padres dan a sus hijos.

A mi hermana Karol Chamorro Rodríguez por todo el apoyo brindado día a día en el transcurso de cada año de mi carrera Universitaria.

A Xavier Rosero por además de ser mi tutor es un gran amigo en el que se puede confiar, por permitirme trabajar junto a su dirección, por confiar en mí y por brindarme consejos sabios.

Jhonny Chamorro

Dedicatoria

Esta tesis está dedicada a mis padres César Augusto Chamorro Ortega y Consuelo del Rosario Rodríguez Mejía por el apoyo incondicional que tuve de su parte, depositando su entera confianza en cada reto que se me presentaba sin dudar ni un solo momento en mi inteligencia y capacidad. Todo lo que soy hoy es gracias a ellos. Su tenacidad y lucha insaciable han hecho de ellos el gran ejemplo a seguir y destacar, no solo para mí, sino para mi hermana también.

Le doy gracias a mi hermana Karol Gisell Chamorro Rodríguez por estar siempre para escucharme y ayudarme a no darme por vencido, es uno de los pilares importantes de mi vida, además de ser mi hermana ha sido una amiga que está ahí por si quiero rendirme.

Jhonny Chamorro

Resumen

La detección y el conteo de personas es un método ampliamente utilizado para el monitoreo y la videovigilancia. El objetivo de este proyecto es realizar el conteo de personas a partir de imágenes planas, y con una cámara ubicada en una posición cenital en un entorno interior.

El algoritmo usado para el conteo de personas se desarrolló mediante procesos morfológicos basado en imágenes planas, ya que la cámara se encuentra en posición cenital y en un entorno interior, las mismas que sirven para modificar el comportamiento del sistema. El sistema de conteo se divide en dos, una parte de monitoreo de imagen y una de adquisición de imagen.

El sistema usado para conteo de personas contiene condiciones; En primer lugar hay que comprobar si la persona está en la zona de detección, en segundo lugar, contiene límites en la parte superior e inferior para comprobar si la persona entra o sale.

El sistema de conteo de personas realizado en Python a través de la biblioteca OpenCv muestra los datos procedentes de la cámara, indicando la cantidad de personas que entraron o salieron de un sitio en particular. El algoritmo que se desarrolló se evaluó en una base de datos registrada y etiquetada.

Abstract

The detection and counting of people is a method widely used for monitoring and video surveillance. The objective of this project is to perform the counting of people process from flat images, with a camera located on top of an interior environment.

The algorithm used for the counting of people process was developed by morphological processes based on flat images, since the camera is on top of an interior environment, the same ones that serve to modify the system behaviour. The counting system is divided into two components; one component is for image monitoring and the other one for image acquisition.

The system used for counting people has some conditions that need to be met in order for it to work properly; First of all, it is important to check if the person is in the detection zone; secondly, the system has limits at the top and bottom areas to check if the person enters or leaves the detection zone.

The people counting system was coded in Python through the OpenCv library which shows the data coming from the camera, indicating the number of people who entered or left a particular site. The algorithm that was developed for this counting system was evaluated in a classified and registered database.

Índice general

1. Introducción	1
1.1. Objetivos	1
1.1.1. Objetivo Principal	1
1.1.2. Objetivos Específicos	1
1.2. Antecedentes	2
1.3. Problema	2
1.4. Justificación	3
1.5. Alcance	3
2. Revisión literaria	4
2.1. Visión por Computador	4
2.2. Python-OpenCV	4
2.3. Cámara	6
2.4. Representación de imágenes	7
2.4.1. Definición	7
2.4.2. Tipos de imágenes	8
2.4.3. Características de una imagen en niveles de gris	8
2.5. Preprocesamiento	10
2.5.1. Substracción de fondo	10
2.6. Procesamiento morfológico de imágenes	12
2.6.1. Erosión binaria	12
2.6.2. Dilatación binaria	12
2.6.3. Procesamiento morfológico en niveles de grises	13
2.6.4. Apertura y cierre	13
2.6.5. Gradiente morfológico	15
2.7. Algoritmo de Canny	16
2.7.1. Obtención del gradiente	16
2.7.2. Supresión no máxima al resultado del gradiente	17
2.8. Propiedades de contorno	18

3. Metodología	19
3.1. Información preliminar	19
3.2. Posicionamiento de la cámara	20
3.3. Contador de personas	21
3.3.1. Ingresos y salidas de personas	21
3.4. Sistema de validación	21
3.5. Segmentación	22
3.5.1. Extracción de fondo	22
3.6. Filtro morfológico	24
3.6.1. Erosión	24
3.6.2. Dilatación	24
3.6.3. Combinación de ambas	26
3.7. Detección de contorno	26
3.8. Detección de personas	27
3.9. Video de salida	28
4. Pruebas y resultados	29
5. Conclusiones y Trabajo futuro	33
A. Listado de códigos	35
A.1. Programa implementado	35

Índice de figuras

2.1. Cámara Web HD Logitech c270, Video 720p, Resolución de 1280x720, USB	7
2.2. Representación imagen a color	9
2.3. Representación imagen en grises	10
2.4. Representación imagen a color	11
2.5. Igualación de histograma	11
2.6. Representación imagen a color	14
2.7. a) Imagen segmentada, b) erosión con elemento estructurante de disco de radio 2, c) apertura con elemento estructurante de disco de radio 2	14
2.8. Apertura y cierre en niveles de grises	15
2.9. Gradientes	15
2.10. Máscara Gaussiana	17
2.11. Máscara Gaussiana	17
3.1. Diferentes posicionamientos de la cámara: a) enfoque frontal, b) enfoque diagonal, c) enfoque cenital	20
3.2. Conteo de personas	21
3.3. Área de detección	22
3.4. Diagrama de bloques de la aplicación	22
3.5. Respuesta del algoritmo MOG ante primer plano dinámico	23
3.6. Respuesta del algoritmo MOG ante primer plano estático	23
3.7. Ejemplo aplicando erosión	25
3.8. Ejemplo aplicando dilatación	25
3.9. Contorno de una persona	26
3.10. Detección de una persona	27
3.11. Detección de una persona	28
4.1. Imágenes de video de resolución 480x480 píxeles	29
4.2. Imágenes de video de resolución 320x240 píxeles(VGA)	30
4.3. Imágenes de video de resolución 320x240 píxeles(VGA)	31
4.4. Imágenes de video de resolución 320x240 píxeles(VGA)	32

Índice de cuadros

4.1. Resultados primer video	30
4.2. Resultados segundo video	30
4.3. Resultados tercer video	31
4.4. Resultados tercer video	32

Listings

A.1. Conteo de Personas	35
-----------------------------------	----

Capítulo 1

Introducción

Actualmente las aplicaciones que se basan en el conteo de personas son muy usadas por lo que existe una gran demanda, debido a su sistema de seguridad y videovigilancia que permite conocer la cantidad de personas que entran y salen de un lugar [4].

Algunos métodos para la detección de personas están basados en la substracción de fondo, en el seguimiento de los objetos en primer plano se realiza a través de el área de ocupación para diferenciar el número de personas y un vector de color para realizar el seguimiento de cada persona. Otro tipo de sistema es donde la cámara se encuentra inclinada, estos sistemas necesitan la obtención de una perspectiva de la persona para llevar acabo su detección [1].

La mayoría de sistemas que realizan el conteo de personas lo hacen con una sola cámara que se orienta de forma cenital y a una cierta altura. El sistema posee dos regiones de interés definidas en la parte superior e inferior de la imagen. Se calcula el histograma de dichas regiones de la imagen de movimiento y mediante la comparación de un umbral detectan el nivel de ocupación. El conteo bi-direccional se basa en una algoritmia determinista que es basada en el cruce de las regiones superior e inferior. Este tipo de sistema con orientación cenital no puede realizar un análisis del objeto detectado, ya que la vista no ofrece suficientes características para detectar a la persona.

1.1. Objetivos

1.1.1. Objetivo Principal

- Implementar un sistema de conteo de personas usando visión artificial, bajo una plataforma de software libre.

1.1.2. Objetivos Específicos

- Abordar el estado del arte sobre conteo de personas con visión por computador.

- Seleccionar una cámara con las características idóneas para su funcionamiento con software libre.
- Desarrollar el sistema de conteo de personas.
- Realizar pruebas de funcionamiento.

1.2. Antecedentes

En la actualidad existe una gran demanda de sistemas de conteo de personas que ofrezcan una alta efectividad y precios muy bajos. Estos sistemas de conteo de personas son muy utilizados para tareas como videovigilancia, seguridad, análisis estadístico del acceso de personas a un recinto, etc.

Se han publicado muchas investigaciones para resolver este problema, Que es contar a las personas usando cámara de video. Hay algunas situaciones difíciles de resolver incluso con las velocidades actuales de la computadora ya que el algoritmo opera en tiempo real y por lo tanto limita la complejidad de los métodos de detección y rastreo.

Basados en ese punto, el conteo de personas presenta algunos problemas referentes a la iluminación, tiempo de respuesta, entre otros mas concretos. Cada vez se han ido mejorando los métodos para la detección de persona usando el algoritmo adecuado para cada tipo de aplicación en diferentes entornos [7].

1.3. Problema

En la actualidad empresas, instituciones y hogares han optado por la utilización e implementación de videocámaras principalmente como medios de videovigilancia y monitoreo, por lo que su uso se ha vuelto de gran importancia en la seguridad [1].

Con la implementación de técnicas de visión por computador aplicadas al monitoreo se ha logrado potencializar el uso de las videocámaras permitiendo generar aplicaciones tales como, el tráfico vehicular, realizar el conteo de personas, objetos o efectuar reconocimiento facial, datos que son usados ampliamente por otras ramas de la ingeniería [2],[?].

El conteo de personas a través de videocámaras, es uno de los más usados en la actualidad, debido a que permite conocer la cantidad de personas que entran y salen de un lugar. En lo que concierne a la realidad de la Universidad Técnica del Norte, particularmente a la facultad de ingeniería en ciencias aplicadas en su laboratorio de automática y robótica no existe un sistema de conteo de personas, Se pretende realizar una aplicación que permita medir la afluencia de

personas al ingreso al laboratorio de automática y robótica, con el fin de potencializar el uso de las instalaciones del mismo.

1.4. Justificación

Hoy en día los sistemas de monitoreo y video vigilancia son usados ampliamente debido a su fácil instalación y bajo costo [5]. Con ello y con la aplicación de técnicas de visión por computador se intenta potencializar estos equipos para poder utilizarlos en múltiples aplicaciones orientadas a servicios [4].

La relevancia del sistema de conteo de personas propuesto permitirá realizar un análisis de la concurrencia de personas, datos que son útiles especialmente en el manejo de servicios [6], por ejemplo, para vigilar lugares públicos y obtener el óptimo desempeño de las instalaciones.

1.5. Alcance

El software que se va a desarrollar tiene como objetivo detectar características de personas, a través de algoritmos computacionales de visión artificial ya existentes. Lo expuesto anteriormente y cierta interfaz se desarrollarán bajo la plataforma de software libre.

Capítulo 2

Revisión literaria

2.1. Visión por Computador

La visión por ordenador ayuda a ver el mundo que nos rodea, por lo tanto es la ciencia que permite obtener mejores resultados que una máquina o computadora. La visión computacional surgió con la idea de conectar una cámara de video a un computador, para que las imágenes capturadas fueran manipuladas y así obtener una representación simbólica de los objetos presentes en las escenas analizadas [8].

La visión por computadora principalmente realiza una extracción automática, es decir el análisis de la información necesaria de una sola imagen o de una secuencia de imágenes(video). Para lograr este proceso se necesita una base teórica y algorítmica para una mejor comprensión visual automática.

2.2. Python-OpenCV

Los primeros trabajos para la detección y conteo de personas se basan en el uso de cámaras de color. Como [5] hace uso de la información 3D de la escena mediante cámaras estéreo y cámaras RGB-D, éstas mostraron buenos resultados precisos acerca de la detección y seguimiento de personas. Otros trabajos se basan en la detección de caras [2], o por la clasificación de puntos de interés(RoI) [3]. Las propuestas presentadas tienen buenos resultados en condiciones controladas, pero si generan conflicto en escenarios con un alto grado de oclusión, para tratar de minimizar las oclusiones que se presentan existen otras alternativas al ubicar la cámara de forma que se tenga una vista cenital de la escena.

En la actualidad empresas, instituciones y hogares han optado por la utilización e implementación de videocámaras principalmente como medios de videovigilancia y monitoreo, por lo que su uso se ha vuelto de gran importancia en la seguridad [1]. Dicho proceso consiste en detectar personas y determinar su ubicación, existen algunos factores que pueden afectar el

proceso de detección, entre ellos: iluminación, posición, escala, ubicación, orientación.

Para los sistemas de visión donde no se obtienen datos significativos para poder identificar al individuo [6], los autores de este trabajo ubican cámaras de baja resolución alejadas de la persona con el fin de no invadir la privacidad del mismo. Por otro lado, hay múltiples trabajos que proponen el uso de sensores de profundidad basados en tiempo de vuelo TOF para la detección y conteo de personas [7],[4]. En varios trabajos emplean sensores de profundidad en donde la cámara va ubicada de forma cenital con el objetivo de reducir el efecto de las oclusiones creando así una imagen mas clara para el seguimiento, detección y conteo de personas [15].

A continuación se muestran algunos métodos de segmentación entre regiones estáticas y zonas con movimiento basado en las técnicas de extracción del fondo. Para que la clasificación sea lo más clara posible se ha dividido las técnicas posibles en dos categorías, modelo de fondo y varios modelos de fondo.

El primer grupo depende de las aproximaciones que hagan con la máscara de primer plano obtenidas:

- **Análisis imagen a imagen:** Esta categoría describe los métodos que emplean modelos de segmentación frente-fondo bastante comunes, seguidos de otro tipo de análisis. Dependiendo del tipo análisis pueden aparecer varias categorías:
 - **Aproximaciones clásicas:** Basadas en el uso de técnicas sencillas de segmentación frente-fondo y un postprocesado de la máscara de foreground seguido a su vez de alguna otra etapa de análisis.
 - **Basados en la acumulación de máscaras de primer plano:** Dicha acumulación se realiza frame a frame y con ella se puede moldear una máscara final de primer plano, de donde se obtienen las regiones estáticas.
 - **Basados en algunas propiedades del modelo de fondo utilizado:** Considerando las transiciones entre los diferentes estados de un modelo de mezcla de Gaussianas o observando el valor de algunos parámetros como, por ejemplo, el peso de las Gaussianas.
- **Análisis de máscaras de primer plano muestreadas:** Las aproximaciones intentan detectar regiones estáticas analizando la secuencia de video a diferentes velocidades, aprovechándose de las ventajas espacio-temporales que ello conlleva.

El segundo grupo se basa en las aproximaciones que combinan mas de un modelo de fondo para cada pixel.

- **Aproximaciones basadas en el análisis imagen a imagen:** Se combina las propiedades de los diferentes modelos de fondo que se utilizan.
- **Aproximaciones basadas en el sub-muestreo:** Estas aproximaciones detectan regiones estacionarias analizando la secuencia de vídeo a través de los diferentes modelos de fondo debido a que cada modelo de fondo se muestrea con una tasa binaria diferente.

Las técnicas que se mencionaron anteriormente son capaces de detectar objetos en movimiento por lo tanto tienen que resolver los siguientes problemas:

- **Ruido:** Es importante que el algoritmo sea capaz de eliminar el ruido procedente de la cámara de vídeo pues puede provocar la detección de zonas de foreground incorrectas. Este ruido es debido al sensor de la cámara o al medio de transmisión de la señal, que se manifiesta en píxeles aislados que toman un valor diferente al de sus vecinos.
- **Sombras y reflejos:** Es uno de los principales problemas ya que no pertenecen ni al background ni al foreground y en la mayoría de los casos generan interferencias que hacen que el algoritmo no actúe de forma adecuada, por ello hay que eliminarlos. Esto se realiza mediante técnicas de filtrado y operaciones morfológicas.
- **Cambios de iluminación:** Son variaciones de la iluminación, estas variaciones pueden tener lugar tanto si la escena ha sido capturada en el exterior (cambio de luminosidad en todo el día) como si ha sido tomada en el interior. La iluminación juega un papel importante en la visión artificial pues simplifica de manera considerable el análisis y el procesamiento de la escena captada. Es un factor que influye en gran medida en la complejidad del algoritmo, y como es más sencillo modificar la iluminación que un complejo algoritmo, hay que dedicar el tiempo necesario para conseguir una buena iluminación de forma que el algoritmo no se complique más de lo debidamente necesario.
- **Actualización del fondo de escena:** Su importancia radica principalmente en dos puntos. El primero de ellos es que la inicialización del fondo de la secuencia, generalmente no coincide con el fondo ya que puede haber algún objeto o persona en movimiento, por lo que el algoritmo debe ser capaz de no clasificarlo como fondo, el segundo punto por lo que la actualización de fondo es importante es debido a que pueden haber cambios importantes en la secuencia.
- **Determinar los parámetros de funcionamiento de los algoritmos:** Es una de las tareas más complejas ya que dependiendo de los parámetros utilizados los algoritmos mostrarán unos resultados u otros, que en la mayoría de los casos difieren en gran medida.

2.3. Cámara

Para un funcionamiento adecuado de visión por ordenador hay que tomar muy en cuenta la calidad de la imagen sobre la que se va a trabajar, en algunos casos viene a ser la parte fundamental para el éxito de la aplicación más que el propio algoritmo.

La cámara es capaz de producir un vídeo natural y gracias a su enfoque de gama alta las imágenes tomadas son siempre nítidas.

Especificaciones

- Videoconferencias HD (1280 x 720 píxeles) con el sistema recomendado.

- **Captura de video:** Hasta 1280 x 720 píxeles.
- Tecnología Logitech Fluid Crystal™.
- **Fotos:** Hasta 3.0 megapíxeles (mejora por software).
- **Micrófono:** Micrófono integrado con reducción de ruido.
- **Conexión USB:** Certificación USB 2.0 de alta velocidad (es recomendado).
- **Montaje:** Clip universal para monitores LCD, CRT o portátiles.

A continuación se muestra la imagen de la cámara que se usó para la obtención de imágenes en tiempo real.



Figura 2.1: Cámara Web HD Logitech c270, Video 720p, Resolución de 1280x720, USB

2.4. Representación de imágenes

La información que contienen las imágenes son representadas de diversas maneras. Se podrá observar la representación espacial, excluyendo algunas representaciones también muy útiles como la representación en número de ondas, que se obtiene al aplicar la Transformada de Fourier a una representación espacial. También es necesario tener conocimiento sobre como puede ser manejada la representación por computadora.

2.4.1. Definición

Una imagen representa una distribución espacial de la irradiación en un plano por lo tanto la distribución de irradiación espacial puede ser descrito como una función continua de dos variables espaciales.

De manera general, la imagen se define como:

$$f : \Omega \rightarrow C \quad (2.1)$$

donde se tiene que en $\Omega \subset R^2$ y C es llamado el espacio característico.

El espacio característico puede ser:

- Un intervalo, por ejemplo $[0, 255]$ o $[0, \infty]$, para imágenes en escala de grises.
- Un subconjunto de R^3 , por ejemplo $[0, 1]^3$, para imágenes a color RGB.
- Para imágenes en escala de grises, la función f puede ser vista como una superficie tridimensional.

Las computadoras no pueden manejar imágenes continuas, pero si números o arreglos de ellos, por lo tanto se necesita representar las imágenes como arreglos bidimensionales de puntos [9].

2.4.2. Tipos de imágenes

Imágenes a color

Son imágenes compuestas por cuatro valores almacenados en tres planos de píxeles o matrices, véase Figura 2.4, que se encuentran definidos por los modelos de color, entre los que se encuentran: el modelo RGB (red, green, blue), el modelo de color de Hering, el modelo HSL (hue, saturation y luminance), el modelo HSV (hue, saturation, value) y el diagrama de cromaticidad establecido por la CIE (Comisión Internationale de l'Éclairage).

En el modelo RGB cada color aparece en sus componentes espectrales primarias rojo, verde, azul. La razón por la que se toman estos tres colores tiene que ver con la teoría tricromática de color que indica que tres estímulos cromáticos diferentes, combinados en diferentes proporciones, pueden producir el efecto equivalente a muchos más estímulos que una persona percibirá como muchos colores diferentes [14].

Imágenes en niveles de gris

Son imágenes compuestas de un simple plano o matriz de píxeles, en donde sus valores están representados entre $[0, 255]$ para una imagen de 8-bit, en donde el 0 representa el color negro y el 255 el color blanco, dado que en ese intervalo se ubican todas las intensidades o grises disponibles en la imagen, véase Figura 2.3.

2.4.3. Características de una imagen en niveles de gris

- Tamaño de las imágenes: Las dimensiones de una imagen se representan por medio de los pares ordenados de píxeles (M, N) , donde M representa el número de filas de y N el número de columnas. En la mayoría de los casos se opta por tener M igual a N dado que la igualdad entre filas y columnas simplifica el cálculo y presenta grandes ventajas computacionales, cuando las imágenes a procesar tienen tamaños en potencias de dos

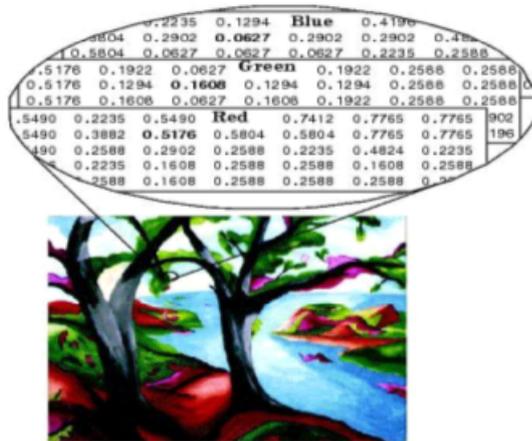


Figura 2.2: Representación imagen a color

(128x128, 256x256, 512x512). De tal forma que al momento de almacenar estas imágenes el número de bits necesarios para ser almacenadas se calcula con la siguiente expresión:

$$b = N \times N \times m \quad (2.2)$$

Donde m representa los distintos niveles de gris de la imagen, por ejemplo para una imagen de 512x512 con 256 niveles de gris ($m = 8$) son necesarios 2,097,152 bits o 262,144 bytes; es por esto que a mayor número de bytes empleados para el almacenamiento de una imagen, mayor será su resolución.

- **Resolución de la imagen:** La resolución de una imagen expresa cuanto detalle el observador puede ver en ella. Esta característica depende directamente de los valores de N y de M , dado que tener un mayor número de píxeles y niveles de gris en una imagen, da como resultado una mejor visualización de sus detalles; pero a su vez incrementa el tamaño de almacenamiento. Véase Figura 2.4
- **Histograma:** El histograma de una imagen es una función discreta que representa el número de píxeles en la imagen ubicados en el eje y , en función de los niveles de intensidad o gris presentes en la imagen ubicados en el eje x . En donde el tipo de distribución encontrada en el histograma representa el grado de contraste presente en la imagen; si los datos presentan una distribución uniforme en todo el rango del niveles de gris, se obtendrá un alto contraste que permitirá observar mejor los objetos presentes en la imagen, véase Figura 2.5

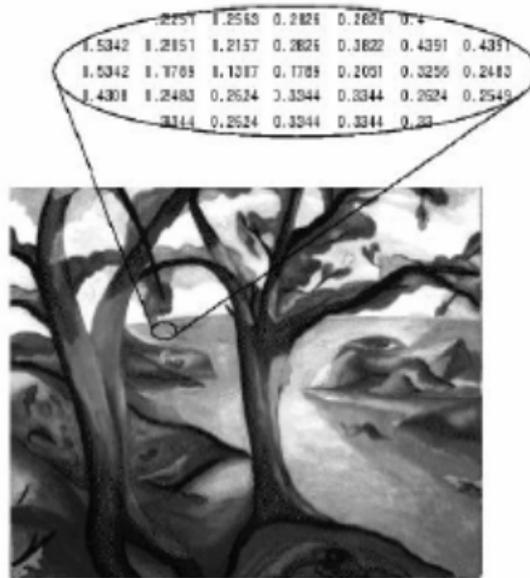


Figura 2.3: Representación imagen en grises

2.5. Preprocesamiento

En el área de procesamiento de imágenes y visión por computador y han sido estudiados para diferentes aplicaciones, como la vigilancia visual de la actividad del hombre en distintos lugares. Sujetos a rastreo y la mayoría de las actividades humanas al aire libre. En otra clase de estudio, el humano está sujeto a seguimiento en un entorno en el que existe una interfaz basada en la visión que tiene el usuario para interactuar con entornos virtuales.

2.5.1. Substracción de fondo

El método ViBe (Barnich and Van Droogenbroeck, 2009, 2011) utiliza un modelo del fondo que almacena para cada pixel, N muestras seleccionadas aleatoriamente de frames ya procesados. Si consideramos $m = m1, m2, \dots, mN$, las muestras correspondientes al pixel actual pt , cada mi para $i = 1..N$, consiste en un vector $mi = (r, g)$ que usa dos componentes del espacio RGB. Esto puede adaptarse fácilmente para extenderlo al espacio RGB, escala de grises u otro. El hecho de excluir el componente azul se debe a que en las secuencias de videovigilancia este componente suele encontrarse más saturado que los otros componentes. Luego, se define la condición de intersección $I(pt, mi)$ igual a 1, cuando la distancia euclídea evaluada entre el valor del pixel actual y una muestra del modelo menor o igual a un determinado valor de radio R , es decir:

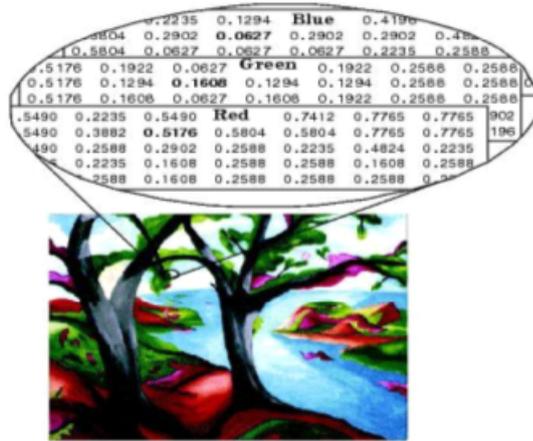
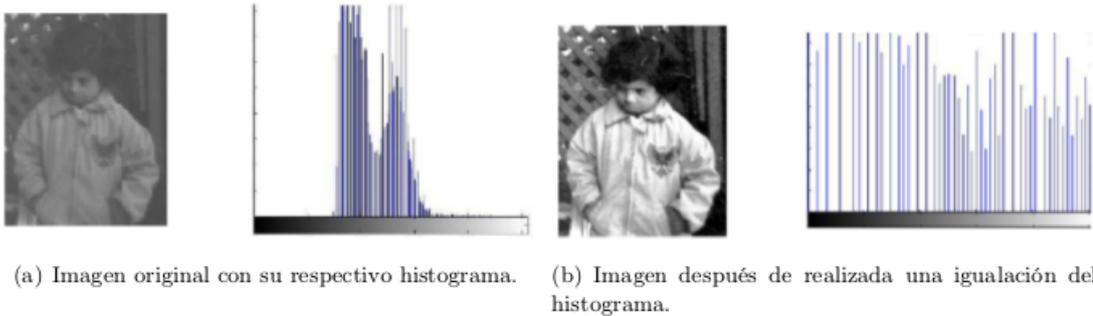


Figura 2.4: Representación imagen a color



(a) Imagen original con su respectivo histograma. (b) Imagen después de realizada una igualación del histograma.

Figura 2.5: Igualación de histograma

$$I(pt, mi) = \begin{cases} 1 & \text{si } \sqrt{(pt(r) - mi(r))^2 + (pt(g) - mi(g))^2} \leq R \\ 0 & \text{en otro caso} \end{cases} \quad (2.3)$$

radio R, es decir:

$$I(pt, mi) = (1 \text{ si } p((pt(r) - mi(r))^2 + (pt(g) - mi(g))^2) \leq R) \quad (2.4)$$

En otro caso la clasificación consiste en comparar cada pixel de la imagen actual con las n muestras del mismo que se encuentran almacenadas en m , y mediante un indicador determinar si es fondo o primer plano.

En la siguiente función se tiene una máscara binaria de clasificación para cada pixel del frame de entrada, Min es la cantidad mínima de veces que debe ser verdadera la 2.3.

$$Clasificar(pt, m) = \begin{cases} Fondo & \text{si } \sum_{i=1}^N I(pt, mi) \geq Min \\ Primerplano & \text{en otro caso} \end{cases} \quad (2.5)$$

Por cada imagen del video se genera una salida binaria correspondiente a la clasificación de cada pixel. Luego de realizar la detección, el modelo se actualiza en forma estocástica para adaptar gradualmente la representación a los diferentes cambios que ocurren a lo largo de una secuencia. Es deseable que sólo aquellos píxeles clasificados como fondo se inserten en el modelo, ya que la inserción de píxeles que pertenecen a objetos en movimiento, que resultan inciertos en la clasificación pueden alterar significativamente los resultados de la detección. Lo mencionado anteriormente es muy importante ya que ha permitido la introducción de mejoras al método original logrando mayor tasa de detección. Entre ellas podemos citar, el módulo que nos permite insertar píxeles que alternan permanentemente su clasificación entre fondo y primer plano (Blink Discard), el módulo de postprocesamiento (Extract Contour + Fill Object) que corrige la clasificación de aquellos píxeles clasificados según 2.5 como fondo, pero que en realidad pertenecen al interior de objetos en movimiento [10].

2.6. Procesamiento morfológico de imágenes

El objetivo de las transformaciones morfológicas es la extracción de estructuras geométricas en los conjuntos sobre los que se opera, mediante la utilización de otro conjunto de forma conocida, al que se le denomina elemento estructurante. El tamaño y forma del elemento estructurante se elige de acuerdo con la morfología sobre la que va a operar y la obtención de formas que se desea extraer.

2.6.1. Erosión binaria

La transformación de la erosión es el resultado de comprobar si el elemento estructurante B está completamente incluido dentro del conjunto X . Cuando esto no ocurre, el resultado de la erosión es el conjunto vacío:

$$E_B(X) = X \ominus B = \{B_X \subseteq X\} \quad (2.6)$$

Cuando los objetos de la escena sean menores que el elemento estructurante, éstos desaparecerán. Otra interpretación de la erosión supone tomar el valor mínimo de la imagen en el entorno de vecindad definido por el elemento estructurante.

2.6.2. Dilatación binaria

La dilatación es la transformación dual a la erosión. El resultado de la dilatación es el conjunto de elementos, tal que al menos algún elemento del conjunto estructurante B está contenido

en el conjunto X , cuando B se desplaza sobre el conjunto X :

$$\delta_B(X) = X \oplus B = \{X \cap B_X \neq \emptyset\} \quad (2.7)$$

Esta operación representa un crecimiento progresivo del conjunto X . Al pasar el elemento estructurante dentro del conjunto, éste no se modificará. Sin embargo, en la frontera del conjunto X , al desplazar a B , el conjunto resultado se expansionará. La aplicación iterada de este operador haría degradar la imagen, haciendo coincidir el conjunto dilatado con la imagen.

$$X \subseteq \delta_B(X) \quad (2.8)$$

La dilatación también se interpreta como el valor máximo del entorno de vecindad definido por el elemento estructurante.

2.6.3. Procesamiento morfológico en niveles de grises

Las operaciones de erosión y dilatación son crecientes, respectan el orden presente en la estructura del conjunto. Para dos imágenes f y g :

$$Si \quad f \leq g \quad \Rightarrow \quad E(f) \leq E(g) \quad (2.9)$$

$$Si \quad f \leq g \quad \Rightarrow \quad \delta(f) \leq \delta(g) \quad (2.10)$$

Gracias a esta propiedad, las operaciones morfológicas pueden extenderse del caso binario a la escala de grises, puesto que estas últimas señales pueden expresarse como una suma ponderada de imágenes binarias. La erosión de una función en un punto será el valor mínimo de la función dentro de la ventana definida por el elemento estructurante B , cuando el origen de B esté situado en dicho punto. La dilatación de una función será la operación dual. Retornará el valor máximo de la función definido por el elemento estructurante: [11]

$$E_s(f) = f \ominus B = \min(f(x-y), y \in B) \quad (2.11)$$

$$\delta_s(f) = f \oplus B = \max(f(x-y), y \in B) \quad (2.12)$$

2.6.4. Apertura y cierre

Usando los operadores elementales de erosión y dilatación se pueden diseñar operaciones de realce de las formas de los objetos. La erosión binaria suele utilizarse para eliminar pequeños objetos, práctica que suele ser utilizada en la fase de postprocesado, después de la segmentación. Sin embargo, tiene el inconveniente de disminuir el tamaño del resto de los objetos. Este efecto

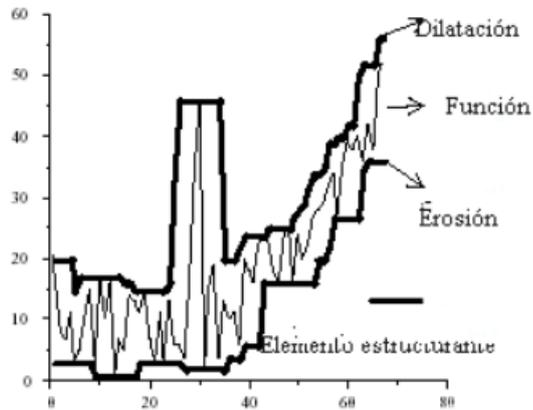


Figura 2.6: Representación imagen a color

puede ser utilizado con una aplicación en cascada de erosión y dilatación binaria con igual elemento estructurante. A esta operación se la llama apertura:

$$\Upsilon_s(X) = X \circ B = \delta_s(E_s(X)) \quad (2.13)$$

La apertura binaria elimina todos los objetos que no están completamente contenidos en el elemento estructurante, pero además no disminuye el tamaño a los objetos que superen la erosión. Sin embargo, la imagen resultante no recupera la misma forma de los objetos filtrados de la imagen de entrada. Esta operación puede ser ideal para la eliminación de ruido, aunque no preserva la geometría de los objetos ya que los bordes serán suavizados.



Figura 2.7: a) Imagen segmentada, b) erosión con elemento estructurante de disco de radio 2, c) apertura con elemento estructurante de disco de radio 2

La dilatación binaria realiza un crecimiento de los objetos, cerrando los agujeros y las grietas. El ensanchamiento de los objetos puede ser reducido mediante la aplicación seguida de una erosión. La operación combinada de dilatación y erosión es llamada cierre:

$$\varphi_s(X) = X \bullet B = E_s(\delta_s(\cdot)X) \quad (2.14)$$

El cierre binario morfológico produce que la dilatación rellene las estructuras que la erosión no puede excluir. Los contornos de los objetos también serán suavizados, pero habiendo rellenado las fisuras. En imágenes en niveles de grises, la apertura eliminará formas claras menor que el elemento estructurante; mientras el cierre, los objetos oscuros, serán atenuados.



Figura 2.8: Apertura y cierre en niveles de grises

2.6.5. Gradiente morfológico

El residuo de dos operaciones Ψ y ζ es su diferencia. Para el caso de conjunto, esta diferencia se define como:

$$\rho_{\Psi\zeta}(X) = \frac{\Psi(X)}{\zeta(X)} \quad (2.15)$$

El primer residuo que se puede definir en Morfología es el gradiente. Debida a las relaciones de extensión y antiextensión definidas sobre los operadores morfológicos se establece el concepto de gradiente morfológico. El gradiente morfológico o de Beucher es la diferencia entre la imagen original y su erosión o entre una dilatación y la imagen original o entre una dilatación y una erosión:

$$\begin{aligned} g^-(X) &= X \setminus (X \ominus B) && \text{gradiente por erosión} \\ g^+(X) &= (X \oplus B) \setminus X && \text{gradiente por dilatación} \\ g(X) &= (X \oplus B) \setminus (X \ominus B) && \text{gradiente simétrico} \end{aligned}$$

Figura 2.9: Gradientes

Su utilidad está en la localización de los bordes o perímetro de los objetos. La elección del tipo de gradiente morfológico dependerá de la geometría de los objetos.

2.7. Algoritmo de Canny

El objetivo de Canny era descubrir el algoritmo óptimo de detección de bordes. En su trabajo, siguió una lista de criterios para mejorar los métodos actuales de detección de bordes. La primera es la baja tasa de error. El segundo criterio es que los puntos de borde estén bien localizados, esto quiere decir que la distancia entre los píxeles de borde debe ser mínima, tal como se encuentran por el detector y el borde real. Un tercer criterio es tener una sola respuesta a un solo borde. Esto fue implementado porque el primer y segundo método no eran sustentables al eliminar completamente la posibilidad de respuestas múltiples a un borde.

El detector de bordes Canny primero suaviza la imagen para eliminar el ruido. Luego encuentra el gradiente de la imagen para resaltar las regiones con altas derivadas espaciales. El algoritmo sigue rastros a lo largo de las regiones y suprime cualquier pixel que no esté al máximo. La matriz de gradiente es reducida ahora mas por histéresis, la histéresis se utiliza para rastrear a lo largo de los píxeles restantes que no han sido eliminados, además utiliza dos umbrales y si la magnitud está por debajo del primer umbral se pone a cero. Si la magnitud está por encima del umbral alto, se hace una arista. Y si la magnitud está entre los 2 umbrales, entonces se pone en cero a menos que haya una trayectoria de este pixel a un pixel con un gradiente sobre T_2 . Para que un detector de bordes se considere en estado óptimo debe cumplir los siguientes puntos:

Buena detección: el algoritmo debe marcar el mayor número real en los bordes de la imagen como sea posible. **Buena localización:** los bordes de marca deben estar lo más cerca posible del borde de la imagen real. **Respuesta mínima:** el borde de una imagen sólo debe ser marcado una vez, y siempre que sea posible, el ruido de la imagen no debe crear falsos bordes.

Canny es un algoritmo multi-paso que puede detectar bordes con ruido sumprimido al mismo tiempo, a continuación se muestran los pasos que realiza dicho algoritmo.

2.7.1. Obtención del gradiente

Con el fin de implementar el algoritmo de detector de borde Canny, una serie de pasos deben seguirse. El primer paso es eliminar cualquier ruido en la imagen original antes de intentar localizar y detectar los bordes. Y debido a que el filtro gaussiano se puede calcular usando una máscara simple, se utiliza exclusivamente en el algoritmo de Canny. Una vez que se ha calculado una máscara adecuada, el suavizado gaussiano se puede realizar usando métodos de convolución estándar. Una máscara de convolución suele ser mucho más pequeña que la imagen real. Como resultado, la máscara se desliza sobre la imagen, manipulando un cuadrado de píxeles a la vez. Cuanto mayor es el ancho de la máscara gaussiana, menor es la sensibilidad del detector al ruido. El error de localización en los bordes detectados también aumenta ligeramente a medida que aumenta el ancho de Gauss. La máscara gaussiana se muestra a continuación.2.10

$$g(m, n) = G_{\sigma}(m, n) * f(m, n) \quad (2.16)$$

donde

$$G_{\sigma} = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{m^2 + n^2}{2\sigma^2}\right) \quad (2.17)$$

$\frac{1}{115}$ $\sigma=1.4$	2	4	5	4	2
	4	9	12	9	4
	5	12	15	12	5
	4	9	12	9	4
	2	4	5	4	2

Figura 2.10: Máscara Gaussiana

%myequations Función de máscara gaussiana

2.7.2. Supresión no máxima al resultado del gradiente

Las dos imágenes generadas en el paso anterior sirven de entrada para generar una imagen con los bordes adelgazados. El procedimiento es el siguiente: se consideran cuatro direcciones identificadas por las orientaciones de 0° , 45° , 90° y 135° con respecto al eje horizontal. Para cada pixel se encuentra la dirección que mejor se aproxime a la dirección del ángulo de gradiente.

-1	0	+1
-2	0	+2
-1	0	+1

Gx

+1	+2	+1
0	0	0
-1	-2	-1

Gy

Figura 2.11: Máscara Gaussiana

2.8. Propiedades de contorno

A continuación se muestran algunas propiedades de uso frecuente de objetos como solidez, diámetro, equivalente, imagen de máscara, intensidad media, etc.

- **Radio** Proporción entre ancho y altura del rectángulo del objeto.

$$Radio = \frac{Ancho}{Altura} \quad (2.18)$$

- **Extensión** Proporción entre el área del contorno y área del rectángulo delimitador.

$$Exten = \frac{Area \ del \ contorno}{Area \ rectangulo \ delimitador} \quad (2.19)$$

- **Solidez** Proporción entre el área del contorno y área convexa del contorno.

$$Solidez = \frac{Area \ del \ contorno}{Area \ convexa \ del \ contorno} \quad (2.20)$$

- **Diámetro Equivalente** Es el diámetro del círculo cuya área es igual que el área del contorno.

$$DiámetroEquivalente = \sqrt{\frac{4x \ Area \ del \ contorno}{\pi}} \quad (2.21)$$

- **Orientación** Orientación es el ángulo en el que se dirige el objeto. Con el siguiente método también da las longitudes del eje mayor y del eje menor.

Capítulo 3

Metodología

Consideramos al sistema de ingreso y egreso de personas constituido por un conjunto de formaciones y se tomará cada formación como unidad de estudio. Los motivos de esta decisión son que los datos medidos y sus errores dependen de:

- **Enfoque:** Hay que tener muy en cuenta el enfoque y escena en donde va a ser implementado el conteo de personas ya que tiene mucho que ver la luminosidad que exista para la respuesta óptima del algoritmo. En nuestro caso el enfoque será de manera cenital ya que nos ayuda a tener mayor enfoque en la parte de la cabeza y realizar una detección y conteo con mas precisión.
- **Recorrido de las personas:** Tiene mucho que ver el recorrido de las personas al momento de su detección. Se puede tener algunos casos de ingreso o salida de personas como por ejemplo, cuando pasa una persona junto a otra, cuando se queda parado en la detección o cuando pasan agachados, entre otros. En algunos casos el sistema responde adecuadamente pero en otros posee errores referentes a la luminosidad y a la calidad de video que se esta manejando.
- **Día/Hora:** Tiene que ver el día y hora porque según eso se tiene una problemática diferente. También tiene que ver en cuánto a la luminosidad existente ya que si existen cambios bruscos de iluminación el algoritmo tarda en responder adecuadamente y por lo tanto no se tiene un resultado lo más óptimo posible.

3.1. Información preliminar

El software que se va a desarrollar tiene como objetivo detectar características de personas, a través de algoritmos computacionales de visión artificial ya existentes.

- **Cantidad de Personas:** En lo referente a la cantidad de personas no habría inconveniente ya que el software posee contar varias personas tanto hombres y mujeres que ingresen o salgan del sitio, pero el proceso de entrada es diferente al proceso de salida, por lo que se espera tener distintos valores y distribución del error en cada caso.

- **Ocupación máxima en la región:** El software está diseñado para el conteo de una persona, al momento de pasar dos personas al mismo tiempo por el área de detección el algoritmo lo comprenderá como uno solo, por lo tanto es una condición del sistema.
- **Parada de ocupación nula o mínima:** Hay que tener en cuenta la posición y parada que tiene la persona a detectar, ésto generaría errores cuando la persona permanezca en el área de detección en cuánto al conteo, por lo tanto es recomendable avanzar y no quedarse en el área de detección para evitar conflictos con la respuesta del sistema.
- **Recorrido:** en lo referente al recorrido se podría tener inconvenientes al momento del paso de la persona a diferentes velocidades, ésto hace que tarde en responder adecuadamente el algoritmo. Sin embargo, se puede reducir el margen de error mediante los límites del área de conteo y detección de personas.

3.2. Posicionamiento de la cámara

- **Para cada posición se conoce diferente algoritmo:** se tiene que según la escena o entorno en la que se va a implementar el sistema tiene la persona diferente posicionamiento, por lo tanto, para cada posición de enfoque o de la persona a detectar y contar hay algoritmos existentes para su implementación. En este caso, se ubicará la cámara a 90° con referencia al piso para su respectivo conteo, es decir, se ubicará de manera cenital (enfoque arriba hacia abajo). Véase fig 3.2



Figura 3.1: Diferentes posicionamientos de la cámara: **a)** enfoque frontal, **b)** enfoque diagonal, **c)** enfoque cenital

3.3. Contador de personas

Para el conteo y detección de personas se implemento un algoritmo adecuado. En la parte de programación se tiene un área de detección, posee límites de entrada y salida(como se muestra en la imagen 3.3) de la persona, éstos límites varían según la calidad del video ya que si se tiene videos de alta y baja resolución en donde el área de detección debe ser mas amplia o reducida según el entorno para poder detectar sin menor rango de error.



Figura 3.2: Conteo de personas

3.3.1. Ingresos y salidas de personas

Para el ingreso y salida de personas hay que tener en cuenta la velocidad a la que pasan, si se detienen, cambian de posición, se agachan o ingresan con algún objeto, porque al tener algunas condiciones el sistema implementado no tendría tanta precisión en su funcionamiento. También tiene que ver la distancia a la que está la cámara de la persona a detectarse ya que por la altura se pueden generar errores ya que hay personas que son mas altos que otros.

3.4. Sistema de validación

El sistema de validación consta de un área de detección y los límites superior e inferior que determinan los límites de entrada y salida de la persona. El proceso del sistema de detección consta en que, al momento de que la persona es enfocada por la videocámara es detectada pero al momento que ingresa en el área de detección que es en forma rectangular y es calculada referente al ancho y largo del video, ésta área restringe la detección a otros elementos que no posean las características que tienen las personas. Esto se lo realizó mediante un detector de bordes de Canny que lo primero que hace es suavizar la imagen para eliminar el ruido. Luego encuentra el gradiente de la imagen para resaltar las regiones con altas derivadas espaciales. El algoritmo sigue pistas a lo largo de estas regiones y suprime cualquier pixel que no esté al máximo (supresión no máxima). Al distinguir las características de la persona el sistema la detecta y entra en una condición tanto de entrada o de salida ya que puede dirigirse a cualquiera de los dos sentidos sin que afecta el sistema. La validación se da cuando la persona salga del área de detección y decida entrar o salir del sitio, al darse ésto el sistema realiza el conteo tanto de ingreso como de salida según corresponda. véase imagen 3.4 A continuación se muestra un

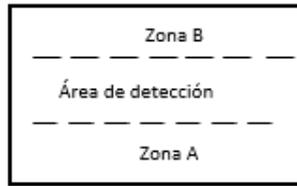


Figura 3.3: Área de detección

diagrama de flujo de la aplicación que se está realizando, la cuál presenta todas las etapas que tiene el algoritmo así como la relación entre ellas. Por lo tanto se requiere analizar cada módulo y investigar la teoría de las técnicas utilizadas en la aplicación.

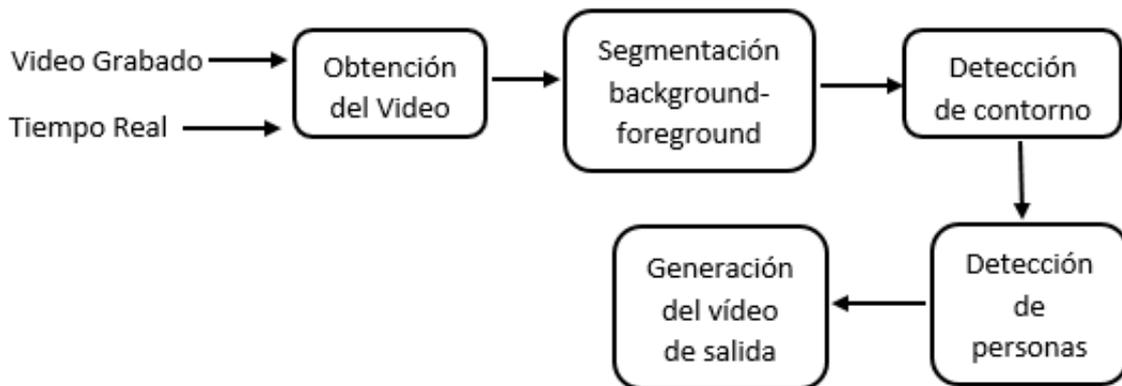


Figura 3.4: Diagrama de bloques de la aplicación

3.5. Segmentación

3.5.1. Extracción de fondo

Se usó el extractor de fondo MOG basándose en la concordancia de cada una de las gaussianas de la mezcla, se determina si corresponden a los colores de fondo. Luego se modelan los valores de un pixel particular como una mezcla de distribuciones gaussianas y los valores de pixel que no correspondan con colores background son considerados foreground hasta que exista una gaussiana que los incluya.

En primer lugar se debe saber que si un pixel resulta de una superficie concreta bajo una iluminación determinada, una distribución gaussiana simple sería suficiente para modelar el valor del pixel. Por otro lado, si solamente se producen cambios de iluminación durante horas, una distribución gaussiana simple y adaptativa sería suficiente. Pero en la práctica, son múltiples superficies las que aparecen a menudo en la vista de un pixel particular y las condiciones de

iluminación cambian. Por lo tanto, en ese caso, son necesarias gaussianas múltiples adaptativas. Este método usa entonces una mezcla de gaussianas adaptativas para aproximar este proceso. A continuación se muestran conclusiones del algoritmo utilizado:

- Fondo de la escena(cualquier objeto que luego de un tiempo se incluye en el fondo): El algoritmo tiene respuesta óptima con los parámetros establecidos en escenas de interiores.
- Fondo dinámico: En muchas mas ocasiones se detecta dicho fondo dinámico como primer plano.

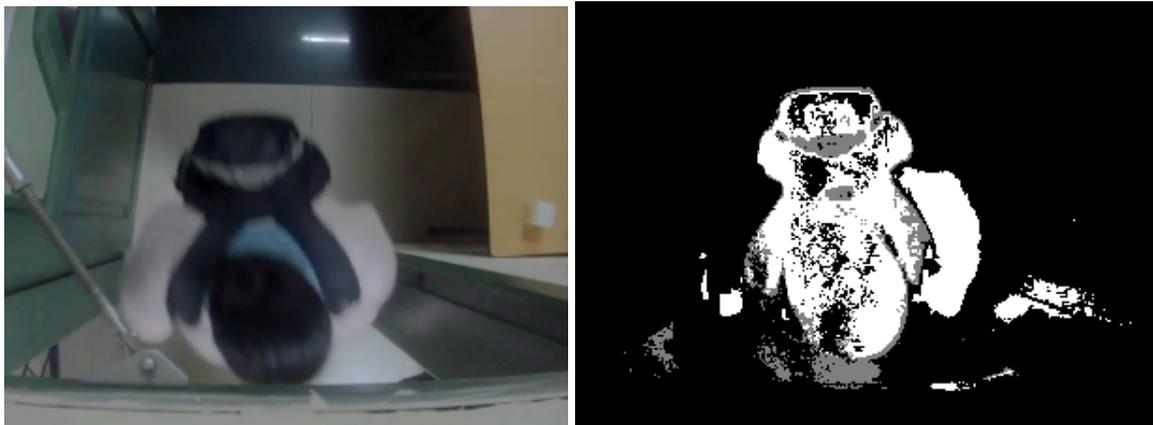


Figura 3.5: Respuesta del algoritmo MOG ante primer plano dinámico

- Cambios de iluminación: En este caso en ocasiones si se detectan las zonas afectadas por los cambios de iluminación como primer plano.
- Foreground estático: En el caso de cuando la persona se encuentre parada y realiza pequeños movimientos en el área de detección el algoritmo tiene menor tiempo de respuesta y el comportamiento es mejor.

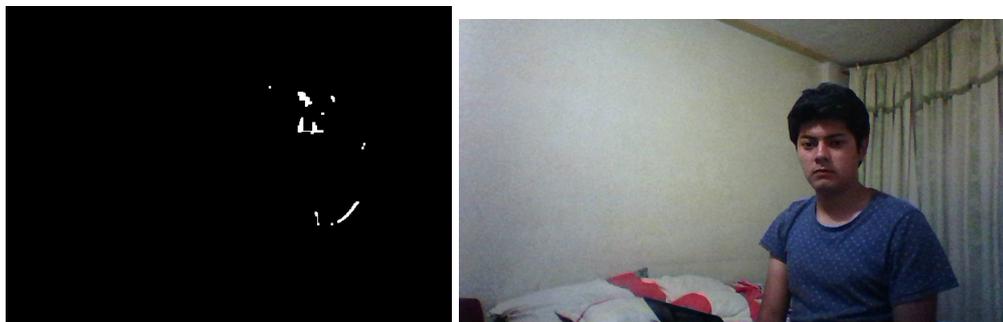


Figura 3.6: Respuesta del algoritmo MOG ante primer plano estático

En este caso se realizaron pruebas a distancias medias y cortas, en dichas distancias el algoritmo no posee problemas para la detección de movimiento, en medias distancias el algoritmo presenta mayor ruido debido a la sombra que se genera y por lo tanto detectará dos personas ya sea una encima de otra o alguien con mayor altura, pero su correcto funcionamiento sería a distancias cortas ya que presenta menor ruido y mejor comportamiento al momento de detectar movimiento [12].

3.6. Filtro morfológico

Como se explicó en el capítulo anterior sobre el filtro morfológico y las operaciones que se pueden aplicar para solucionar el problema tenemos que, las transformaciones morfológicas tienen como objetivo modificar la forma de los objetos que se observan en una imagen. En esta aplicación es muy necesario, puesto que la imagen de foreground resultante del módulo anterior contiene partes que se ven afectadas por la generación de regiones o píxeles con ruido. Por lo tanto, se puede tener zonas de movimiento falsas donde realmente no existe ningún tipo de movimiento, o cuando el movimiento es mínimo, por lo que se le consideraría una región despreciable. Para solucionar este problema se aplica las operaciones de erosión y dilatación que posee el filtrado morfológico.

3.6.1. Erosión

Es la degradación progresiva de uno de los campos que puede ser 0 o 1. Un elemento del campo a degradar seguirá perteneciendo al mismo si está rodeado de elementos semejantes, caso contrario pasará al otro campo. Matemáticamente se expresará de la siguiente forma: si se toma el elemento estructural simétrico respecto al origen de B , la erosión de un conjunto X respecto al elemento B es:

$$X \ominus B = \{B_X \subset X\} \quad (3.1)$$

Lo que es igual a una transformación acierta o falla donde B es el conjunto vacío. El símbolo \ominus representa la resta de Minkowski.

3.6.2. Dilatación

Es el crecimiento progresivo de uno de los campos (0 o 1). Un elemento del campo contrario a crecer será convertido si posee algún vecino perteneciente al campo que se expansiona. En caso contrario, permanecerá igual. Los elementos pertenecientes al campo a expansionar evidentemente no se modifican. Si se aplicase un numero elevado de veces terminaría por destruir la imagen ya que todos lo píxeles estarían a nivel alto. Matemáticamente se puede expresar la dilatación como:

$$X \oplus B(X^C \ominus B^C) \quad (3.2)$$



Figura 3.7: Ejemplo aplicando erosión



Figura 3.8: Ejemplo aplicando dilatación

3.6.3. Combinación de ambas

Una de las principales características de la dilatación y la erosión es que no cumplen la propiedad conmutativa, es decir, no es lo mismo realizar una dilatación seguida de una erosión, que una erosión seguida de una dilatación. En este proyecto se ejecuta en primer lugar dos repeticiones para la erosión y a continuación una para la dilatación. Cuando las operaciones se realizan en este orden el proceso recibe el nombre de apertura. Se denomina así ya que al empezar por una erosión se tiende a romper las piezas en sus partes constitutivas. Las principales causas por las que se elige este orden y no el inverso es que se consigue suavizar los contornos del objeto, se rompen enlaces delgados y se eliminan pequeñas protuberancias.

3.7. Detección de contorno

En cada frame se agrupan los píxeles en movimiento que se encuentran continuos entre sí como una sola región. A continuación se tiene que la imagen en la que se está trabajando los píxeles blancos representan las zonas que están en movimiento y los negros representan las zonas inmóviles. Dado que la máscara de foreground es una imagen binaria no es necesario convertir la imagen que tiene varios niveles de gris en una nueva con solo dos (Umbralización). Existe un método en el que va siguiendo la imagen empezando por la esquina superior izquierda hacia mano derecha hasta acabar de recorrer la imagen, cuando se localiza el primer píxel que se encuentre en nivel alto se le asigna un 1, indicando que se ha encontrado el primer objeto. Se continúa revisando detenidamente si los vecinos se encuentran también en nivel alto, si lo están se les asigna el mismo número. Otro caso es cuando un píxel no sea vecino de uno ya anteriormente etiquetado y si se encuentra en nivel alto, se le asigna el número siguiente que sería el 2 respectivamente.



Figura 3.9: Contorno de una persona

3.8. Detección de personas

A partir de la detección de contorno se determina si éste pertenece a personas o no, tomando en cuenta que se requiere detección de personas como objetivo. Primeramente se realiza un filtrado de áreas, tomando en cuenta que se filtren las áreas que estén sobre o por debajo de un valor. Tanto que el objetivo principal es la detección y conteo de personas en un determinado rango de distancia, y posee ciertos límites de áreas mayores o menores. Luego se aplica el siguiente filtro que tiene que ver con la relación largo y ancho de la persona detectada, tanto ésta se encuentre de manera erguida o sentada, si en un caso se requiere ingresar personas en fila cojidas de las manos se ampliaría el límite ya sea superior o inferior para evitar conflictos en la detección. Debido a la restricción de límites que posee se puede limitar la detección de objetos que tengan características similares al de la persona en lo referente a su ancho y a la proporción de su alto. También hay que tener muy en cuenta en el área de detección que la misma englobe el contorno de la persona y si no se lo hace no se detectaría la persona o provocaría falsas detecciones. Otro punto importante es la resolución del video, ya que a menor resolución del video se tiene una detección y conteo mas óptimo en comparación con resoluciones altas. Con todo lo comentado anteriormente se obtiene un muy buen porcentaje de detecciones correctas, pero también se obtiene cierto margen de error por lo tanto no es totalmente fiable.

Finalmente luego de haber explicado como se realizó la detección de personas cabe recalcar que el programa esta designado al conteo de una sola persona y el proceso consta en detectar la posición de la persona y localizar cada frame de la misma. Por otro lado cuando existe mas de una sola persona en el área de detección, el programa no puede realizar la detección de x personas en x posiciones [13].



Figura 3.10: Detección de una persona

3.9. Video de salida

Esta sección es la que se ocupa de la generación del video de salida, siendo el resultado final de la aplicación. El video de salida es ejecutado constantemente y en éste se muestra cada resultado al momento de su funcionamiento. Este proceso inicia en la obtención de imágenes por medio de una videocámara, dicho paso es sencillo ya que es realizado automáticamente. La aplicación consta de dos ventanas, en la primera ventana se aprecia la imagen de entrada ya sea en forma de video grabado o en tiempo real, teniendo en cuenta que el área de detección englobe a la persona detectada en cada posición. Al tener la ventana en la que se visualiza la entrada y salida de las personas se observa el momento en el que la persona es detectada y al pasar los límites ya sea superior o inferior se tiene el conteo, ésto se lo hace para mejorar la visualización del usuario. Se lo realiza con el fin de que el usuario visualice su zona de videovigilancia y tenga constancia de lo que sucede en ese momento y como se comporta el algoritmo en entornos y escenas internas. Se visualiza de igual manera el número de personas detectadas en cada instante tanto entrantes como salientes, la información se almacena para luego utilizarla como el usuario desee. Por último también se puede observar en la segunda ventana cuando se aplica el filtro morfológico, con ésto la persona puede tener constancia de lo que sucede en la aplicación y si la detección y conteo de personas es preciso o no.

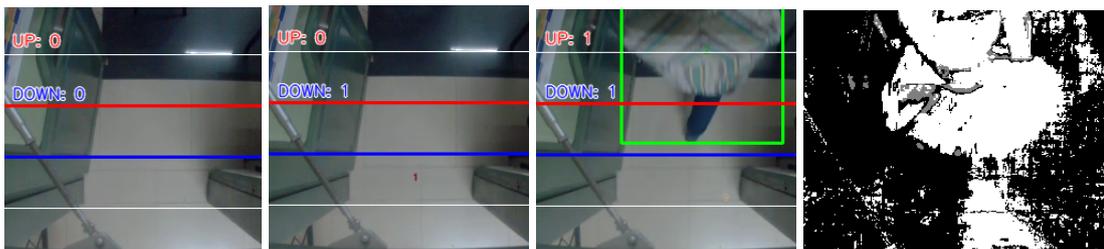


Figura 3.11: Detección de una persona

Capítulo 4

Pruebas y resultados

Para las pruebas y resultados se grabaron secuencias de imágenes con diferentes resoluciones con una cámara Motion Pro Action Camera la cuál nos permite grabar en distintas resoluciones el video, la cámara esta ubicada a una altura de (2.30m) y a 90° con referencia del suelo, en un entorno interior. En el primer video se grabó a una resolución de 480x480 píxeles, y veinte y cuatro frames por segundo. La implementación del contador de personas realizado



Figura 4.1: Imágenes de video de resolución 480x480 píxeles

sobre la Fig 4 se obtuvo como resultado lo siguiente:

- Se grabó un video en el cuál se observa el paso de personas tanto entrando como saliendo, en dicho video entran 4 personas y salen 4, en el cuál se obtiene falsos positivos, falsos negativos y la detección correcta de personas. Para la mejora del conteo se agrando el área de interés reduciendo asi los falsos positivos y negativos en un entorno interior.
- En ésta representación de imágenes se tiene que para la entrada se detectó tres personas además de detectar cuatro falsos positivos y un falso negativo que representa el error en el conteo de una persona.
- Para la salida se detectarán cuatro personas, cuatro falsos positivos y ningún falso negativo. Se tiene que en la salida existe menor error en cuanto al conteo de personas pero no en cuánto a lo referente con falsos positivos y negativos.

Cuadro 4.1: Resultados primer video

Personas	Personas Detectadas	Falsos Positivos	Falsos Negativos	Total Conteo
Sal.5	5	5	0	10
Ent.6	4	5	1	10

En el segundo video se grabó a una resolución de 320x240(VGA) píxeles y treinta frames por segundo. La implementación del contador de personas realizado sobre la fig 4 se obtuvo como



Figura 4.2: Imágenes de video de resolución 320x240 píxeles(VGA)

resultado lo siguiente:

- Se grabó un video en el cuál se observa el paso de hombres y mujeres entrando y saliendo de dicho entorno, en el video se tiene que entran 6 personas y salen 5 en el cuál se obtiene falsos positivos, falsos negativos y las detecciones correctas. Para la mejora del conteo se agrando el área de interes para el conteo reduciendo asi los falsos positivos y negativos en la escena propuesta.
- En ésta representación de imágenes se tiene que para la entrada se detectó cinco personas además de detectar ocho falsos positivos y un falso negativo que representa el error en el conteo de una persona.
- Para la salida se obtuvo que se detectó cinco personas, tres falsos positivos y ningún falso negativo. Se tiene que en la salida existe menor error en cuanto al conteo de personas referente a las características del video.

Cuadro 4.2: Resultados segundo video

Personas	Personas Detectadas	Falsos Positivos	Falsos Negativos	Total Conteo
Sal.5	5	3	0	8
Ent.6	5	8	1	13

El tercer video se grabó a una resolución de 320x240 píxeles(VGA) y treinta frames por segundo. La implementación del contador de personas realizado sobre la fig 4 se obtuvo como

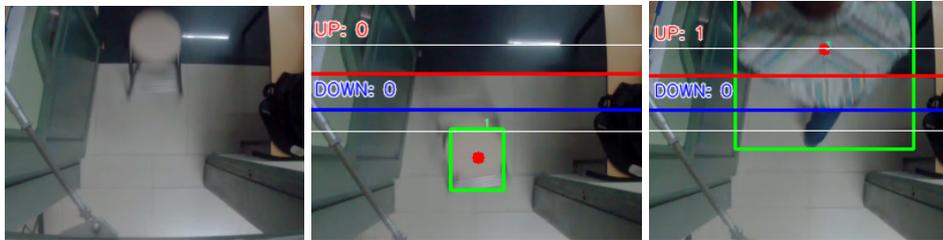


Figura 4.3: Imágenes de video de resolución 320x240 píxeles(VGA)

resultado lo siguiente:

- En este video se realizaron pruebas referentes a conteo de objetos y personas, el resultado que se obtuvo fue que realizó el conteo de personas y objetos, pero para corregir dicho error lo que se hizo fue disminuir los rangos entre línea de entrada y línea de salida, al hacer esto se tiene que el objeto no es contado pero si la persona.
- En ésta representación de imágenes se tiene que entran seis personas y salen cinco personas, para la entrada se detectó cinco personas pero también se detectó tres falsos positivos y un falso negativo.
- Para la salida se obtuvo que se detectó cinco personas, siete falsos positivos y ningún falso negativo. Se tiene que en la salida existe menor error en cuanto al conteo de personas pero se tiene mas falsos positivos que en la entrada.

Cuadro 4.3: Resultados tercer video

Personas	Personas Detectadas	Falsos Positivos	Falsos Negativos	Total Conteo
Sal.5	5	7	0	12
Ent.6	5	3	1	9

En este video se presencia la misma resolución hablada en el primer video pero con diferentes condiciones, por lo tanto se obtendrá diferentes resultados. La implementación del contador de personas realizado sobre la fig 4 se obtuvo como resultado lo siguiente:

- En este video se realizaron pruebas referentes a conteo de objetos y personas en el cuál los rangos de área de interés son los mismos probados anteriormente pero con resultados diferentes.

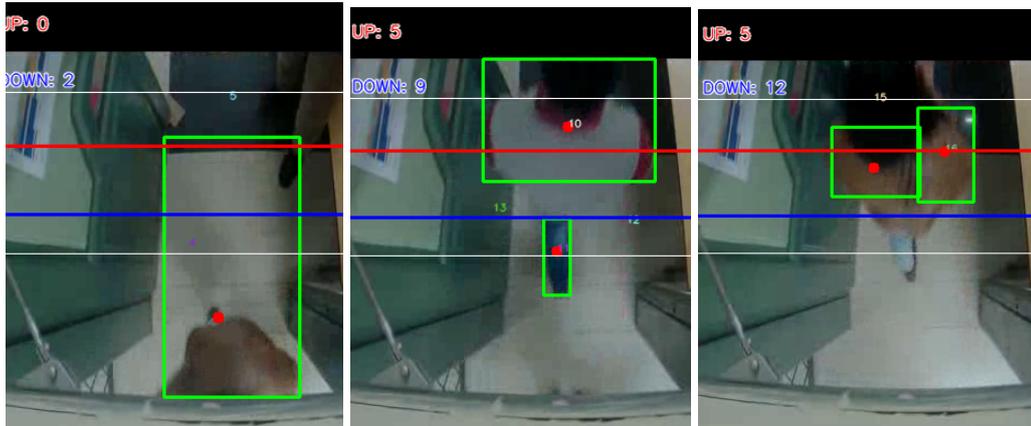


Figura 4.4: Imágenes de video de resolución 320x240 píxeles(VGA)

- En ésta representación de imágenes se tiene que entran cuatro personas y salen cuatro personas, para la entrada se detectó dos personas pero también se detectó un falso positivo y dos falsos negativos.
- Para la salida se obtuvo que se detectó tres personas, ocho falsos positivos y un falso negativo. Se tiene que tanto en la entrada como en la salida se tiene mas errores de conteo que las anteriores pruebas.

Cuadro 4.4: Resultados tercer video

Personas	Personas Detectadas	Falsos Positivos	Falsos Negativos	Total Conteo
Sal.5	4	8	2	14
Ent.6	3	1	1	5

Una última prueba se realizó en tiempo real con una cámara Logitech ubicada en la misma posición y altura que la anteriormente nombrada, en los resultados obtenidos de dicha prueba se tuvieron mayores inconvenientes con la detección de falsos positivos y negativos, pero claramente se noto que fue efecto de la resolución que posee por defecto la cámara, la segunda prueba que se realizó fue con una resolución de 160x120 píxeles, ante éste cambio los resultados tienen un menor porcentaje de error en el conteo de personas.

Capítulo 5

Conclusiones y Trabajo futuro

El presente trabajo presenta un sistema de conteo de personas a través de una cámara ubicada en posición cenital, diferenciando entre conteo de entrada y salida. Se propone un modelo para la detección de cabezas, suavizando la imagen para así eliminar el ruido y detecciones erróneas, para luego usar el filtro de Canny como base para realizar el conteo y seguimiento de personas. Por medio de las especificaciones del filtro de Canny posee buena detección, buena localización y respuesta mínima, resolviendo así problemas de oclusión temporal que se producen en los videos de test.

La aplicación presenta un funcionamiento autónomo ya que el usuario en ningún momento manipula variables durante la ejecución de la aplicación. Además se tiene un bajo costo en la etapa de adquisición, y se demostró que mediante una cámara de baja calidad el sistema tiene un buen funcionamiento, y la obtención de imágenes se las realiza en tiempo real y por medio de videos anteriormente grabados, es mejor realizarlo en tiempo real para evitar pérdidas de información y falsas detecciones.

La técnica de sustracción de fondo permitió obtener una mejor detección de movimiento ya que permite segmentar objetos de interés que se encuentren en la escena. Con una resolución de 160x120 píxeles, así se tiene información suficiente para detectar personas que pasan por el área de conteo. Además se implemento una metodología para el seguimiento de las personas reconocidas en la secuencia de video, mediante la predicción de movimiento de sus centroides por lo que los transeúntes no presentaban un mismo patrón entre frame y frame, adicionalmente el seguimiento y detección ayuda a tener un registro de el monto de personas que entran o salen de la escena.

El sistema presentó algunos errores en lo referente a calidad de imagen en donde se presentan falsas detecciones cuando no existe una diferencia considerable en el contraste entre el fondo y la persona, la solución para esto es reducir la calidad del video hasta tener un número adecuado de frames por segundo, así se evitan errores y se mejora el funcionamiento del sistema.

Trabajo futuro

En el caso de buscar una solución al momento de detectar dos o más personas al mismo tiempo se deberían implementar otras técnicas mas robustas que logren definir correctamente

los contornos de las personas y las identifique independientemente, además para mejorar en lo referente al seguimiento de personas se puede aplicar, filtro de kalman que es usado para la predicción de movimiento, tracking en base a colores y trabajar con imágenes en distintos instantes de tiempo y tratar de relacionarlas entre si.

Si se requiere implementar el algoritmo en un ambiente exterior se debe investigar un modelo de fondo que sea mas robusto a objetos que se encuentren en movimiento y pertenezcan al fondo, cuando se tengan cambios fuertes de iluminación, tener sombras de objetos que no son de interés. Si se tiene éstos objetos y ingresan en la escena, el sistema realizaria el seguimiento además de el conteo del objeto. Para evitar dichas falsas detecciones se realiza una extracción de características que permita una clasificación que se base en la posición, silueta y la dirección en la que se mueve el objeto que se empezó a seguir.

Apéndice A

Listado de códigos

En este apéndice incluye el programa implementado para el conteo de personas.

A.1. Programa implementado

Listing A.1: Conteo de Personas

```
import numpy as np
import cv2
import Person
import time

cnt_up = 0
cnt_down = 0

#cap = cv2.VideoCapture(1)
cap = cv2.VideoCapture('/home/jhonny/Prueba.MP4')

cap.set(3,160)
cap.set(4,120)

for i in range(19):
    print i, cap.get(i)

w = cap.get(3)
h = cap.get(4)
frameArea = h * w
areaTH = frameArea / 250
print 'Area Threshold', areaTH

line_down = int(3 * (h / 5))
line_up = int(2 * (h / 5))
```

```

up_limit = int(1 * (h / 5))
down_limit = int(4 * (h / 5))

print "Red line y:", str(line_down)
print "Blue line y:", str(line_up)
line_down_color = (255, 0, 0)
line_up_color = (0, 0, 255)
pt1 = [0, line_down];
pt2 = [w, line_down];
pts_L1 = np.array([pt1, pt2], np.int32)
pts_L1 = pts_L1.reshape((-1, 1, 2))
pt3 = [0, line_up];
pt4 = [w, line_up];
pts_L2 = np.array([pt3, pt4], np.int32)
pts_L2 = pts_L2.reshape((-1, 1, 2))

pt5 = [0, up_limit];
pt6 = [w, up_limit];
pts_L3 = np.array([pt5, pt6], np.int32)
pts_L3 = pts_L3.reshape((-1, 1, 2))
pt7 = [0, down_limit];
pt8 = [w, down_limit];
pts_L4 = np.array([pt7, pt8], np.int32)
pts_L4 = pts_L4.reshape((-1, 1, 2))

# Substractor de fondo
fgbg = cv2.createBackgroundSubtractorMOG2(detectShadows=True)

# Elementos estructurantes para filtros morfologicos
kernelOp = np.ones((3, 3), np.uint8)
kernelCl = np.ones((11, 11), np.uint8)
font = cv2.FONT_HERSHEY_SIMPLEX
persons = []
max_p_age = 5
pid = 1

while (cap.isOpened()):
    ret, frame = cap.read()

    for i in persons:
        i.age_one()

    # PRE-PROCESAMIENTO
    fgmask = fgbg.apply(frame)
    fgmask2 = fgmask.apply(frame)

    # Binarizacion
    try:
        ret, imBin = cv2.threshold(fgmask, 200, 255, cv2.THRESH_BINARY)
        ret, imBin2 = cv2.threshold(fgmask2, 200, 255, cv2.THRESH_BINARY)
        mask = cv2.morphologyEx(imBin, cv2.MORPH_OPEN, kernelOp)
        mask2 = cv2.morphologyEx(imBin2, cv2.MORPH_OPEN, kernelOp)
        mask = cv2.morphologyEx(mask, cv2.MORPH_CLOSE, kernelCl)
        mask2 = cv2.morphologyEx(mask2, cv2.MORPH_CLOSE, kernelCl)
    except:
        print('EOF')
        print 'UP:', cnt_up
        print 'DOWN:', cnt_down
        break

    # CONTORNOS
    _, contours0, hierarchy = cv2.findContours(mask2, cv2.RETR_EXTERNAL, cv2.CHAIN_APPROX_SIMPLE)
    for cnt in contours0:

```

```

area = cv2.contourArea(cnt)
if area > areaTH:

    # TRACKING
    M = cv2.moments(cnt)
    cx = int(M['m10'] / M['m00'])
    cy = int(M['m01'] / M['m00'])
    x, y, w, h = cv2.boundingRect(cnt)

    new = True
    if cy in range(up_limit, down_limit):
        for i in persons:
            if abs(cx - i.getX()) <= w and abs(cy - i.getY()) <= h:
                new = False
                i.updateCoords(cx, cy)
                if i.going_UP(line_down, line_up) == True:
                    cnt_up += 1;
                    print "ID:", i.getId(), 'crossed going up at', time.strftime("%c")
                elif i.going_DOWN(line_down, line_up) == True:
                    cnt_down += 1;
                    print "ID:", i.getId(), 'crossed going down at', time.strftime("%c")
                break
            if i.getState() == '1':
                if i.getDir() == 'down' and i.getY() > down_limit:
                    i.setDone()
                elif i.getDir() == 'up' and i.getY() < up_limit:
                    i.setDone()
            if i.timedOut():
                index = persons.index(i)
                persons.pop(index)
                del i
        if new == True:
            p = Person.MyPerson(pid, cx, cy, max_p_age)
            persons.append(p)
            pid += 1

    cv2.circle(frame, (cx, cy), 5, (0, 0, 255), -1)
    img = cv2.rectangle(frame, (x, y), (x + w, y + h), (0, 255, 0), 2)

#DIBUJAR TRAYECTORIAS
for i in persons:

    cv2.putText(frame, str(i.getId()), (i.getX(), i.getY()), font, 0.3, i.getRGB(), 1, cv2.LINE_AA)

#IMAGANES
str_up = 'UP: ' + str(cnt_up)
str_down = 'DOWN: ' + str(cnt_down)
frame = cv2.polylines(frame, [pts_L1], False, line_down_color, thickness=2)
frame = cv2.polylines(frame, [pts_L2], False, line_up_color, thickness=2)
frame = cv2.polylines(frame, [pts_L3], False, (255, 255, 255), thickness=1)
frame = cv2.polylines(frame, [pts_L4], False, (255, 255, 255), thickness=1)
cv2.putText(frame, str_up, (10, 40), font, 0.5, (255, 255, 255), 2, cv2.LINE_AA)
cv2.putText(frame, str_down, (10, 40), font, 0.5, (0, 0, 255), 1, cv2.LINE_AA)
cv2.putText(frame, str_down, (10, 90), font, 0.5, (255, 255, 255), 2, cv2.LINE_AA)
cv2.putText(frame, str_down, (10, 90), font, 0.5, (255, 0, 0), 1, cv2.LINE_AA)

cv2.imshow('Frame', frame)

```

```
k = cv2.waitKey(30) & 0xff
if k == 50:
    break
cap.release()
cv2.destroyAllWindows()
```

Bibliografía

- [1] Z. Z. Y. X. Bin Li, Jian Zheng, “A People Counting Method Based on Head Detection and Tracking”, Jul. 2011.
- [2] S. S. Sumati Manchanda, “Analysis of Computer Vision based Technique for Motion Detection”, Jun. 2016.
- [3] R. G. Jiménez, “Detección y conteo de personas, a partir de mapas de profundidad cenitales capturados con cámaras TOF”, Dec 2015.
- [4] E. M. J. Diego Alberto Vera Meaurio, “Conteo de personas en imagen y video mediante la técnica de Viola-Jones a través de clasificadores Haar utilizando software libre”, Feb. 2014.
- [5] C. S. K. Fabiana da Silva Guizi, “Real-time People Detection and Tracking using 3D Depth Estimation”, Apr. 2016.
- [6] K.-W. L. Young-Ho Suh, Sang Keun Rhee, “Continuous Location Tracking of People by Multiple Depth Cameras”, Sep. 2015.
- [7] J. M. ValenciaLópez, “Registro de transeúntes en tiempo real utilizando un sistema de visión artificial sobre un ambiente controlado”, 2007.
- [8] H. Ma and M. Z. Huchuan Lu, A Real-time Effective System for Tracking Passing People, Jul. 2008.
- [9] I. A. S. M. Analí J. Alfaro Alfaro, “Diseño de un Algoritmo de Segmentación de Imágenes aplicando el Funcional de Mumford-Shah para mejorar el desempeño de los Algoritmos Clásicos de Segmentación”, Mar. 2010.
- [10] R. B. L. Gervasoni, J.P.D amato, “Un Método eficiente para la sustracción de fondo en videos usando Gpu”, Sep. 2014.
- [11] R. G. y R.e.Wood, Digital Image Processing, Dec. 1993.
- [12] D. Lefloch, “Real-Time People Counting system using Video Camera”. Jan, 2007.
- [13] U. Gonzáles Benítez, “Detección de personas a partir de visión artificial”, Sep. 2010

- [14] G. Córdova Manobanda, “Recuento de personas en espacios cerrados mediante sensores RGB+D”. *España*, 2016.
- [15] [Online]. Available: <http://docs.opencv.org/master/d9/df8/tutorial/root.html> #Opencvmaster