Universidad de Alcalá
Escuela Politécnica Superior

Grado en Ingeniería Electrónica de Comunicaciones

Trabajo Fin de Grado

Detección y conteo de personas, a partir de mapas de profundidad cenitales capturados con cámaras TOF.

Autor: Raquel García Jiménez

Tutores: Cristina Losada Gutiérrez y Carlos Andrés Luna Vázquez

2015

UNIVERSIDAD DE ALCALÁ

ESCUELA POLITÉCNICA SUPERIOR

Grado en Ingeniería Electrónica de Comunicaciones

Trabajo Fin de Grado

Detección y conteo de personas, a partir de mapas de profundidad cenitales capturados con cámaras TOF.

Autor: Raquel García Jiménez

Directores: Cristina Losada Gutiérrez y Carlos Andrés Luna Vázquez

Tribunal:

Presidente: Javier Macías Guarasa

Vocal 1º: Jesús Ureña Ureña

Vocal 2º: Cristina Losada Gutiérrez

Calificación: .....

Fecha: .....

Agradecimientos

Recuerda, todo empezó con un ratón.

Walt Disney.

Después de 5 años que en un principio parecían una eternidad, hoy puedo decir ¡¡Muchas gracias!! a toda la gente que me ha acompañado, que han conseguido hacer de estos años los más cortos y divertidos que puedo recordar.

Resumen

El objetivo de este proyecto es la detección y conteo de personas, a partir de imágenes de profundidad obtenidas mediante un sensor basado en tiempo de vuelo (ToF), situado en posición cenital en un entorno interior.

Para conseguir este objetivo se han estudiado diferentes trabajos en este área y se ha implementado una solución basada en la extracción de características relacionadas con la superficie de la persona vista desde el sensor, y la clasificación posterior mediante el análisis de componentes principales (PCA).

El algoritmo desarrollado se ha evaluado sobre una base de datos grabada y etiquetada para ello, obteniendo una tasa de aciertos en torno al 95 %.

Palabras clave: Cámaras de tiempo de vuelo, detección personas, PCA

Abstract

The aim of this project is the detection and counting of people, using depth images obtained by a sensor based on time of flight (TOF), located in zenith position in an indoor environment.

To achieve this, several works in this area have been studied, and a solution has been proposed. It is based on the extraction of features from the surface of the person seen from the sensor, and their clasification using the principal components analysis(PCA).

The developed algorithm has been tested with a database recorded and labelled for it, and a success rate of 95 % was obtained.

Keywords: Time of fligth cameras, people detection, PCA.

Resumen extendido

Este trabajo tiene como fin el desarrollo de un sistema capaz de detectar y contar personas a partir de imágenes de profundidad extraidas de sensores basados en tiempo de vuelo (ToF). El principal objetivo del detector implementado es diferenciar personas de cualquier otro objeto presente en una escena.

Este proyecto parte del estudio realizado en [1], en el cual se propone un contador de personas mediante cámaras de tiempo de vuelo (ToF) empleando como detector el filtro basado en sombrero mejicano. Trás un análisis del mencionado trabajo se exponen los incovenientes de su uso y se propone una nueva técnica basada en la extracción de las características principales que tendría la forma de una persona vista desde una posición cenital.

Para implemenar el detector se emplea el sensor de tiempo de vuelo (TOF) implementado en Kinect II, utilizando para el desarrollo del algoritmo una programación en lenguaje C/C++ empleando la librería de tratamiento de imágenes *OpenCV*, sobre una plataforma Linux.

Este tipo de sensores se basa en la medición de la profundidad, entendiendo como profundidad la distancia existente entre el sensor y un punto determinado de la escena. Este proceso se realiza mediante la emisión de un haz de luz modulada a una determinada frecuencia, su posterior captación por parte de un array de receptores y el análisis de la diferecia de fase entre la señal emitida y la recibida mediante técnicas de correlación. Una vez obtenida la diferencia de fases se genera una matriz de profundidades cuyo tamaño vendrá establecido por el tamaño del array de receptores, la cual contendrá en cada uno de sus píxeles la información relativa a la profundidad del punto equivalente en la escena.

El sensor Kinect II proporciona tres imágenes diferentes, imagen de amplitud, imagen de escala de grises e imagen de profundidad, como se muestra en (Figura 2). Debido a problemas relacionados con la privacidad de la identidad de las personas, la única información utilizada será la relacionada con la imagen de profundidad.

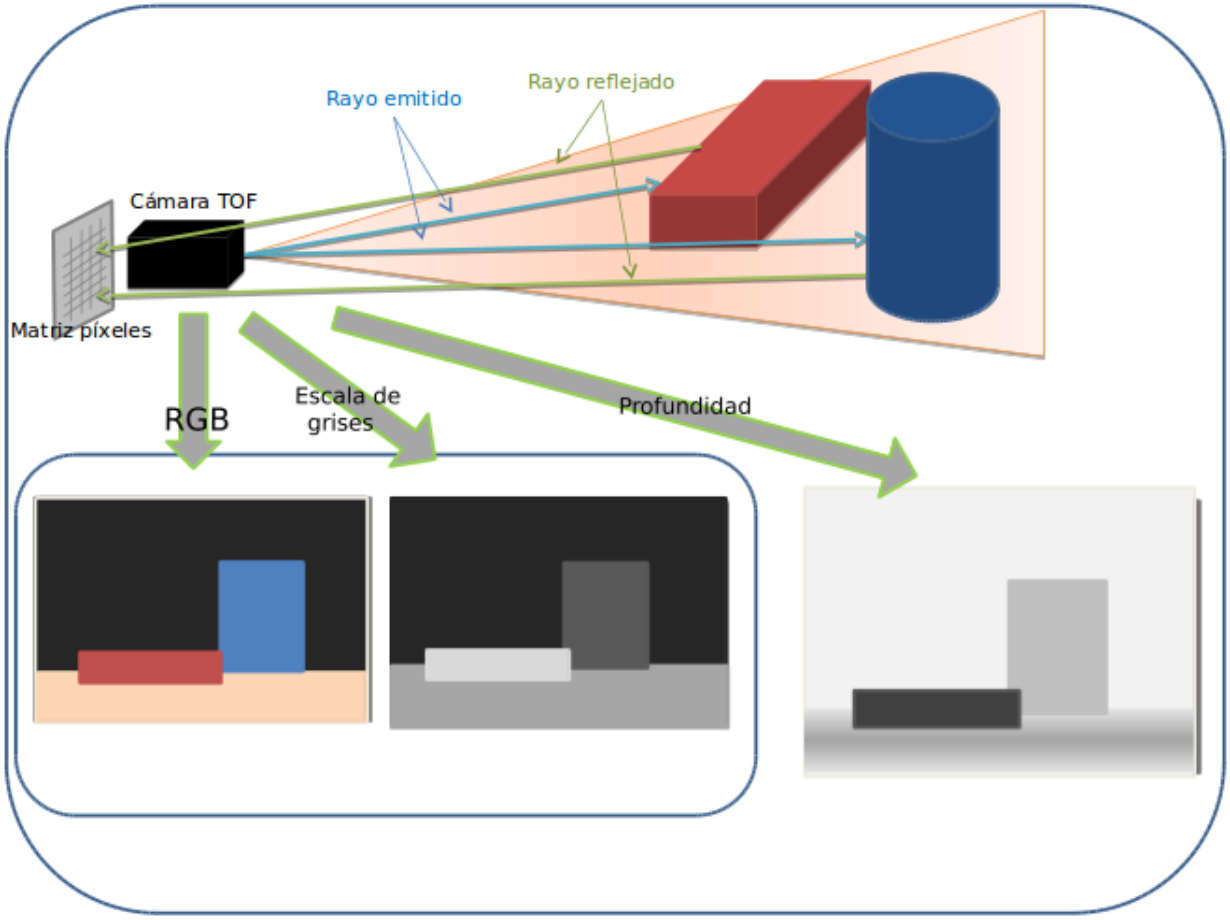


Figura 2: Tipos de imágenes obtenidas a partir del sensor Kinect II

El sistema de detección de personas propuesto está formado por dos procesos diferentes, *"On-line"* formado por seis bloques relativos al tratamiento y análisis de la imagen, y un proceso *"Off-line"*, donde, mediante el uso de una base de datos grabada y etiquedata previamente, se entrenan las diferentes clases que se utilizarán en el clasificador.

1. Obtención de la matriz de alturas como una imagen captada por el sensor. Los valores de los píxeles contendrán los valores inversamente proporcionales a la profundidad.
2. Filtrado de la imagen con el fin de eliminar los diversos ruidos por los que se ven afectadas las señales recibidas, incluyendo además, una estimación de valores de los píxeles erroneos indicados por el sensor.
3. Detección de los máximos de profundidad dentro de la matriz. Cada uno de estos máximos será evaluado en las etapas siguientes, ya que se podrá considerar como un candidato a persona,
4. Extracción del vector de características de cada máximo localizado. Este vector de características definirá tanto la forma de la persona u objeto vista desde el sensor, como la relación de cicularidad de su parte superior.
5. Clasificador de los diferentes vectores de características mediante el agoritmo de extracción de componentes principales PCA empleando clases definidas durante el entrenamiento.
6. Contador de los candidatos a personas aceptados como personas en la etapa de clasificación.

La respuesta obtenida de nuestro sistema será la posición en la escena de cada uno de los máximos detectados como personas. Para evaluar los resultados obtenidos se han utilizado secuencias pertenecientes a la base de datos de imágenes ToF, empleando secuencias con personas aisladas y varias personas andando aleatoriamente o en grupo.

Una vez analizados los resultados se ha concluido que:

- El algoritmo tomado como punto de partida basado en el filtrado mediante sombrero mejicano, no cumple una función como detector de personas sino como detector de contornos, debido a esto no sería válido para escenas donde aparecen otros elementos a parte de personas o en secuencias donde los individuos que aparecen se ecuentran muy juntos.
- Los resultados obtenidos tienen una alta tasa de detecciones correctas, en torno al 95 %, para personas aislas en la escena y para situaciones donde la oclusión de unas personas con otras no se produce de frente o de espaldas. En situaciones donde las personas se mueven en grupo o se encuentran de frente o espaldas unos a otros, la tasa de detecciones correctas se sitúa entorno al 85 %. En ambos casos las deteciones de otros elemetos como personas son inferiores al 1 %.

Índice general

Índice de figuras

Índice de tablas

Lista de acrónimos

CCD	Charged Coupling Devices.
CMOS	Complimentary Metal Oxide Semiconductor.
CWM	Continuous Wave Camera.
LoG	Laplacian of Gaussian.
ROI	Region Of Interest.
SoC	System on Chip.
TOF	Time Of Flight.

Capítulo 1

Introducción

Actualmente las aplicaciones basadas en la detección y el conteo de personas dentro de un espacio presenta gran interés en diferetes aplicaciones, tales como video vigilancia, control de accesos, análisis de flujos de movimiento de personas o control de aforos. Además, este tipo de aplicaciones han adquirido cada vez más importancia en los útimos años, dadas las necesidades de prevención y detección de situaciones potencialmente peligrosas.

Para tratar de resolver este problema se han planteado númerosas propuestas con elementos invasivos, como la implementación de tornos de control de acceso, además también se han planteado soluciones no invasivas mediante la utilización de diferentes sensores de visión. Muchas de estas propuestas emplean algoritmos basados en la localización de patrones de persona y análisis de la escena mediante el empleo de cámaras RGB, esto supone una controversia al utilizarlos en determinados ámbitos donde la privacidad de la persona debe ser guardada. Debido a esto y a la creación de nuevas técnicas dentro de la visión artificial, ha comenzado a abrise la solución de este problema al campo de los sensores de profundidad.

Estos sensores de profundidad se basan en la obtención de las distancias de cada punto de una escena respecto a la posición del sensor. Dentro de los diferentes métodos de obtención de medidas de profundidad existentes, expuestos en 2.2.1, este trabajo se centrará en el uso de cámaras de tiempo de vuelo *Time Of Flight (TOF)* , cuyo principio de fucionamiento se expone en 2.2.1.1.

1.1 Objetivos

El objetivo de este trabajo es realizar un algoritmo capaz de contar el número de personas dentro de un entorno interior, únicamente mediante el uso de cámaras de profundidad basadas en sensores TOF situados en una posición cenital respecto al plano del suelo. La siguiente imagen 1.1 muestra el diagrama de bloques general en el cual se basa el algoritmo.

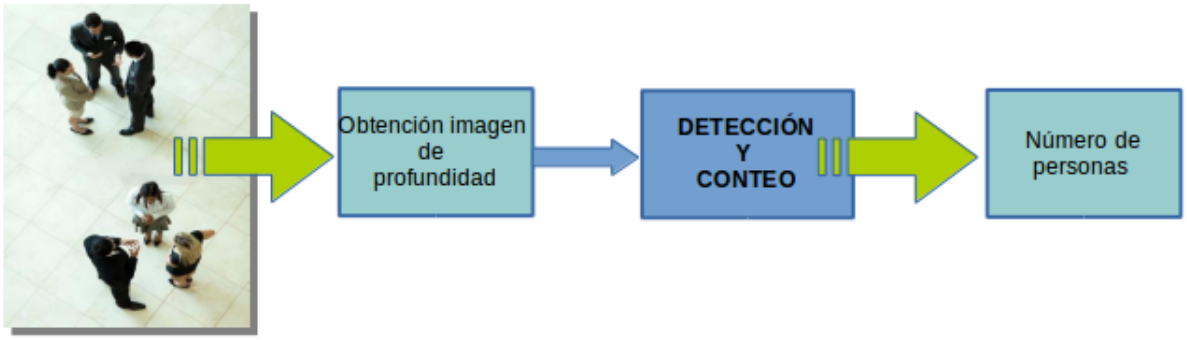


Figura 1.1: Análisis de una escena cenital para el conteo de personas en la misma.

El desarrollo del sistema se basa en la implementación de un algoritmo en lenguaje C/C++ mediante el uso de la libreria OpenCV, que detecte y cuente el número de personas presentes en la escena a partir de la información proporcionada por el sensor Kinect II (2.2.2). Este sistema consta de las siguientes etápas, que se explican con mayor detalle en el capítulo (3).

Una vez obtenida la imagen de profundidad mediante el sensor TOF colocado en el techo de la escena (3.3), se procede al análisis de la información de profundidad. Este análisis se divide en los diferentes bloques descritos a continuación.

1. Etapa de filtrado para eliminar la gran cantidad de ruido introducido por el sensor, más la aportación del ruido ambiente, desarrollado en el apartado 3.4.

2. Etapa de localización de los máximos de altura existentes en la escena, los cuales se corresponderán con todos los posibles candidatos a persona de la imagen, descrito en 3.5.

3. Etapa de obtención de las regiones de la imagen que forman parte de cada uno de los máximos entontrados, desarrollado en la sección 3.6.

4. Etapa de extracción del vector de características, explicada en 3.7, que se utilizará para definir dicho máximo y evaluar su pertenencia o no la clase persona, definida previamente a partir del aprendizaje de una base de datos (3.2).

5. Etapa de clasficiación del vector de características 3.8.

1.2 Estructura del documento

Este trabajo se dividirá en cuatro capítulos organizados de la siguiente manera:

- Fundamentos teóricos. En este capítulo se expondrá el problema de la detección y conteo de personas, analizando los diferentes trabajos previos y, con más profundidad, el trabajo tomado como punto de partida, el funcionamieto y fundamentos de las cámaras basadas en sensores de tiempo de vuelo (ToF) y del clasficador empleado.

- Detección y conteo de personas a partir de información de profundidad. En este capítulo se explicarán con mas detalle cada uno de los diferentes bloques que forman el sistema propuesto.

- Resultados experimentales. En este capítulo se expondrán los diferentes resultados obtenidos empleando secuencias grabadas previamente.
- Conclusiones y lineas futuras.

Capítulo 2

Fundamentos teóricos

El objetivo del este proyecto es la detección y el conteo de personas dentro de una escena basándose únicamente en información de profundidad proporcionada por un sensor de profundidad en posición cenital.

Dentro de este apartado se expodrá el problema comentado en mayor profundidad, así como la solución propuesta en otros trabajos anteriores que se han tomado como punto de partida para este proyecto. A continuación se detallarán también los diferentes tipos de sensores de profundidad que puede encontrarse en el mercado centrándose con más detalle en las cámaras TOF y su funcionamiento, finalmente se explicarán los fundentos teóricos del clasficador empleado en este trabajo.

2.1 Detección y conteo de personas

La detección y el conteo de personas en imágenes es una tarea de gran interés, debido a las diferentes aplicaciones, tales como vigilancia, control de accesos, análisis de flujo de personas, análisis de comportamiento o control de aforos. Este tipo de aplicaciones ha adquirido cada vez más importancia en los últimos años, dadas las necesidades de prevención y detección de situaciones potencialmente peligrosas. Se habla de conteo de personas como análisis del número de personas que se encuentran en un determinado área a evaluar.

Dado el creciente interés de la comunidad científica en este tema, existen numerosos trabajos en la literatura que tratan de realizar la detección y conteo de personas de forma robusta, y no invasiva (sin necesidad de incorporar tornos u otros elementos para el control de acceso) en diferentes escenarios. Los primeros trabajos en esta línea se basan en el uso de cámaras de color. En [3] se plantea un sistema basado en el aprendizaje de modelos de persona, por otro lado en [4] se presenta una propuesta para el conteo de personas en tiempo real a partir de la eliminación robusta del fondo y posterior segmentación de las personas. Otros trabajos están basados en la detección de caras [5] o la clasificación de puntos de interés [6].

Las propuestas anteriores presentan buenos resultados en condiciones controladas, sin embargo tienen problemas en escenarios con un alto grado de oclusiones. Para tratar de reducir las oclusiones,existen otras alternativas que ubican la cámara de forma que se tenga una vista cenital de la escena.

En los últimos años ha aumentado significativamente los sistemas de visión en los cuales la privacidad de la persona no es invadida, es decir, donde no se obtienen datos significativos para poder identificar al individuo. Este es el caso del trabajo presentado en [7], en el que los autores proponen el uso de cámaras

de baja resolución,alejadas de las personas para su monitorización sin invadir su privacidad. Sin embargo, la solución planteada únicamente puede aplicarse en entornos que permitan la ubicación de las cámaras en posiciones alejadas. Por otro lado, en los últimos años han aparecido múltiples trabajos [8], [9], [10] que proponen el uso de sensores de profundidad basados en tiempo de vuelo TOF o cámaras 2.5D [11], [2] para la detección y conteo de personas.

Los trabajos que emplean sensores de profundidad para la detección,seguimiento y conteo de personas [8], utilizan en todos los casos una cámara cenital con el objetivo de minimizar el efecto de las oclusiones, creando así una imagen más clara de la forma de la persona.

En [10] se plantea el uso de un sensor de profundidad en lugar de un par estéreo para el seguimiento de personas incluso en condiciones en las que la iluminación sea pobre. En este trabajo se emplea tanto la imagen de profundidad, como una imagen de intensidad (en escala de grises) proporcionada por el sensor utilizado, por lo que puede presentar problemas en aplicaciones en las que existan restricciones relacionadas con la preservación de la privacidad. Además, se impone la condición de que las personas entren separadas en la imagen, para evitar problemas de oclusión, por lo que no funciona bien si el número de personas en la escena es elevado o si estas entran en grupos. Por otro lado, en [9] y [1] los autores presentan un sistema que emplea una cámara ToF cenital. La propuesta se basa en la detección de máximos en la escena y posterior uso de un filtro para la separación de los posibles usuarios. Esta propuesta presenta mejores resultados que [8] cuando la densidad de personas en la imagen es elevada. Finalmente, en [10] se describe una alternativa para la detección y seguimiento de personas,así como una estimación de la pose (sentado o de pie) de las personas en un espacio inteligente. Los autores realizan la detección de personas y obtienen la pose de las mismas en función de la altura de un conjunto de puntos segmentados.

Los trabajos mencionados [9] y [10], permiten realizar la detección preservando la privacidad, sin embargo, presentan problemas de robustez en situaciones en las que puedan existir otros agentes en el entorno (objetos, animales, etc.) además de personas, ya que detectan cualquier agente presente en la escena, sea o no una persona, por lo que pueden presentar falsos positivos.

En este trabajo, se propone un sistema que permite la detección y seguimiento de múltiples personas, de forma fiable y robusta, empleando únicamente la información de profundidad proporcionada por un sensor de profundidad basado en tiempo de vuelo. La solución propuesta, funciona incluso en situaciones en las que el número de personas es elevado y éstas se encuentran próximas entre sí, además, mejora la robustez frente a trabajos anteriores al permitir discriminar entre personas y otros agentes presentes en el entorno.

Dado que el punto de partida de este trabajo se basó en el estudio de [1], a continuación se describe con mayor profundidad.

2.1.1 Detector de personas basado en el filtro wavelet del Sombrero mexicano (Laplaciana de la Gaussiana)

La solución presentada en este artículo se basa en la implementación de un sistema basado en el diagrama de bloques mostrado en la figura 2.1

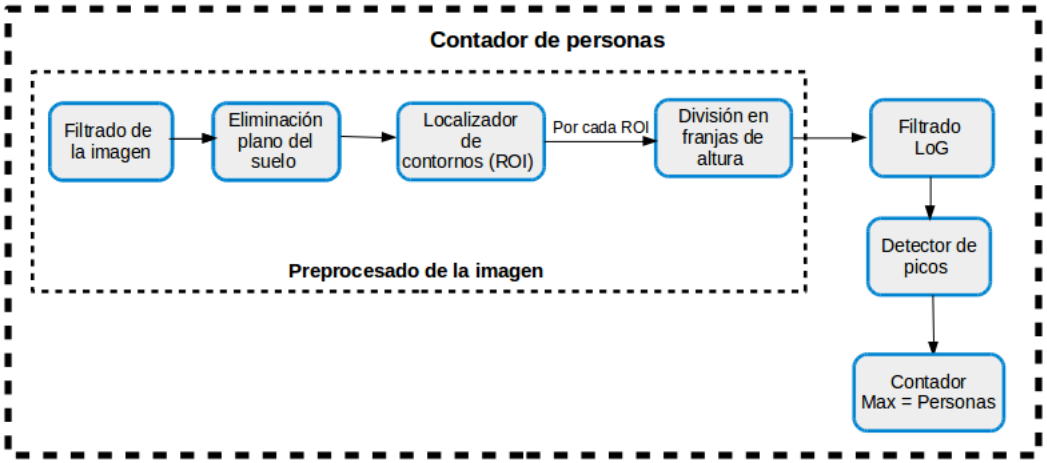


Figura 2.1: Diagrama de bloques del algoritmo desarrollado en [1]

El análisis del conteo de personas ha sido analizado como una localización e identificación de personas empleando un filtro *Laplacian of Gaussian (LoG)* , se emplea esta solución basándose en la similitud existente entre la forma de una persona vista en posición cenital y la forma del filtro (tal y como se puede apreciar en la figura LoG 2.2)

Este tipo de filtro se forma al realizar la segunda derivada de la función gaussiana, recibiendo la denominación LoG . El principal motivo de esta operación es la gran sensibilidad al ruido que presenta el operador Laplaciano por si mismo, de esta forma los bordes de la imagen quedan suavizados, detectándo únicamente como bordes aquellos en los que su conjunto de puntos formen un máximo local debido al uso de la segunda derivada.

- Función Gaussiana:

$$\psi(r) = \frac{1}{(\sqrt{2\pi(\sigma)^3})}\left(1 - \frac{r^2}{(\sigma)^2}\right) e^{\frac{-r^2}{2(\sigma)^2}} \tag{2.1}$$

- Fución LoG:

$$\psi(r) = \frac{2}{\sqrt{3\sigma}\pi^{1/4}}\left(1 - \frac{r^2}{\sigma^2}\right) exp\left(-\frac{r^2}{2\sigma^2}\right) \tag{2.2}$$

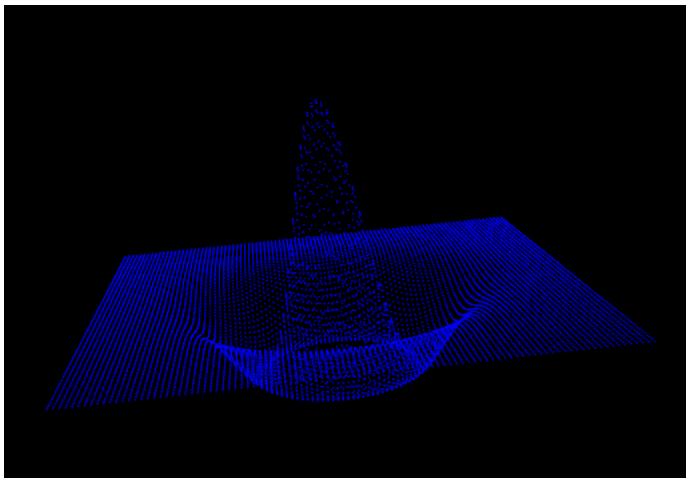


Figura 2.2: Filtro Sombrero mexicano

Pese a tratarse únicamente de un detector de bordes, lo estudios relativos al paper a analizar o emplean como un detector de personas debido a su semejanza con la parte superior de la persona visto desde una posición cenital, para ello se normaliza la altura del sombrero en función de la altura a la que esté colocada la cámara, lo que equivaldría a la altura relativa de los candidatos a persona. Una cámara situada a una distancia muy elevada tendría un campo de visión (FOV) mayor y por lo tanto los objetos aparecerían más pequeños.

2.1.1.1 Preprocesado de la imagen

Dado que la imagen a analizar se trata de una imagen de distacias, para aumentar la efectividad del filtro se crea una imagen equivalente $\hat{F}$, donde los puntos se agruparan en conjuntos de distancias con diferentes valores, dividiendo la imagen en franjas en función de la altura a la que correspondan. Los pasos a realizar para obtener dicha imagen equivalente serían:

1. Eliminación del plano del suelo g(x,y,z). (Figura 2.3c)

$$F = f(x, y, z) - g(x, y, z) \tag{2.3}$$

   Esta operación consiste en un filtrado por alturas, elimando alturas inferiores a 60cm. Este umbral fue seleccionado debido a que serán considerados candidatos a persona aquellos objetos que se encuentren en la imagen cuya altura sea superior a 60cm. Por lo tanto, si el escenario a analizar se encontrase vacio, el resultado de esta operación sería

$$F = 0 \tag{2.4}$$

2. Para cada objeto o contorno existente en la imagen (empleo del algoritmo de búsqueda de contornos analizado anteoriormente) se definirá una *Region Of Interest (ROI)* . El tamaño de dicha ROI se definirá en función del radio del contorno encontrado. Por lo tanto, si el contorno localizado abarca varios candidatos a persona, dichos elementos se analizarán de manera conjunta. (Figuras 2.3d, 2.3e)

3. División de imagenes formadas por cada ROI en s franjas, siendo s un valor variable. En este caso emplearemos un valor s = 8 (valor indicado en [1]). Para crear dichas secciones será necesario localizar el valor máximo y mínimo de cada ROI y de esta forma calcularemos el "salto" o "step" $S_{i,step}$ mediante el cual se dividirá la imagen.

$$S_{i,step} = \frac{F(n_t, m_{max}) - F(n_t, m_{min})}{s} \tag{2.5}$$

4. Una vez calculado el "step"se asigará a cada pixel de la imagen un nuevo valor:

$$S_{i,l} = \lfloor \frac{d - F(n_t, m_{min})}{S_{i,step}} \rfloor + 1 \tag{2.6}$$

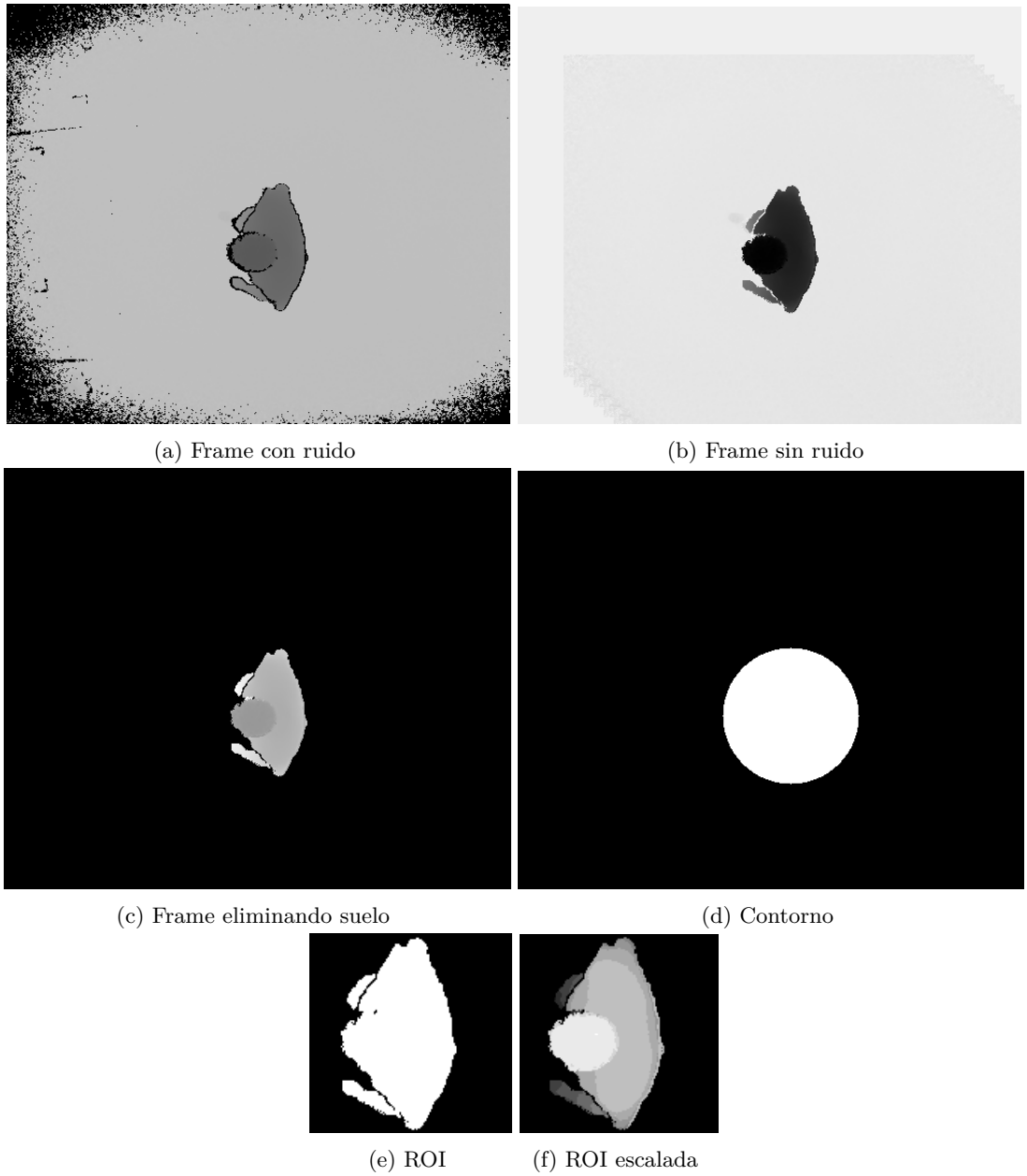
   De esta forma se invertirá la intensidad en la image resultante, tomando valores más próximos a 1 (255) los elementos con mayor altura y más proximos a 0, los elementos más bajos. (Figura 2.3f)

5. Creación de una imagen independiente para cada ROI obtenida y preprocesada, donde todos los

píxeles no pertenecientes a la ROI toman valor 0.

$$\hat{F}(n_t, n_x, ny) = \begin{cases} S_{i,l}, & \text{if } \Re \in \Re_i \\ 0, & \text{else} \end{cases} \tag{2.7}$$

A continuación se muestra el preprocesado descrito sobre una imagen de profundidad donde aparece una única persona en el centro de la escena:



(a) Frame con ruido (b) Frame sin ruido

(c) Frame eliminando suelo (d) Contorno

(e) ROI (f) ROI escalada

Figura 2.3: Preprocesado de una persona sola en la escena, algoritmo perteneciente a [1]

2.1.1.2 Filtrado de la imagen

La elección de los parámetros del filtro se basan en pruebas realizadas con una cámara TOF Kinect II a una altura de 340cm del suelo, debido a esto el tamaño que abarca una persona vista desde una posición cenital puede estimarse en un rango de 80 -60 píxeles. Dado que el empleo de este filtro es la similitud

con la figura humana los valores más óptimos serían los siguientes:

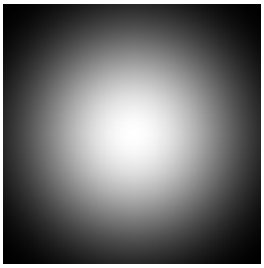


Figura 2.4: Filtro LoG empleado como detector de personas

$$\sigma_x = 40px \ \sigma_y = 40px \text{ Tamaño Kernel} = 100x100px \tag{2.8}$$

Si se quiere asemejar el sombrero mexicano a la forma de un ser humano, este no debería tener simetria axial, sino mantener una forma elipsoidal, sin embargo, ya que no se conoce la pose de la persona, resulta más sencillo utilizar un filtro completamente circular.

Una vez termiado el preprocesado de la imagen, es decir, el rescalado del candidato a persona si en el frame analizado apareciera, se procede a convolucionar la imagen resultante con el filtro adapado a la altura de la cámara. Esta correlación se realiza respecto a las variables n_x y n_y (coordenadas del plano x-y de la imagen de profundidad) de la imagen para un mismo frame.

$$C(n_t, n_x, n_y) = \frac{1}{N_x N_y} \sum_{k_x=0}^{N_x-1} \sum_{k_y=0}^{N_y-1} \hat{F}(n_t, k_x, k_y)\psi(n_x + k_x, n_y + k_y) \tag{2.9}$$

En el trabajo [1] se analiza a continuación la imagen C obtenida, afirmando que los picos de intensidad se corresponden con altos valores de semejanza entre el objeto a analizar y el filtro empleado, siendo el valor de dichos picos independientes de la altura del objeto a analizar gracias a la segmentación realizada previamente. Por lo tanto, y ya que una misma ROI tiene la posibilidad de contener varios candidatos a persona, empleando un detector de picos con un umbral para agilizar el proceso se detectarían todas las personas de la imagen.

2.2 Cámaras de profundidad

En este apartado se expondrán las diferentes técnicas de medición de distancias analizadas en [12]. La visión de los seres humanos (también de los animales más desarrollados) permite obtener una medida de profundidad, la cual se pierde al plasmar imágenes con cámaras RGB. Los sensores de profundidad, entendiendo como profundidad la distancia entre el fondo y un punto tomado como referencia, en este caso el sensor, permiten obtener una matriz de la escena captada cuyos valores se corresponderán con cada una de las medidas de profundidad obtenidas.

En la actualidad existen diferentes métodos utilizados para la obtención de medidas de profundidad en distintos escenarios, a continuación se expone una breve exposición de los mismos centrándose más detalladamente en aquellos basados en tiempo de vuelo.

2.2.1 Métodos para la obtención de medidas de profundidad

A continuación se describirán los principales métodos utilizados para la obtención de medidas de profundidad.

1. **Interferometría**
   La interferometría consiste en el análisis de patrones de intensidad. Se basa en la superposición de dos frentes de onda monocromáticas con diferente amplitud (α_1 y α_2) y fase (ϕ_1 y ϕ_2), pero en la misma frecuencia *f*.
   La onda resultante debido a la reflexión de las emitidas en un determinado punto, tendrá una amplitud (α_3) diferente en función del resultado de las ondas enviadas, constructivas o destructivas.

   El resultado de esta medida es interpretada por el interferómetro, instrumento que emplea las interferencias de ondas monocromáticas para medir longitudes de ondas (λ). Contando el número de máximos y mínimos producidos en la interferencia se puede estimar la distancia entre el elemento emisor y la superficie con una precisión estimada de λ.

2. **Triangulación**
   Los métodos de triangulación obtienen información de profundidad utilizando el triángulo que se forma entre el punto sobre la superficie en cuestión y el eje óptico. Existen dos tipos de métodos de triangulación en función del número de receptores.

   (a) Triangulación pasiva: Empleo de dos receptores pasivos. Es el método de medida de profundidad más conocido, se basa en el mismo principio que la visión humana (visión en estéreo). En este método es necesario el uso de dos cámaras situadas a una distancia *b* una de la otra, para que sea posible el cálculo de la medida de profundidad de un punto, dicho punto debe encontrarse en el campo de visión de ambas cámaras. El principal problema de este método es la búsqueda de la correspondencia de puntos en ambas imágenes.

   (b) Triangulación activa: Empleo de un receptor pasivo y un emisor de luz, separados entre sí de la misma forma que en la triangulación pasiva. En este caso la formación del triángulo utiliza el ángulo de emisión del rayo de luz y la posición del rayo relejado en la imágen formada en el receptor pasivo.

3. **Cámaras de tiempo de vuelo TOF**
   El funcionamiento de esta tecnología se basa en la utilización de un emisor y un receptor de luz. El emisor envía una señal modulada, parte de la cual, al rebotar en los objetos que se encuentren en la escena, regresará al receptor. Al tratarse de una señal periódica, modulada a una frecuencia conocida, la diferencia de fase entre la señal enviada y la recibida será un indicador de la distancia a la que se encuentre el objeto.

   (a) Emisor: Normalmente, para la iluminación de la escena se emplea un laser de estado sólido o un LED operando a frecuencias cercanas al infrarojo (850nm, no visible a partir del ojo humano). En función del tipo de señal emitida por el emisor aparecen dos tipos de clasificaciones:

      i. Modulación pulsada (2.5)

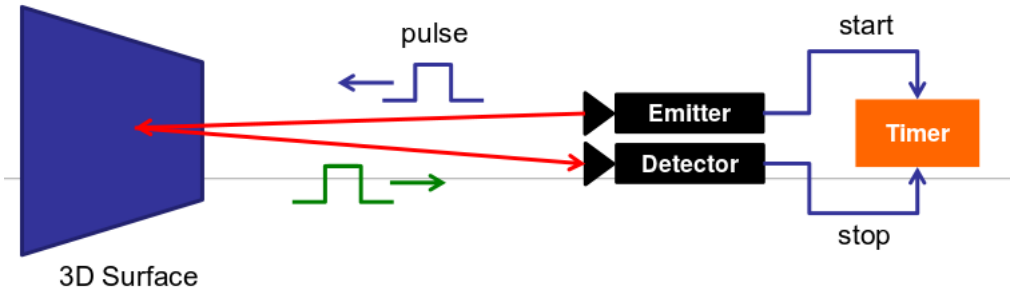


Figura 2.5: Funcionamiento TOF mediante modulación pulsada

Este tipo de modulación, aunque permite una medida directa de la distancia a partir del pulso y se ve menos influenciada por la iluminación del fondo, posee inconvenientes como pueden ser la necesitad de una alta precisión en la medición del tiempo, la importante incursión de efectos de *scattering* que pueden variar la medida del pulso de luz recibido y la dificultad de generar pulsos a alta frecuencia durante periodos prolongados de tiempo.

ii. Modulación de onda continua (CWM) (2.6)

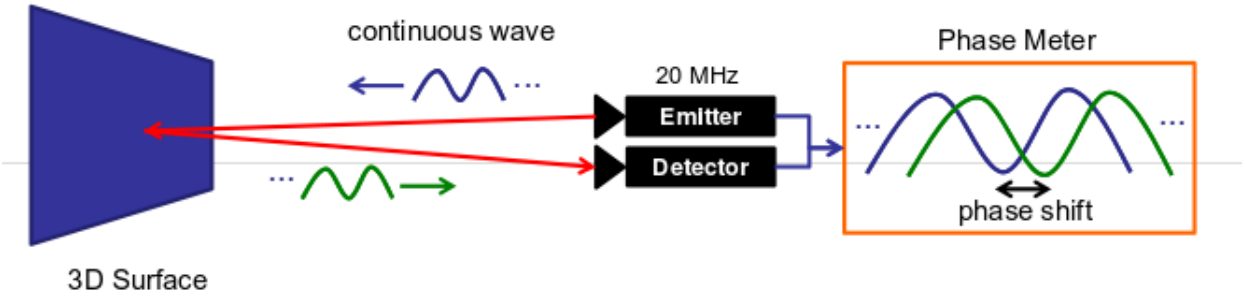


Figura 2.6: Funcionamiento TOF mediante modulación CWM

Este tipo de modulación basada en la iluminación continua del espacio, midiéndose en el receptor la diferencia de fase de la onda recibida respecto de la emitida, el cual será el indicativo de la distancia a la que se encontrará el objeto. Este procedimiento es más utilizado en este campo gracias a su facilidad para una mayor aplicación de diferentes técnicas de modulación.

(b) Receptor: La luz recibida por el receptor posee una componente ambiente y una reflejada por lo tanto el factor señal disminuirá frente al ruido. Para que sea posible la medida de la diferecia de fase entre el rayo enviado y su reflejado es necesario que se trate de una fuente de iluminación de luz pulsada o modulada, siendo la modulación más usada una modulación de onda cuadrada debido a su comodidad de uso en circuitos digitales.

Las cámaras TOF actuales se basan en un sistema formado por 4 desplazamientos de fase (separados 90º)por cada píxel, es decir, se empleará una ráfaga de luz durante un tiempo T, donde se distinguirán los cuatro desplazamientos de fase

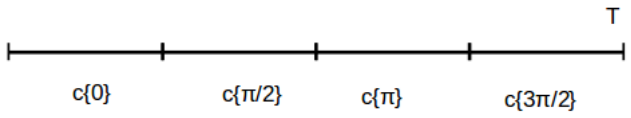


Figura 2.7: Desplazamientos de fase para la señal emitida por cada píxel

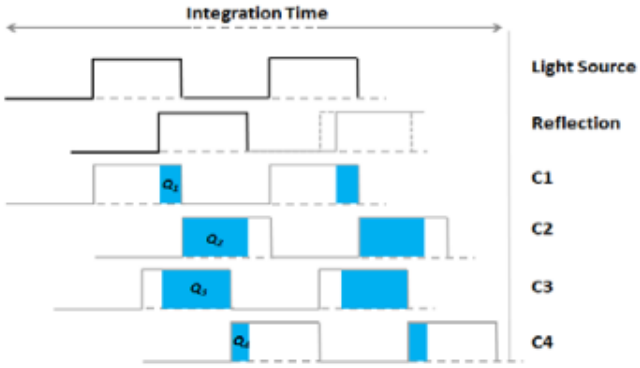


Figura 2.8: Proceso de emisión y recepción de una señal en cámaras TOF

A partir de la figura 2.8 podemos ver el funcionamiento de la medida de la diferencia de fase del rayo emitido y relfejado, β, y la distancia d, las cuales se calcularán de la siguiente forma:

$$\beta = \arctan(\frac{Q_C - Q_D}{Q_A - QB}) \tag{2.10}$$

$$d = \frac{c}{4\pi f_{mod}}\beta \tag{2.11}$$

Siendo $c = 3 * 10_8$ m/s (Velocidad de la luz en el vacio), f_{mod} la frecuencia de modulación y Q_A, Q_B, Q_C, Q_D las cargas acumuladas durante las diferentes muestras en cada uno de los receptores Mosfet.

La máxima distancia fiable (d_{max})posible de calcular viene determinada por la frecuencia de modulación del emisor. Dado que la medición de fase se situa en torno a 2π, una vez superada dicha distancia el valor obtenido en d se repetirá periodicamente (d + d_{max} = d).

$$d_{max} = \frac{c}{2f_{mod}} \tag{2.12}$$

Siguiendo la ecuación 2.12 si se quisiera ampliar la distancia máxima de la cámara bastaría con disminuir la frecuencia de modulación de la onda.

A continación se explicará con más detalle el funcionamiento de los sensores basados en modulación de onda continua (CWM) ya que será la tecnología empleada en el sensor utilizado en este trabajo (Kinect II).

2.2.1.1 Principio de funcionamiento cámaras ToF basados en modulación de onda continua (CMW)

El funcionamiento de las cámaras *Continuous Wave Camera (CWM)* de tiempo de vuelo se basan en la existencia de cuatro elementos básicos: (i) Fuente incoherente de luz infrarroja, (ii) modulador de

frecuencias, (iii) *Charged Coupling Devices (CCD)* / *Complimentary Metal Oxide Semiconductor (CMOS)* array de píxeles correlados, (iv) detector de desplazamientos de fase. La fuente de luz no debe ser necesariamente una fuente coherente, debido a que la medición no se ve influida por las interferencias, en lugar de eso se emplea una amplitud modulada a frecuencia f_{mod}, de esta forma se evita que sea necesario el uso de una fuente de luz específica. El pricipio de fucionamiento de estas cámaras se basa en el envío de una onda de luz modulada de la forma:

$$e_t = A_o \sin(2\pi f_{mod} t) \tag{2.13}$$

siendo f_{mod} la frecuencia de modulación y A_o la amplitud de la potencia emitida.

La señal recibida por el array de receptores CCD /CMOS se puede representar de la siguiente forma:

$$r_t = o_t + A_o K \sin(2\pi f_{mod} t - \beta) \tag{2.14}$$

A partir de la señal definida se pueden diferenciar las diferentes características añadidas a la señal emitida:

1. $o(t)$ offset añadido a la señal recibida debido a la iluminación ambiente, normalmente este offset, ya que la iluminación de fondo suele producirse a frecuencias bastante más bajas que la frecuencia de modulación, suele considerarse constante de la forma $o(t) = O_o$.
2. K es la atenuación de la potencia de la onda recibida debido tanto a efectos de propagación y reflexión de la onda, como a efectos de la optica (lentes, filtros...)
3. β desplazamiento de fase que contiene la información de distancia. A partir de este, como se muestra anteriormente en 2.11, desplazamiento de fase se obtendría la medida de la distancia:

$$\beta = 2\pi f_{mod} t_d = 2\pi f_{mod} \frac{2D}{c}. \tag{2.15}$$

siendo t_d el tiempo empleado por la onda en ir y volver al receptor.

Debido a la difucultad para medir directamente el desfase entre ambas ondas, se emplean sistemas de correlación en cada sensor del receptor. Dentro de cada pixel se realiza una correlación entre la onda recibida y una señal de referencia $u_m(t)$(generalmente cuadrada) en fase con la señal original emitida $e(t)$ 2.13, durante un tiempo de integración T_{int}. Aunque, como puede verse, con las cámaras TOF no se obtiene una medida directa de la profundidad de la escena, si se obtiene un factor de correlación proporcional a ella.

Se puede definir entonces la función de correlación ρ entre ambas ondas como:

$$\rho = K_{int} \frac{1}{T} \int_0^T u_m(t+\tau) r(t) dt = K_{int} \left(\frac{A_o}{\pi} \cos(\beta + \tau) + \frac{O_o}{2}\right) \tag{2.16}$$

donde las variables A_o, β y O_o son valores desconocidos. La variable K_{int} es un acumulador del número de periodos T sobre los que se ha realizado la correlación durante el tiempo de integración T_{int}.

De esta forma la diferencia de fase calculada en 2.10 puede calcularse sustituyendo los valores de las cargas (Q_A, Q_B, Q_C, Q_D) por los valores de las funciones de correlación en los diferentes instantes $(\rho_0, \rho_{\pi/2}, \rho_\pi, \rho_{3\pi/4})$.

Cada uno de los receptores CCD /CMOS formados por transistores MOSFET transforman la señal recibida en cargas $(C_0 -> Q_A / C_{\pi/2} -> Q_B)$ en cada una de las bandas (como explica en 2.7 cada

una de esas cargas se correspode con un tiempo de emisión de la onda), de esta forma se compara la carga acumulada en las diferentes bandas obteniendose una medida diferencial ($Q_A - Q_B$). esta medida diferencial proporciona un valor al pixel que depende del nivel de luz recibido y de su tiempo de llegada respecto del reloj propio del pixel.

2.2.1.2 Fuentes de errores en sensores TOF

Los valores de profundidad obtenidos pueden verse afectados por diversas fuentes de ruido, restándo precisión y fiabilidad a la medida realizada. La diferentes fuentes de ruido a las cuales se ven sometidos los sensores, generan diversos errores que pueden clasificarse en dos grandes grupos [12]:

1. **Errores sistemáticos:** Este tipo de errores suponen un gran impacto en la precisión y la confianza de las medidas de profundidad tomadas. Normalmente este tipo de errores aparecen en el momento de transformar la luz recibida a señal. Dentro de los errores sitemáticos que aparecen en las cámaras TOF podemos destacar (a) "Wiggling error"(debido a señales sinusoidales no perfectas), (b) Ruidos de patrón fijo en determinados píxeles, (c) diferentes amplitudes debido a iluminación y reflectividad no constantes, (d) "*Shotnoise*" ruido de disparo que produce variaciones en el número de electrones de los MOSFET, (e) variaciones en la temperatura que producen desajustes en el material del semiconductor. Para resolver estos errores la mayoría de las cámaras comerciales utilizan compensaciones via hardware.

2. **Errores no sistemáticos:** Dentro de este campo destacan. (a) "*Flying píxeles*" en las discontinuidades de los objetos de la escena, (b) ambigüedad en la distacia cuando se supera la máxima distancia fiable d_{max} 2.12, (c) ambigüedad debida al movimiento, (d) interferencias debido a "Multicamino", diferencia de caminos en la señal recibida para un mismo pixel.

2.2.2 Características Kinect II

El sensor Kinect II proporciona tres imágenes diferentes, imagen de amplitud, imagen de escala de grises e imágen de profundidad, mostradas en la figura 2.9

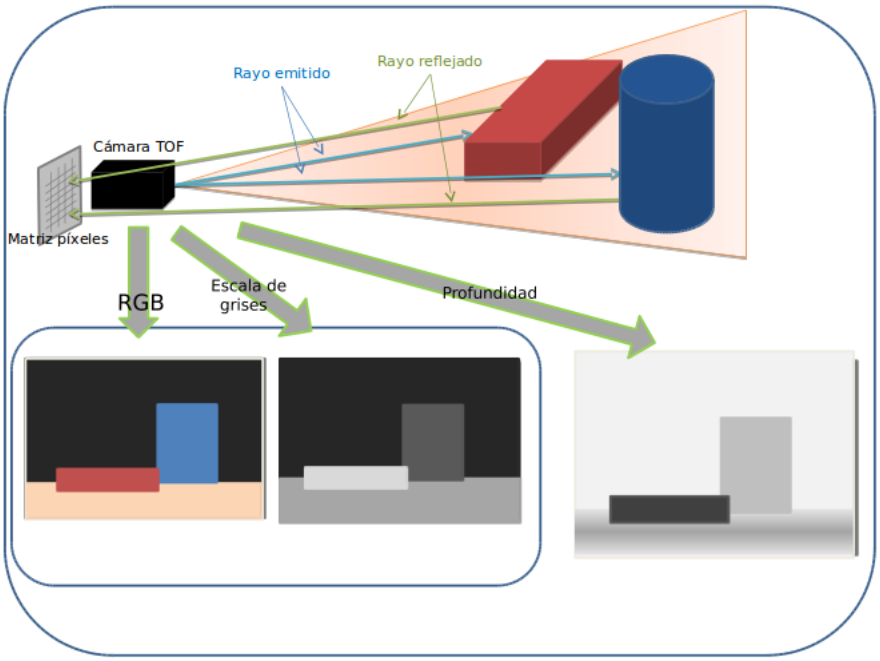


Figura 2.9: Tipos de imágenes obtenidas a partir del sensor Kinect II

Kinect II incluye una cámara de profundidad basada en modulación de onda continua 3(a)ii, siguiendo el principio de funcionamiento descrito en 2.2.1.1. La elección de esta cámara se debe a que proporciona una resolución bastate mayor a otras cámaras del mercado con un precio inferior, si bien el ruido que contiene la imagen obtenida es mucho mayor y requiere fases de filtrado mas bruscas, la relación calidad precio favorece al coste del trabajo.

Las principales características del sensor TOF basado en sensores CMOS que ofrece Kinect II son los siguientes [2]:

1. Ángulo de visión de 70 grados en horizontal por 60 grados verticalmente.
2. Apertural focal F# <1.1
3. Resolución de profundidad dentro del 1 % de la distancia medida.
4. Rango de medidas válidas 0.8 <d <4.2
5. Independiente de la iluminación de fondo.
6. Tiempo máximo de exposición de 14 ms.
7. Transferecia de datos a través de USB 3.0 con una latencia menor de 20 ms.
8. Error en la medida menor de 2 % dentro del rango de operaciones.

La arquitectura del sistema TOF utilizada por Kinect II se muestra en la siguiente imagen:

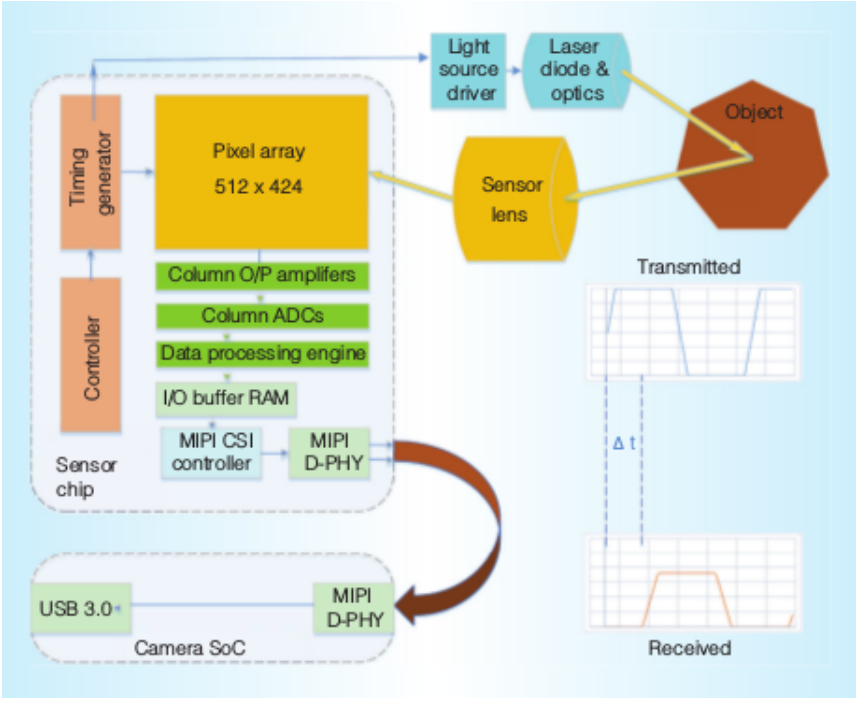


Figura 2.10: Arquitectura interna sensor TOF Kinect II
Imagen obtenida de [2]

El sistema incluye una cámara basada en *System on Chip (SoC)* , sistema de iluminación y los sensores ópticos para la generación de la imagen. Para la modulación de la onda emitida por la fuente de luz y de la onda recibida se utiliza una señal de onda cuadrada generada por *"timing generator"*.

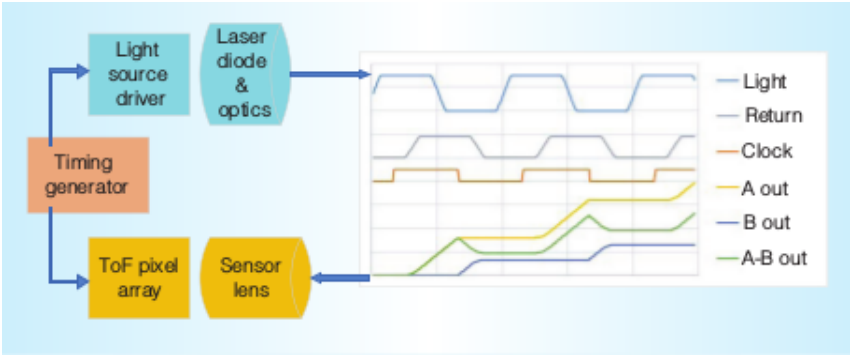


Figura 2.11: Cronograma de señales emitidas y recibidad por cada píxel en sensor TOF de Kinect II
Imagen obtenida de [2]

En 2.11 se muestra el cronograma de formas de onda que utiliza el sensor para obtener la medida de profundidad en un pixel, donde *"Light"* y *"Return"* representan la envolvente de la ondas enviadas y recibidas, *"Clock"* reloj generado para un pixel, *"A out"* y *"B out"* el voltaje recibido en cada uno de los MOSFET de cada pixel y *"A-B out"* la medida diferencial del pixel.

2.3 Clasificadores

Un clasficador es una herramienta utilizada para asigar un elemento no etiquetado a una de las clases predefinidas, este algoritmo permitirá entonces ordenar por clases cada uno de los elementos entrantes a

partir de cierta información característica de ellos. Los datos necesarios para el clasificador se estructuran dentro del llamado conjunto de entrenamiento, esta agrupación debe estar formada por datos conocidos que serán empleados para el aprendizaje del clasificador. Los clasificadores pueden dividirse en dos grandes grupos en función del tipo de aprendizaje al que sean sometidos.

- Aprendizaje supervisado: este tipo de aprendizaje superficial se elabora a partir de un entrenamiento realizado con datos etiquetados previamente, por lo tanto es el usuario quien debe indicar al algoritmo a que clase pertenecen los datos introducidos.
- Aprendizaje no supervisado: este tipo de algoritmos no disponen de un conjuto etiquetado de datos de entrenamiento, así pues se emplean tecnicas de agupamiento o "*clustering*" con la finalidad de agrupar los objetos entrantes en conjuntos semejantes que puedan constituir clases diferenciadas.

Dentro de los clasificadores empleados en temas de visión cabe destacar las Máquinas de soporte virtual (SVM) [13] o *"Random forest"* [14] , en el caso de este trabajo, se empleará el clasficador basado en en análisis de las componentes principales (PCA) explicado a continuación.

2.3.1 Análisis de las componentes principales PCA

"Principal Components Analysis", PCA es una técnica utilizada para reducir la dimensionalidad de un conjunto de datos, hallando las causas de la variabilidad de un conjunto de datos y ordenándolas por importancia. PCA busca la proyección según la cual los datos pueden ser representados en términos de mínimos cuadrados, de esta forma la búsqueda de patrones en los datos se vuelve mucho más sencilla. Una de las funcionalidades principales de PCA es su buen funcionamiento como clasficador, gracias a la extracción de las características pricipales y mediante una etapa de aprendizaje supervisado, pueden establecerse diferentes clases determinadas por las características principales del conjuto de datos. A continuación se expone con mas detalle el funcionamiento y desarrollo matemático del mismo.

2.3.1.1 Funcionamiento PCA

El funcionamiento de PCA se basa en una transformación lineal de los datos para los cuales se forma un nuevo sistema de coordenadas. Para formar dicho sistema de coordenadas es necesario organizar el conjunto datos en vectores de n elementos, y formar la matriz de covarianza de los mismos, de la cual se extraen las componenetes principales, autovectores con autovalores unitarios de la matriz de covarianza. Una vez obtenidos se puede establecer una comparativa de los pesos de cada autovector pudiendo descartar los autovectores con menor autovalor, ya que serán los menos significativos para describir el conjunto de datos, de esta forma se consigue una compresión del conjunto de datos basándose únicamente en los componentes más influyentes.

Esto es, si originalmente se tuvieran n dimesiones, se obtendrían n autovectores y autovalores, de los cuales se seleccionaran m autovectores, el conjunto de datos podría reducirse a m dimensiones. Descartando los autovectores con autovalores más bajos se produce una pérdida de información, pero puede considerarse información poco relevante a la hora de clasificar.

La forma de clasficar se basa en la transformación de cada nuevo vector de datos al espacio transformado, el cual será el indetificador de una determinada clase, y la recuperación del mismo, si la diferencia entre el vector inicial y el vector recuperado se encuentra dentro de los límites establecidos puede considerarse que el nuevo vector pertenece a la clase.

2.3.1.2 Desarrollo matemático

Teniendo una agrupación de datos D formado por un conjunto de N vectores X_p^n (donde n=1 ..., N), cada uno de ellos formados por p componentes (p=1, ... P), se obtiene una matriz T, formada por cada uno de los vectores colocados en columnas, restándole el vector media μ_p del conjunto de datos a cada uno de los componentes se obtendría la matriz R.

$$R = T - \mu = \begin{bmatrix} X_1^1 - \mu_1 & X_2^1 - \mu_2 & ... & X_P^1 - \mu_P \\ X_1^2 - \mu_1 & X_2^2 - \mu_2 & ... & X_P^2 - \mu_P \\ ... & ... & ... & ... \\ X_1^N - \mu_1 & X_2^N - \mu_2 & ... & X_P^N - \mu_P \end{bmatrix} \tag{2.17}$$

A partir de la matriz R se obtendrá la matriz de covarianza C del conjunto de datos (C R^{PxP}).

$$cov(X_1, X_2) = \frac{1}{N} \sum_{k=1}^{N} (X_k^1 - \mu_k) * (X_k^2 - \mu_k) \tag{2.18}$$

$$C = \begin{bmatrix} cov(X_1, X_1) & cov(X_1, X_2) & ... & cov(X_1, X_P) \\ cov(X_2, X_1) & cov(X_2, X_2) & ... & cov(X_2, X_P) \\ ... & ... & ... & ... \\ cov(X_P, X_1) & cov(X_P, X_2) & ... & cov(X_P, X_P) \end{bmatrix} \tag{2.19}$$

Una vez obtenida la matriz C se obtienen los autovectores y autovalores (*eigenvectors - eigenvalues*)propios de la matriz, seleccionando los más significantes se obtiene la matriz de transformación U al nuevo sistema de coordenadas.

$$U = (eig_1, eig_2, eig_3, \dots eig_m) \tag{2.20}$$

La clasificación de un vector nuevo dentro o no de la clase definida por U se realiza siguiendo el siguiente método:

$$dataTransf = U^T * (newData - \mu) \tag{2.21}$$

$$dataRecup = U * dataTransf + \mu \tag{2.22}$$

Una vez obtenido el vector recuperado *dataRecup* 2.22 se comparará con el vector inicial *newData* por el método más eficiente (distancia euclídea, distancia de mahalanovis, ...). Será necesario establecer un umbral con el cual se determinará si el vector pertenece, o no, a la clase bajo estudio.

Capítulo 3

Detección y conteo de personas a partir de información de profundidad

3.1 Descripción del algoritmo

La solución propuesta en este trabajo, para el conteo de personas a partir de la información de una cámara ToF cenital, consta de dos partes diferenciadas, tal como se puede observar en el diagrama de bloques de la figura 3.1.

1. Proceso Off-line: En esta etapa se realizará un proceso de obtención de modelos para cada una de las clases implementadas. Para ello se utilizarán un conjunto de imágenes de profundidad, adquiridas por una cámara Kinect II, y sobre las cuales se han etiquetado las diferentes personas. Para la creación de esta base de datos es importante que las imágenes de profundidad contengan información de diferentes personas con diferentes características como puede ser altura, posición, color de cabello, peinado, etc.

2. Proceso On-line: Este proceso se divide en cinco bloques: captura de la imagen de alturas a través de la cámara TOF, filtrado de la imagen para eliminar ruido, detector de los máximos en la escena (los posibles candidatos a persona), extracción del vector de características de la región de interés (ROI) entorno a cada máximo, clasificador del vector de características en función del aprendizaje Off-line y contador del número de personas. Cada una de estas etapas de describe con mayor detalle a continuación.

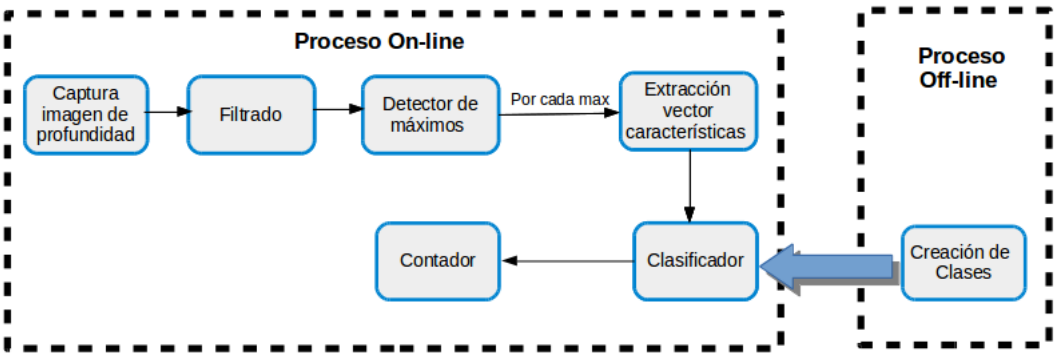


Figura 3.1: Diagrama de bloques perteneciente al algoritmo de conteo de personas

3.2 Base de datos de imágenes de profundidad

Para comprobar y analizar la eficiencia del algoritmo implementado es necesaria la utilización de una base de datos de imágenes de profundidad. Actualmente no existe ninguna que se adapte a las necesidades del proyecto, por lo tanto se ha grabado y etiquetado una base de datos de imágenes de profundidad que cumple las siguietes características:

1. Imágenes de profundidad obtenidas mediante cámaras TOF .
2. Grabaciones en interior.
3. Visión cenital de la escena.
4. Secuencias con personas de diferentes características.
5. Secuencias con personas aisladas, repartidas por la escena y en grupo.

Esta base de datos se ha realizado empleando la cámara TOF de Kinect II situada en posición cenital a una altura h= 3.4m. La finalidad de esta base de datos es tener un banco de pruebas con el cual se pueda contrastar el algoritmo, para ello se han realizado grabaciones de secuencias con una o varias personas repartidas aleatoriamente por la escena y agrupadas. El etiquedado de dichas secuencias se ha realizado indicando la posición de las elipses que forman los hombros y la cabeza de una persona cuando es vista desde una posición cenital.

Las secuencias grabadas se emplearán tanto para generar un conjunto de frames de entrenamiento (extracción del vector de características en frames con personas aisladas para generar las diferentes clases para el clasificador), así como el conjunto de frames de prueba donde se analizará la efectividad del algoritmo tanto con individuos asilados como en grupo, siendo esta última la situación más crítica debido a la dificultad de analizar las regiones de la imagen pertenecientes a cada persona.

3.3 Captura de la matriz de alturas

La cámara utilizada en este sistema (Kinect II) proporciona una imagen RGB, imagen en escala de grises y una imagen con la información relativa a la distacia. Con el fin de preservar la privacidad de las personas, las imágenes de RGB y escala de grises serán descartadas utilizando únicamente la información de la cámara TOF. La ubicación de la cámara en el entorno se muestra en la figura 3.2.

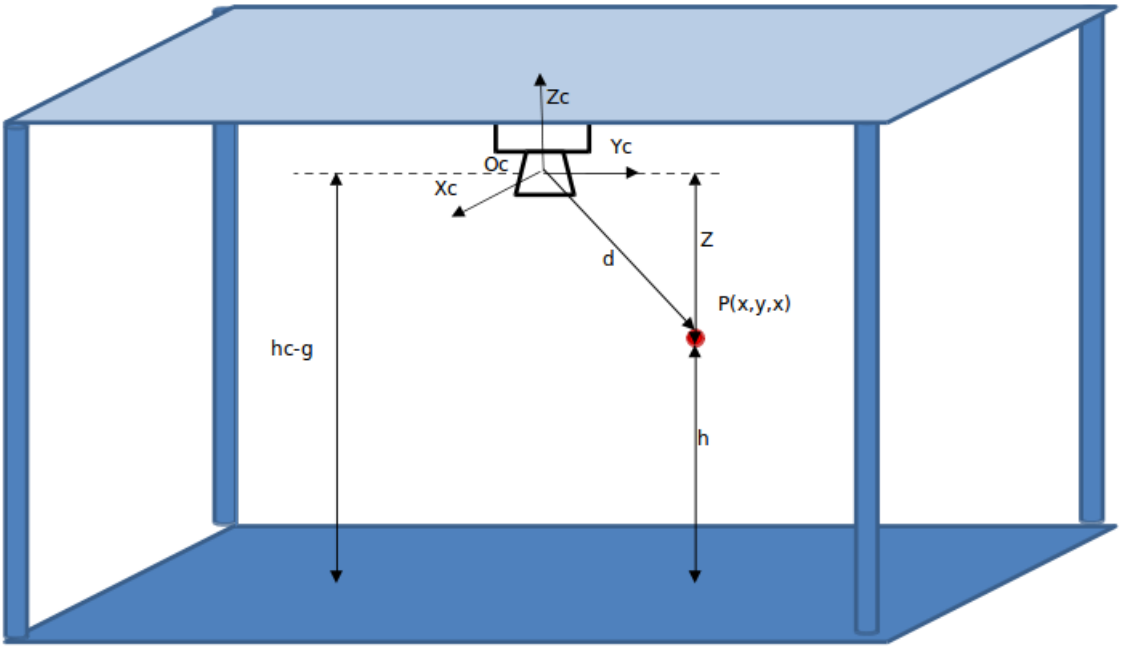


Figura 3.2: Ubicación del sensor Kinect II dentro de la escena

La cámara se sitúa perpendicular al plano del suelo, siendo O_c el eje óptico de la misma, y X_c, Y_c, Z_c los ejes de coordenadas del sistema de referencia de la cámara. El punto de la escena del cual se obtiene la medida se representa con el punto P(x,y,z), siendo d la distancia hasta O_c y Z la distancia al plano de la cámara formado por $X_c - Y_c$. Considerando que el plano del suelo y el plano $X_c - Y_c$ son paralelos, se podrá obtener la altura real del punto de la forma $h = h_g - Z$.

Aunque las cámaras TOF proporcionan para cada punto P, la distancia $d = \sqrt{x^2 + y^2 + z^2}$ y las coordenadas 3D respecto al sistema de coordenadas (X_c, Yc, Zc), en este trabajo únicamente se utilizará la coordenada Z del punto para la identificación de los elementos (empleando las coordenadas X e Y para su ubicación dentro de la escena), esto es así debido a que esta medida proporciona una información de altura independiente de la ubicación (x,y) del elemento en la escena.

Para cada imagen obtenida se generará una matriz de alturas que denomiaremos H_{mea}, suponiendo una cámara con una resolución M x N:

$$\left[H_{mea}\right] = \begin{bmatrix} Z_{1,1}^{mea} & Z_{1,2}^{mea} & \dots & Z_{1,N}^{mea} \\ Z_{2,1}^{mea} & Z_{2,2}^{mea} & \dots & Z_{2,N}^{mea} \\ \dots & \dots & \dots & \dots \\ Z_{M,1}^{mea} & Z_{M,2}^{mea} & \dots & Z_{M,N}^{mea} \end{bmatrix} \tag{3.1}$$

Donde $Z_{i,j}^{mea}$ representa la altura Z media en el pixel i,j de la cámara TOF , respecto al sistema de coordenadas (X_c, Yc, Zc), siendo $H_{mea} \in R^{MxN}$.

3.4 Filtrado de la imagen

Uno de los problemas principales de este tipo de tecnología es el elevado nivel de ruido que presentan las componentes de la matriz H_{mea}. Debido a los errores descritos en 2.2.1.2, aparecen medidas $Z_{i,j}^{mea}$ no válidas dentro de H_{mea}. Estos errores son identificados por el fabricante e indican esta circunstacia

dotando al pixel de un valor significativo (de este modo en Kinect II se asigna $Z_{i,j}^{mea} = 0mm$). A los valores no válidos indentificados por la cámara han de sumarse los valores que cada aplicación considere erróneos (en nuestro caso se considerarán valores no válidos los que superen la altura máxima de una persona, $h_{pmax} = 220cm$).

Con el objetivo de eliminar estos valores erróneos que puedan provocar ambigüedades en el momento de la clasificación se utilizan dos métodos de filtrado:

1. **Estimación de los $Z_{i,j}^{mea}$ no válidos**: En nuestro caso consideraremos medidas no válidas el conjunto de medidas indicadas por la cámara como nulas $\theta[Z_{i,j}^{null-camera}]$ y el conjunto de medidas que no puedan ser valores válidos en la altura de una persona $\theta[Z_{i,j}^{null-h_{max}}]$, considerando el siguiente criterio:

$$if \quad h_g - Z_{i,j}^{mea} > h_{pmax} \Rightarrow Z_{i,j}^{mea} \in \theta[Z_{i,j}^{null-h_{max}}] \tag{3.2}$$

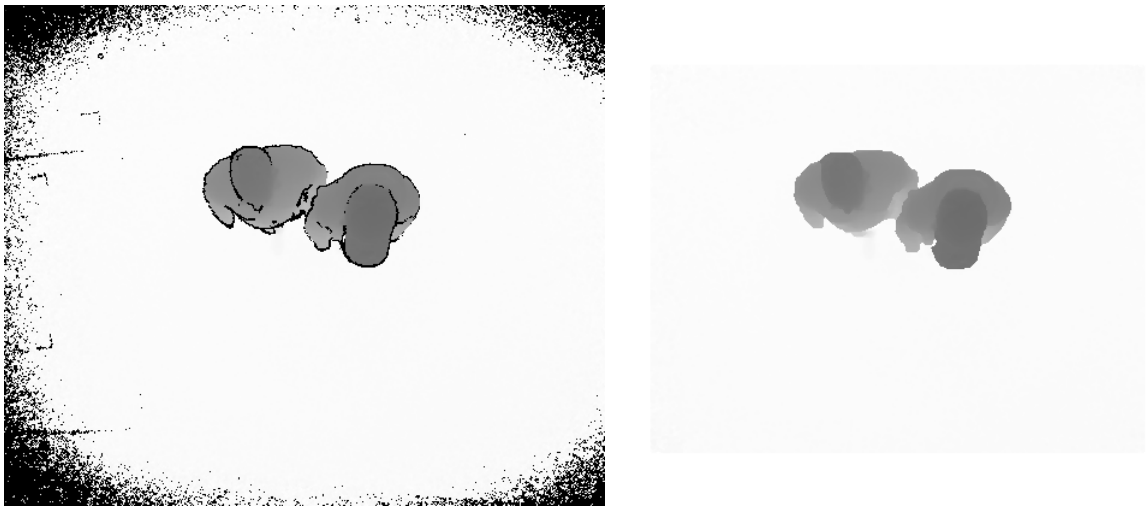
   Tomando como valores no válidos entonces:

$$\theta[Z_{i,j}^{null}] = \theta[Z_{i,j}^{null-camera}] \cup \theta[Z_{i,j}^{null-h_{max}}] \tag{3.3}$$

   Para cada $\theta[Z_{i,j}^{null}]$ se estimará un valor $\hat{Z}_{i,j}$ en función de los valores que contengan sus vecinos, empleando un nivel de vecindad máximo de 5. La matriz resultante después de eliminar los píxeles erroneos se denominará $H^{val} \in \Re^{MxN}$.

2. **Filtro mediana**: Sobre la matriz H^{val} se apicará un filtro de mediana de 9 elementos para eliminar posibles "flying pixels". Una vez realizado este filtrado obtendremos una nueva matiz de alturas:

$$H = median(H^{val}) \tag{3.4}$$



(a) Frame original con ruido (b) Frame filtrado

Figura 3.3: Resultado del filtrado de la imagen obtenida del sensor TOF

3.5 Detector de máximos

El objetivo de este bloque es localizar dentro de la matriz H los posibles elementos que puedan incluirse en el grupo candidatos a persona, para ello el primer paso es localizar todos los máximos (picos) existentes en la matriz H. Dicho algoritmo se detalla a continuación:

1. **División de la matriz H en subregiones (SR's)** Con el fin de poder analizar la imagen en mayor detalle se dividira la imagen en regiones de menor tamaño. El mecanismo de este algortimo se basa en, suponiendo que tengamos una cámara TOF con una resolución M X N, la división de la matriz **H** en n_1 **y** n_2 subregiones cuadradas de tamaño **D x D**, de la forma:

$$n_1 = \frac{M}{D}; n_2 = \frac{N}{D} \tag{3.5}$$

El tamaño de la subregión D se fija en función de las necesidades del algoritmo, es dependiente de la altura a la que se encuentra colocada la cámara, la altura mínima de las personas y el área mínima que se quiera diferenciar. En el caso de este programa se busca un tamaño de subregión lo suficientemente pequeño mediante el cual dentro de una misma persona puedan encontrarse varias de estas subregiones, por ello el caso más crítico será el de las personas de menor tamaño. De esta forma la estructura de una subregión se muestra a continuación:

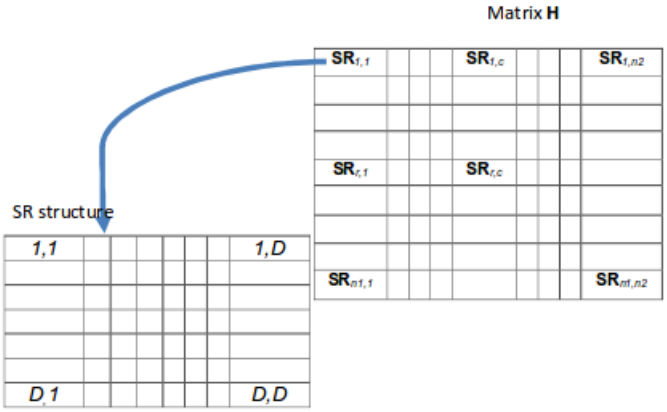


Figura 3.4: División de la Matriz H en subregiones D x D

Cada una de las subregiones formará una nueva matriz $SR_{r,c}$

$$\left[SR_{r,c}\right] = \begin{bmatrix} Z_{(r-1)D+1,(c-1)D+1} & Z_{(r-1)D+2,(c-1)D+1} & \cdots & Z_{(r-1)D+D,(c-1)D+1} \\ Z_{(r-1)D+1,(c-1)D+2} & Z_{(r-1)D+2,(c-1)D+2} & \cdots & Z_{(r-1)D+D,(c-1)D+2} \\ \cdots & \cdots & \cdots & \cdots \\ Z_{(r-1)D+1,(c-1)D+D} & Z_{(r-1)D+2,(c-1)D+1} & \cdots & Z_{(r-1)D+D,(c-1)D+D} \end{bmatrix} \tag{3.6}$$

siendo $r = 1, 2, ..n_1$, $c = 1, 2, ..n_2$ $(SR_{r,c} \in R^{DxD})$

2. **Formación de la matriz de máximos** A partir de todas las **SR** formadas se obtendrá una matriz de máximos denominada H^{maxSR} cuyos elementos se corresponderán con los máximos de cada una de las subregiones $(H^{maxSR} \in R^{n1xn2})$

$$\left[H^{maxSR}\right] = \begin{bmatrix} h_{1,1}^{maxS}R & h_{2,1}^{maxSR} & \cdots & h_{n1,1}^{maxSR} \\ h_{1,2}^{maxSR} & h_{2,2}^{maxSR} & \cdots & h_{n1,2}^{maxSR} \\ \cdots & \cdots & \cdots & \cdots \\ h_{1,n2}^{maxSR} & h_{2,n2}^{maxSR} & \cdots & h_{n1,n2}^{maxSR} \end{bmatrix} \tag{3.7}$$

3. **Localización de los máximos** Una vez obtenida la matriz H^{maxSR} cada uno de los valores de la

misma puede considerarse candidato a persona mientras se cumpla la siguiente condición:

$$h_{pmin} \leq h_{r,c}^{maxSR} \geq h_{r\pm m,c\pm g}; m = 0,1; g = 0,1 \tag{3.8}$$

donde h_{pmin} representa la estatura mínima de las personas a detectar $h_{pmin} = 100cm$. De esta forma cada uno de los elementos de la matriz será un máximo siempre que sun vecinos (nivel de vecindad = 1 en sus ocho direcciones) tengan valores menores (o iguales) al mismo. 3.5. Dichos máximos se indentificarán por $P_{r,c}^{k=1}$. Un ejemplo de esto se muestra en la figura 3.5

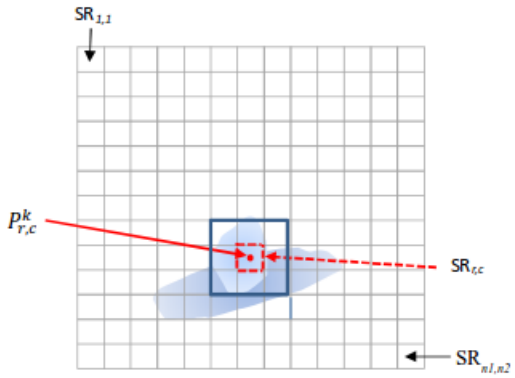


Figura 3.5: Esquema obtención de los máximos en H^{maxSR}

El principal problema existente en este bloque es la localización de varios máximos pertenecientes a una misma persona, esto se debe a que, debido al reducido tamaño de las **SR** es muy probable que **SR's** próximas que pertenezcan a una misma persona sean a su vez $P_{r,c}^{k}$. Se considerarán máximos cercanos los que cumplan la siguiente condición:

$$SR_{r\pm m,c\pm g} m = 0,1,2; g = 0,1,2; 0 \leq g + m \leq 4 \tag{3.9}$$

Es decir, cualquier $P_{r,c}^{k}$ que se ecuentre en un radio de vecindad dentro del rango $(r \pm m, c \pm g)$ de otro $P_{rn,cn}^{kn}$, se analizará como un único máximo en la posición del elemento con el valor superior.

Una vez realizado este proceso cada uno de los $P_{r,c}^{k}$ con $k = 1, \ldots N_p$ localizados será candidato a pertenecer a una persona, por lo tanto N_p será el número de posibles personas en la escena.

3.6 Obtención de las regiones de interés ROI

Se define la ROI como la región perteneciente a un máximo de la cual se extraerán las características, es decir, el área que delimita la pertenencia de un pixel a un máximo o a otro. El objetivo del trabajo es conseguir detectar personas diferenciándolas de otros elementos a partir de una cámara TOF cenital. Respecto a este sistema de coordenadas, la ROI de un máximo candidato a persona deberá comprender cabeza, cuello y hombros. Recordando que estas cámaras se basan en distancias y, teniendo en cuenta las medias antropométricas del cuerpo humano, se puede establecer que estas partes del cuerpo deberán estar incluidas en un rango de no más de 40cm por debajo del punto máximo encontrado h^{max}, de la forma:

$$h^{interest} = 40cm \tag{3.10}$$

Para evaluar los píxeles vecinos a un máximo y decidir cuales de ellos pertenecen a la región de interés

de un máximo $P^k_{r,c}$, se establece un radio de vecindad asociado de N niveles de vecindad(L=1, ...,N) y 8 direcciones. Se evaluará cada una de las 8 direcciones de forma independiente con el fin de obtener todas las SR que abarca la forma del elemento a analizar.

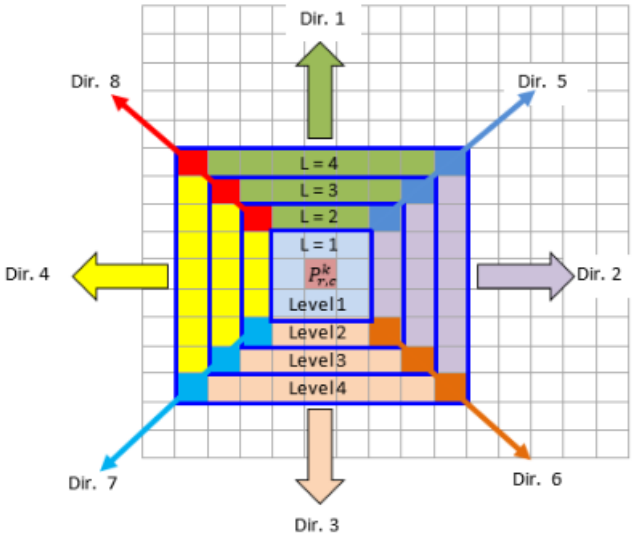


Figura 3.6: Direcciones de busqueda se SR ∈ ROI

La aceptación de una SR dentro del contorno de una persona sigue la siguientes premisas:

1. **Nucleo**:
   Se denominará nucleo a la subregion $SR^k_{r,c}$ donde se ha localizado el máximo $P^k_{r,c}$, esta subregión pertenecerá siempre a $ROI^k_{r,c}$.

2. **Nivel L=1:**
   El nivel L=1 estará formado por los 8 vecinos de la SR donde se sitúe el máximo $P^k_{r,c}$ bajo análisis. Dada una subregión SR situada en el nivel L=1 se considerará perteneciente al máximo si su valor se encuentra dentro de la región de interés $h^{interest}$.

$$h^{max}_{SR} \geq h^{maxSR}_{r,c} - h^{interest} \tag{3.11}$$

3. **Direcciones 1,2,3 y 4**
   Dentro de estas direcciones se analizará las SR pertenecientes a los niveles L=2,3,4 ya que debido a la posición y características de la cámara el espacio que una persona no abarcará mas niveles. La pertenencia o no de una subregión al máximo depende de cuatro premisas:

   - Para que una SR perteneciente a un nivel L (L=2,...4) en una determinada dirección pertenezca a un máximo, es necesario que en el nivel inferior contenga al menos L-1 SR en la misma dirección.
   - Para que una SR perteneciente a un nivel L (L=2,...4) en una determinada dirección pertenezca a un máximo, es necesario que la SR anterior en esa misma dirección pertezca al máximo.
   - El valor máximo, localizado en la matriz de máximos H^{maxSR}, de la SR a analizar debe encontrarse en el rango de altura de interés delimitado por $h^{interest}$.

$$h^{max}_{SR} \geq h^{maxSR}_{r,c} - h^{interest} \tag{3.12}$$

- Dado que el interés de centra en separar el final de un individuo y el comienzo del siguiente, cuando ambos se encuentran juntos. Se establece la condición de que en una determinada dirección el valor máximo de la SR bajo análisis debe ser menor que la SR en el nivel anterior y mayor que el valor de la SR en el nivel siguiente.

$$h_{SR-1}^{max} \geq h_{SR}^{max} \geq h_{SR+1}^{max} \tag{3.13}$$

  Si dicha condición no se cumpliese la SR bajo análisis constituirá un mínimo entre dos máximos, lo que se interpretaría como el fin de las SR pertenecientes a un máximo en esa dirección.

4. **Direcciones 5,6,7 y 8**
   Esta direcciones se corresponden con las cuatro diagonales que parten desde $P_{r,c}^{k}$, como se indica en 3.6 el rango de niveles que se analizarán será L=2,3,4. La pertenencia o no de una subregión al máximo depende de tres premisas:

   - Para que una subregión perteneciente a un nivel L (L=2,...4) en una determinada dirección pertenezca a un máximo, es necesario que la SR anterior en esa misma dirección pertezca al máximo.
   - El valor máximo, localizado en la matriz de máximos H^{maxSR}, de la SR a analizar debe encontrarse en el rango de altura de interés delimitado por $h^{interest}$.

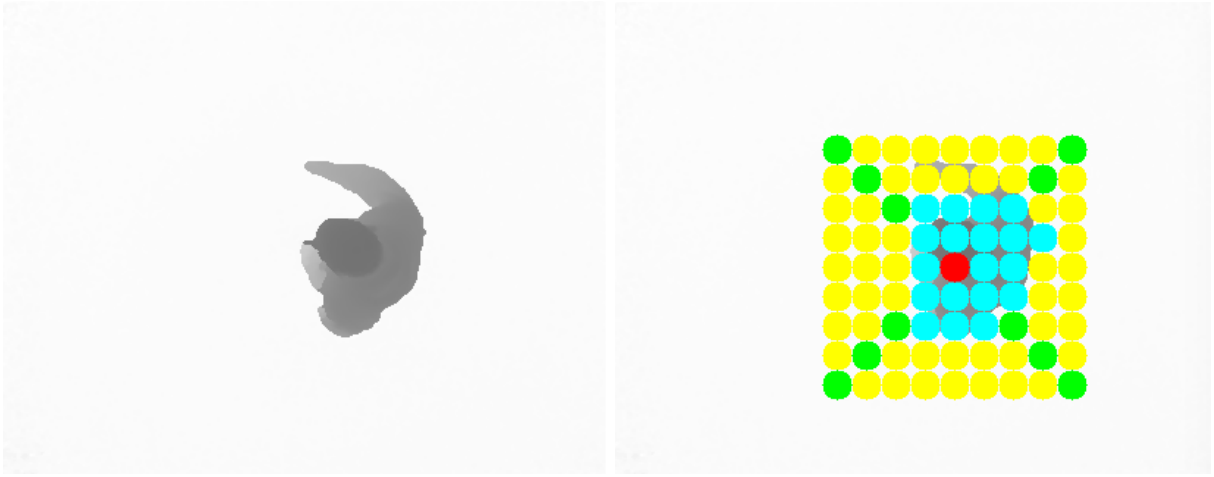
$$h_{SR}^{max} \geq h_{r,c}^{maxSR} - h^{interest} \tag{3.14}$$

   - Se establece la condición de que, en una determinada dirección el valor máximo de la SR bajo análisis debe ser menor que la SR en el nivel anterior y mayor que el valor de la SR en el nivel siguiente.

$$h_{SR-1}^{max} \geq h_{SR}^{max} \geq h_{SR+1}^{max} \tag{3.15}$$

     Si dicha condición no se cumpliese la SR bajo análisis constituirá un mínimo entre dos máximos, lo que se interpretaría como el fin de las SR pertenecientes a un máximo en esa dirección.

En la figura 3.7 se expone un ejemplo del análisis de las regiones (SR) adyacentes a un máximo. Para la obtención de la región de interés se ha estudiado la relación de cada uno de los píxeles en 4 níveles de vencindad y en las 8 direcciones expuestas en 3.6. En 3.7b se muestra con un punto rojo el máximo localizado en la escena con una única persona mostrada en 3.7a. Representado en amarillo se encuentran todas las regiones analizadas correspondientes a las direccciones 1,2,3,4 las cuales no cumplen las condiciones necesarias para pertenecer al máximo correspondiente, y en verde las regiones analizadas en las direcciones 5,6,7 y 8 donde no se cumplen las condiciones necesarias. En color azul se muestran las subregiones correspondientes a las 8 direcciones expuestas, las cuales, dado que las condiciones anteriores si se cumplen, pertecenerían a la ROI del máximo analizado.



(a) Vista cenital persona (b) Subregiones analizadas

Figura 3.7: Obtencion ROI mediante la búsqueda de subregiones pertenecientes a un máximo

3.7 Extracción de las características

Para la clasificación de cada una de las ROI (de cara a determinar si corresponde, o no, a una persona), será necesario obtener un vector de características. Este vector se obtendrá siguiendo el procedimiento que se describe a continuación. Cada $ROI^k_{r,c}$ obtenida perteneciente a $P^k_{r,c}$ puede presentarse como cadidato a persona. A partir de la $ROI^k_{r,c}$ asociada, obtenida en 3.6, se extraerá un vector de características *Vcaract*. Dicho vector estará formado por seis componentes, cinco de las cuales estarán relacionadas con la superficie de los *candidatos a persona* visible desde la cámara a diferentes alturas. La sexta componente del vector identifica la relación entre el diámetro menor y el mayor de la superficie identificada (cabeza).

3.7.1 Obtención del vector de características

Cada una de estas características se basa en la relación existente entre el porcentaje de superficie captada por la cámara y el número de píxeles asignados. Esta superficie comprenderá cabeza, cuello y hombros (debido a la posición cenital) en el caso de una persona, los cuales como se menciona en 3.10 comprenderán una distancia como máximo de 40cm. La obtención de dichas características se basa en agrupar la cantidad de píxeles, es decir, de puntos de la superficie de la persona visibles por la cámara, caracterizando su forma.

1. **Obtención del número de píxeles por alturas.** Dado que, como se ha mencionado, se pretende buscar las características de la $ROI^k_{r,c}$ que correspondan a la cabeza, cuello y hombros de la persona, se establecerá una primera clasificación en franjas de altura de $\Delta h = 2cm$. Esto es, se establecen 20 franjas v_s(s=1,...20) de 2cm partiendo desde $P^k_{r,c}$ recorriendo los 40 centímetros establecidos en $h^{interest}$ contabilizando el número de píxeles dentro de cada franja $v_s \in ROI^k_{r,s}$.

   De esta forma la información útil para la clasificación del elemento, que se localiza en los 40 centímetros inferiores al valor del máximo $P^k_{r,c}$, queda agrupada en las secciones de $\Delta h = 2cm$ descritas descartando el resto de información de $ROI^k_{r,c}$ no útil para la clasficación.

   Se genera entonces para cada $ROI^k_{r,c}$ el vector v, donde el valor de cada componente v_s coincide con el número de medidas de profundidad (pixeles) dentro de los límites $[(s-1)*\Delta h, s*\Delta h)$.

2. **Vector de características.**

   Las componentes del vector v son muy sensibles a las diferentes personas (altura de cuello, tipo de peinado, color de pelo, ...) además su visibilidad se ve afectada por la posición de la persona dentro de la escena. En el la figura 3.8 se representan los valores del vector de secciones de 2cm v, para diferentes alturas de una misma persona.

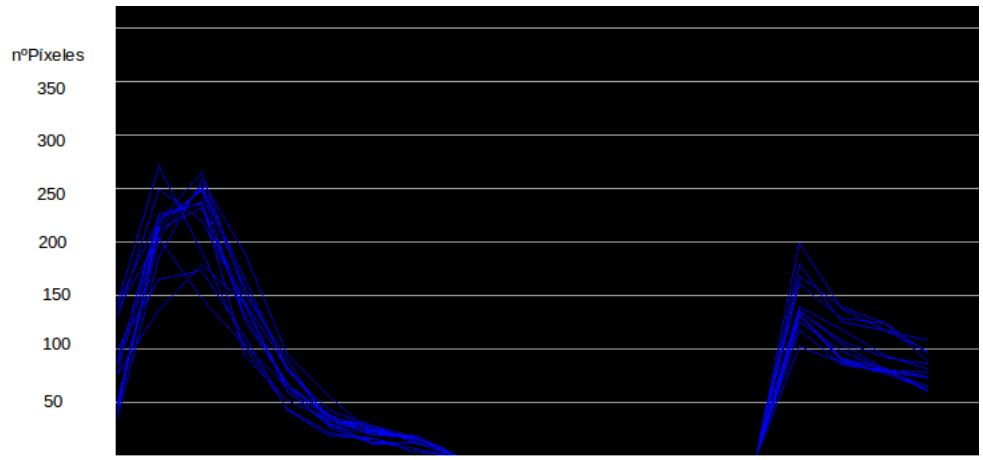


Figura 3.8: Secciones de 2cm correspondientes al vector v de una secuencia con un mismo individuo

Para minimizar los ruidos o variaciones descritas se acumulan las características del vector v en nuevo vector v^R de 5 elementos. Cada uno de los componentes v_i^R (i=1,...5) acumulará los valores de tres componentes de v. El fin de este procedimiento es conseguir obtener un nuevo vector invariante de la altura y características de la persona. Las tres primeras componentes del vector caracterizarán la cabeza, mientras que las 2 ultimas, los hombros de la persona.

Para obtener las tres primeras componentes del v^R se emplea el siguiente procediminto:

$v_i = max\{v_1, v_2, v_3\}, con\ i = 1, 2, 3$

$i = 1 \rightarrow v_1^R = \sum_{s=1}^{3} v_s;\ v_2^R = \sum_{s=4}^{6} v_s;\ v_3^R = \sum_{s=7}^{9} v_s$

$i = 2, 3 \rightarrow v_1^R = \sum_{s=i-1}^{i+1} v_s;\ v_2^R = \sum_{s=i+2}^{i+4} v_s;\ v_3^R = \sum_{s=i+5}^{i+7} v_s$

Mediante este proceso se localiza la posción v_i donde se encuentra el verdadero máximo del elemento a analizar ya que, puede darse la situación en la que el máximo se establezca en un pixel asilado del resto (por ejemplo alguna parte del cabello mas elevada). De esta forma se localiza el elemento v_i donde el número de puntos acumulado sea mayor, el cual pertenecerá a la parte de superior del objeto, o de la cabeza si se tratase de una persona. El valor del nuevo máximo se localizará en la franja v_i de la forma: $h_{r,c}^{maxSR} = h_{r,c}^{maxSR} - \Delta h(i-1)$

De la misma forma que la búsqueda de la cabeza se realizará el proceso de búsqueda de los hombros, partiendo esta vez la búsqueda de la región con mayor número de píxeles (se asume que pertenerá a la parte superior de los hombros) del elemento v_{10}. Por lo tanto la obtención de las características v_4^R y v_5^R, que se corresponden con la localización de los hombros de la persona en la imagen de profundidad, para ello seguirá el siguiente procedimiento:

$v_i = max\{v_{10}, v_{11}, \ldots v_{20}\}, con\ i = 10, 11, \ldots 20$

$v_4^R = \sum_{s=i-1}^{i+1} v_s;\ v_5^R = \sum_{s=i+2}^{i+4} v_s;$

Realizando este proceso las compentes v_4^R y v_5^R se vuelven invariantes a las diferentes longitudes de cuello que pueda tener la persona ya que, se crearán a partir de los píxeles pertenecientes a los hombros (se produce un máximo de medidas en los hombros debido a que su superficie es mayor).

Las cinco componentes descritas del vector v^R aportarían información acerca del área o superficie de los objetos o personas de la escena vistas desde la cámara. La sexta componente del vector incluye información acerca de la forma que tiene el área abarcada por los componentes v_1^R, v_2^R, v_3^R, en los cuales, si se tratara de una persona, estarían localizados los puntos de la cabeza. Esta información se basa en tomar la forma de elipse de la cabeza, para establecer una relación entre el eje mayor y menor de la misma.

Figura 3.9: Extracción de los ejes pertenecientes a la cabeza, componente v_6^R

3. **Normalización del vector de características** El número de pixeles asociados a una determinada franja de altura Δh que la cámara capte tiene una fuerte dependencia con la altura del individuo. Cuanto mayor es su altura, más cerca se encontrará de la cámara, por lo tanto mayor será la cantidad de píxeles que ocupe la forma de la persona en la imagen. Debido a esto las componentes del vector v_i^r $(i = 1 \ldots 5)$ dependerán también de la altura del sujeto, siendo necesaria una normalización de los valores del vector de características en función de la altura.

   En la figura 3.11 se representan los diferentes valores del vector v^R para una misma persona en diferentes alturas. Para demostrar esta variación en la superficie de la persona en función de la altura, se ha empleado una secuencia perteneciente a la base de datos grabada (3.2) donde un mismo individuo se agacha y levanta en el mismo lugar en la escena (Figura 3.10).

(a) (b) (c)

Figura 3.10: Secuencia cambio de altura para evaluar la dependencia de la superficie con la altura

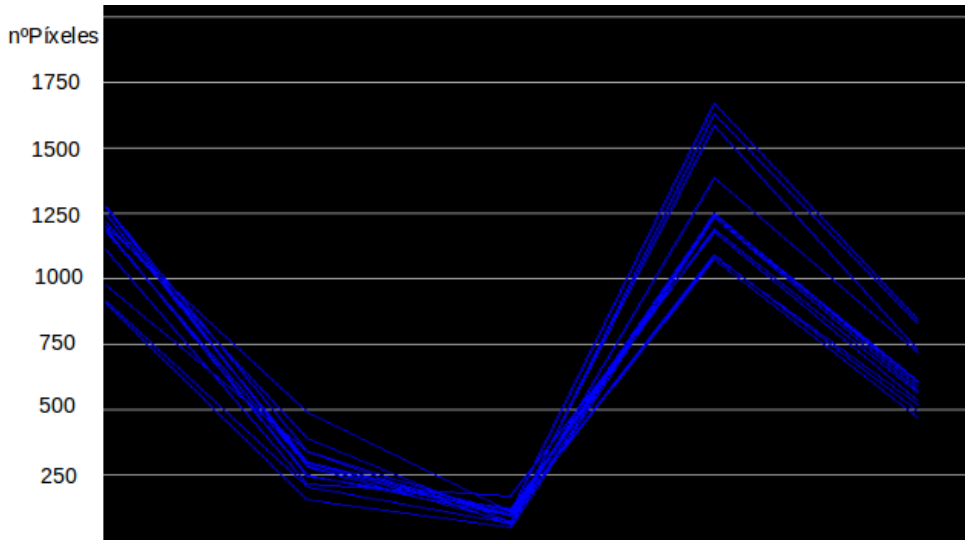


Figura 3.11: Secciones 6cm correspondientes a v^R

Para dicha normalización puede establecerse que la altura de la persona y el número de píxeles que abarca la superficie de su cabeza mantienen una relación lineal (aunque no exacta) que puede utilizarse como recta de normalización del vector v^R. En la figura 3.12 se muestra la relación lineal establecida mediante mínimos cuadrados para una sucesión de puntos de la cabeza asociados a diferentes alturas, de esta forma se podrán obtener los puntos estimados p̂ para una altura h^{max}.

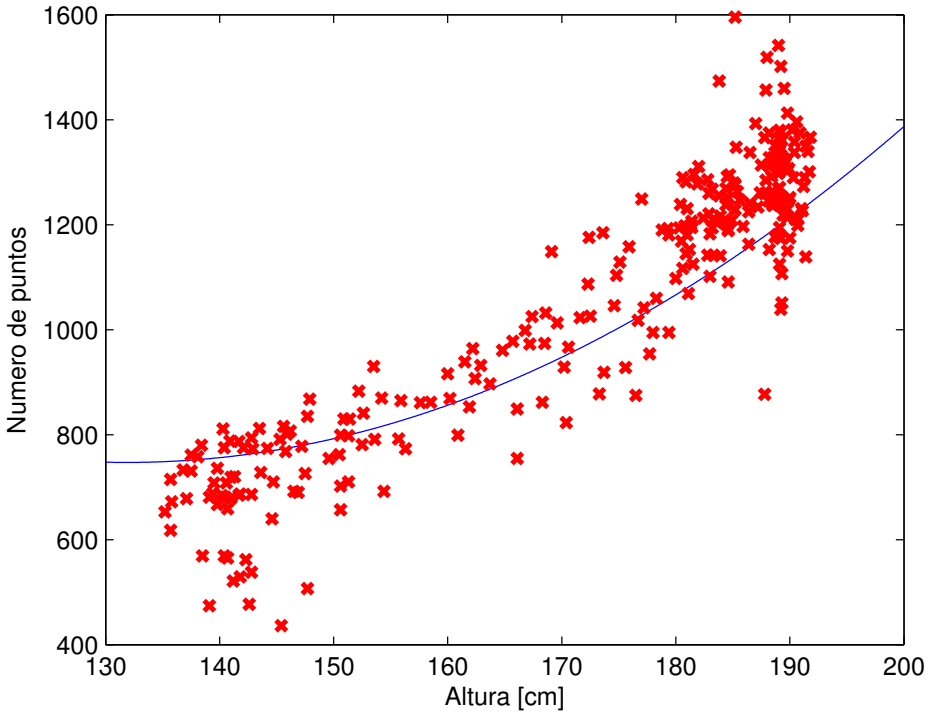


Figura 3.12: Recta de normalización en función de la altura v^R

$$\hat{p} = 0{,}1379 * (h^{max})^2 - 36{,}3939 * h^{max} + 2997{,}1714 \tag{3.16}$$

3.8 Clasificador

En esta etapa del algoritmo se realiza la clasficación de los vectores de características v^R obtenidos con el fin de diferenciar los vectores que identifican la forma de una persona con cualquier otro elemento existente. Para dicha clasficación se emplea el algoritmo basado en el análisis de componentes principales PCA explicado en 2.3.1. Dado que, como se ha mencionado, este tipo de clasificador se basa en un aprendizaje supervisado es necesaria una etapa de entrenamiento con datos conocidos, este proceso de aprendizaje se desarrollará durante el proceso "$Off-line$" (3.1) donde se utilizarán secuencias de datos conocidas y etiquetadas pertenecientes a la base de datos creada (3.2).

A partir de dicho conjunto de imágenes de entrenamiento se generarán las diferentes clases para la clasificación de los vectores de características v^R obtenidos mediante el algoritmo PCA explicado en 2.3.1.

3.8.1 Clases implementadas

Con el fin de abarcar la diversidad de personas (estatura, complexión, cabello ...), así como la posibilidad de que los sujetos lleven elementos en la cabeza (sombreros, gorras ...), ya que debido a esto sus vectores de características pueden tener variaciones en determinadas componentes, se han establecido diferentes clases dentro del clasificador PCA. Cada una de estas clases proporcionará al clasficador la información relativa al vector de características asociado, en la creación de cada una de las clases se han empleado secuecias de personas aisladas en la escena con pelo corto, pelo corto con altura superior a 190cm, pelo largo, sombreros y gorras. A continuación se muestran los vectores de características v^R de las diferentes clases implementadas.

1. **Clase pelo corto** La clase formada por personas de pelo corto es la que mejor representa el vector de características buscado, ya que se tiene gran visibilidad de las dos partes más representativas de la persona desde el ángulo empleado, cabeza y hombros. Clase implementada con 381 vectores de entrenamiento.

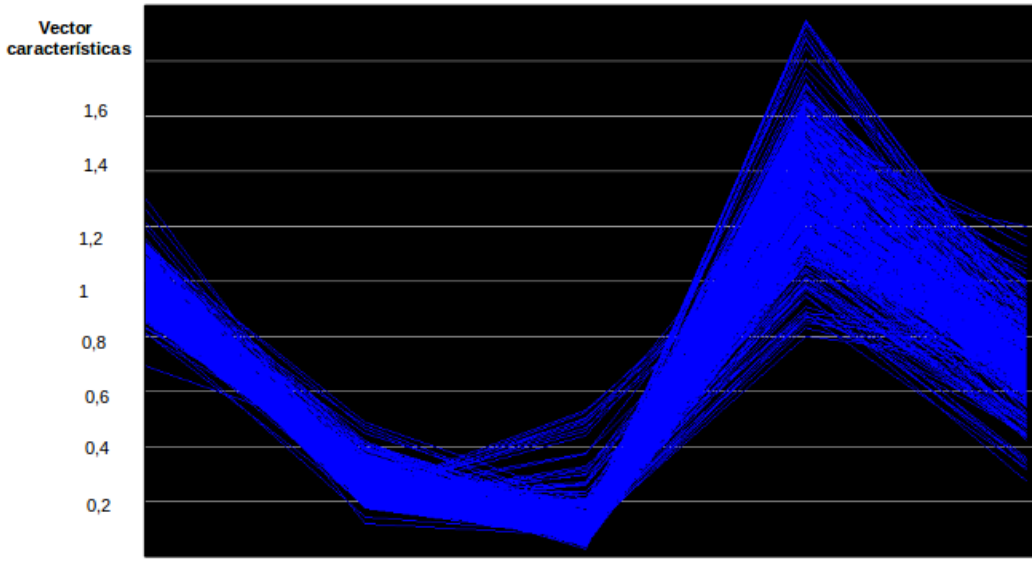


Figura 3.13: Vector v^R clase Pelo corto

Figura 3.14: Secuencia correspondiente a la clase Pelo corto

2. **Clase pelo corto Alturas mayores a 190cm** La creación de esta clase es necesaria debido a que en personas de alturas superiores a 190cm el factor de normalización se separa de la curva establecida.Clase implementada con 209 vectores de entrenamiento.

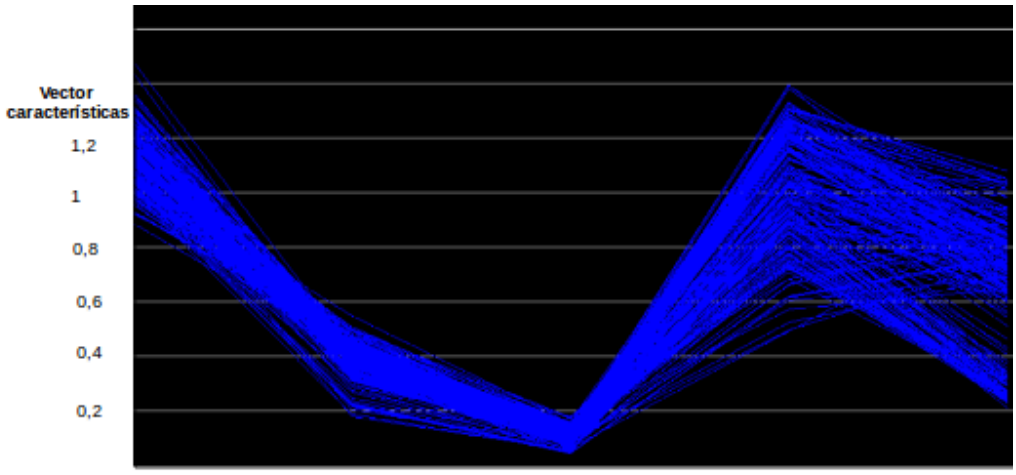


Figura 3.15: Vector v^R clase Pelo corto alturas superiores a 190cm

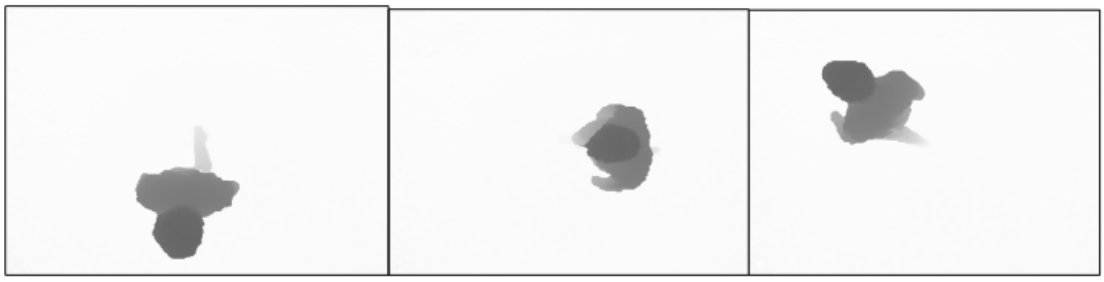


Figura 3.16: Secuencia correspondiente a la clase Pelo corto alturas superiores a 190cm

3. **Clase pelo largo** La creación de esta clase es necesaria debido a su diferencia con la anterior, esto es debido a que al tener el cabello largo, este tapa los hombros lo que supone que el elemento v_4^R posea un valor inferior al correspondiente en la clase "pelo corto". Para la etapa de entrenamiento de esta clase se han empleado 333 vectores pertenecientes a un individuo de pelo largo en el centro de la escena, ya que es en este punto donde más oclusión de los hombros se produce.

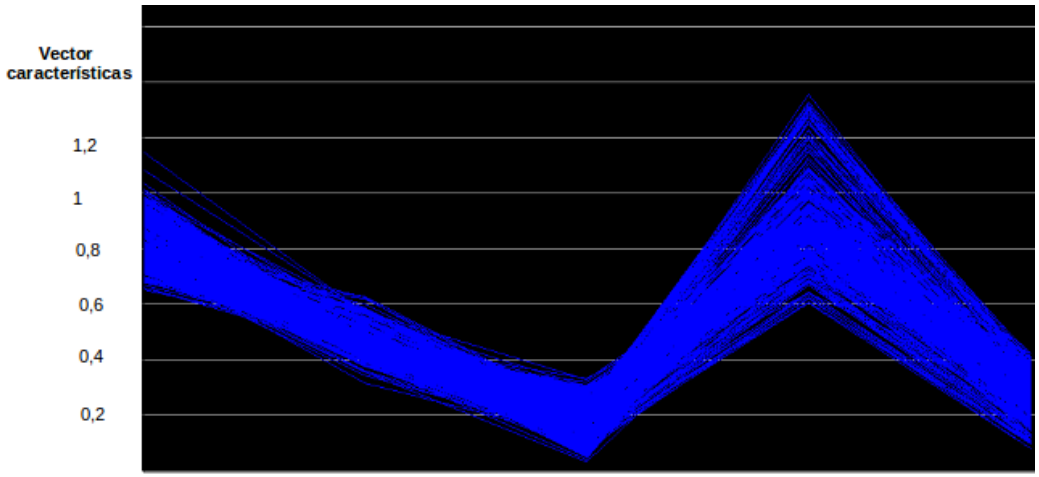


Figura 3.17: Vector v^R clase Pelo largo

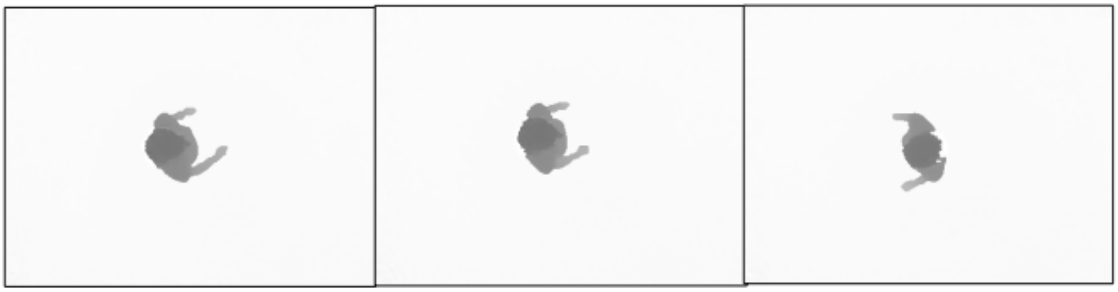


Figura 3.18: Secuencia correspondiente a la clase Pelo largo

4. **Clase Sombrero** Con el fin de abarcar la clasficación de personas con posibles accesorios en la cabeza se ha establecido la clase sombrero, la cual como se observa en la figura 3.19 difiere mucho de las clases anteriores en el valor del vector v_2^R, ya que las alas de sombrero impedirían la visibilidad del cuello de la persona. Clase implementada con 19 vectores de entrenamiento.

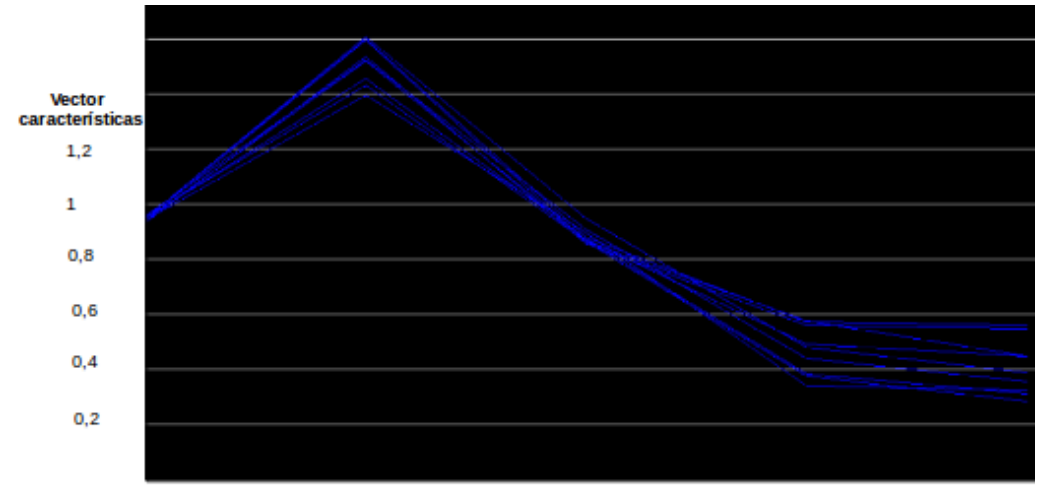


Figura 3.19: Vector v^R clase Sombrero

Figura 3.20: Secuencia correspondiente a la clase Sombrero

5. **Clase Gorra** De la misma forma que la clase anterior, se ha establecido la posibilidad de que los individuos a analizar lleven diferentes gorras sobre la cabeza. En esta clase se observa la diferencia de valor sobre el elemento V_2^R.Clase implementada con 17 vectores de entrenamiento.

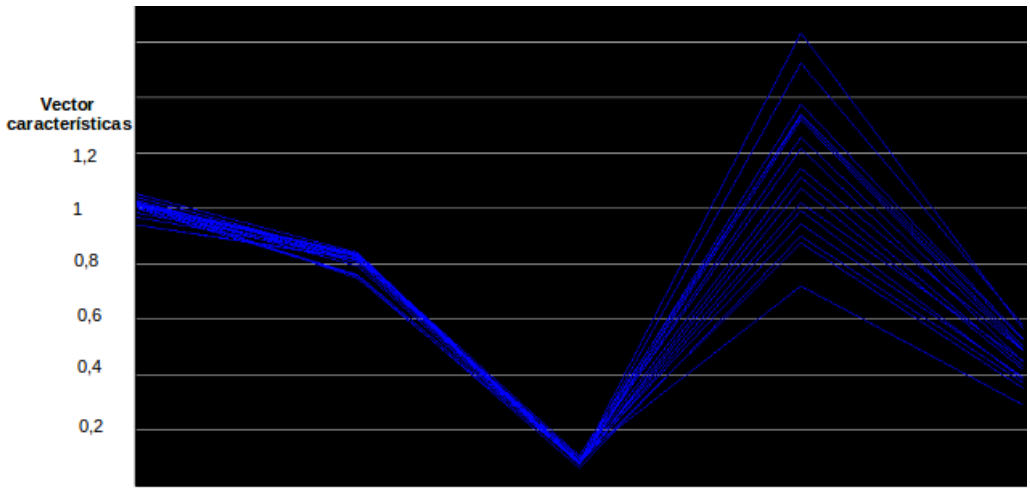


Figura 3.21: Vector v^R clase Gorra

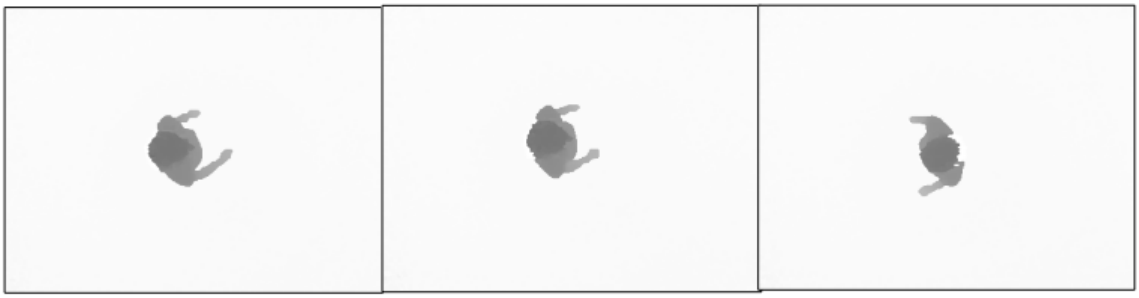


Figura 3.22: Secuencia correspondiente a la clase Gorra

Tras la declaración de dichas clases, se compararán las 3 primeras componentes principales del vector de características extraido (ya que son las que reciben mayor peso) perteneciente al máximo $p_{r,c}^k$ bajo análisis con cada una de ellas. Transformando el vector al sistema de coordenadas de cada clase y comparando el vector incial con el recuperado, se extraerá una medida de error relativa a la pertenenecia o no del vector a la clase. Para considerar un vector perteneciente a una de las clases, y por lo tanto detectar

su máximo correspodiente como persona, el error de recuperación mínimo obtenido tras la proyección con las cuatro clases deberá ser inferior al umbral de error fijado en $e_{umbral=}$0,25.

Capítulo 4

Resultados experimentales

En este capítulo se expondrán los resultados obtenidos con el algoritmo descrito. Todas las secuencias de prueba, así como las secuencias de entrenamiento, pertenecen a la base de datos grabada y etiquetada para esta aplicación 3.2. Se contrastarán los resultados con los generados empleando el detector de personas mediante filtro sombrero mejicano 2.1.1 y se realizará un análisis estadístico del algoritmo implementado sobre diferentes secuencias.

4.1 Análisis y comparación de resultados con sombrero mejicano

En este apartado se realizará una compararión de los resultados obtenidos mediante el algoritmo basado en la detección por el sombrero mejicano y los resultados obtenidos con el algoritmo desarrollado en este trabajo. Para el testeo y comparación de ambos algoritmos se emplearán las mismas imágenes de profundidad. Se utilizarán tres tipos de escenas diferentes donde aparecen un único sujeto en la escena, varios sujetos separados repartidos por la escena y un grupo de sujetos juntos.

1. **Secuencia 1:** un único sujeto en el centro de la escena (Figura 4.1)
    - La figura 4.1b muestra el resultado del análisis de la imagen de profundidad mediante el algoritmo basando en el filtro sombrero mejicano. Una vez realizado el preprocesado 2.1.1.1 la localización de la persona es una tarea sencilla, ya que la figura de la persona supondría el único contorno localizado en la escena y por lo tanto una única ROI en toda la imágen con un único máximo en ella, además no hay elementos que distorsionen la forma de su figura en la encena.
    - La figura 4.1c muestra el resultado del análisis de la imagen empleando el algoritmo de conteo desarrollado, el vector carácterístico v^R del sujeto pertenecería a la clase ”Pelo Corto"definida en el apartado 3.8.1.
2. **Secuencia 2:** varios sujetos repartidos por toda la escena (Figura 4.2). Esta escena a diferencia de la anterior cuenta con diversos individuos por toda la escena pero sin solaparse entre sí.
    - La figura 4.2b muestra el resultado del análisis mediante sombrero mejicano donde, aunque existen numerosos individuos en la escena, no se solapan entre sí, por lo tanto cada uno de ellos supondría un contorno, una ROI, a analizar de la misma forma que en la secuencia 1. Cabe destacar que existen 2 individuos cuya forma no se aprecia completamente en la escena,

debido a que se encuentran entrando, o saliendo, de la misma, pero el algoritmo aquí aplicado los detecta como si su forma se viera completa.

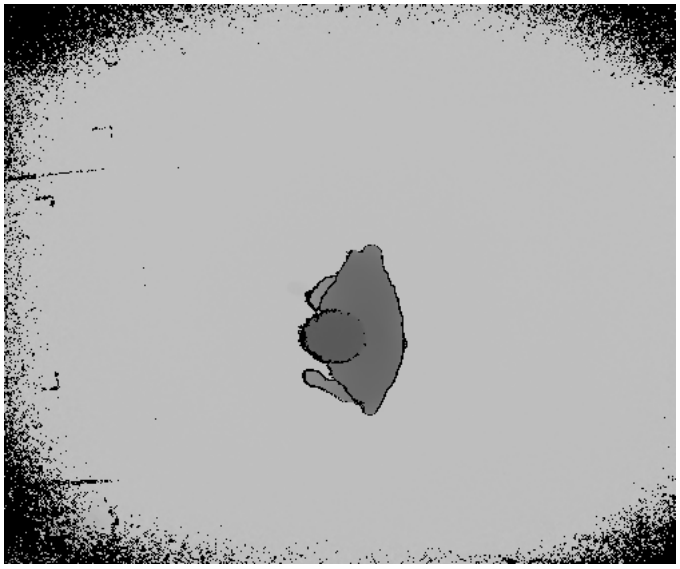
- La figura 4.2c muestra el resultado de la detección empleando el algoritmo desarrollado. La detección de las personas se realiza correctamente solo en aquellas en las que su forma se aprecia completamente. De este modo se demuestra que es necesaria la aparición de la persona completa para detectarla como tal,evitado así posibles falsas detecciones.

3. **Secuencia 3:** La figura 4.3 muestra un escena con varios sujetos situados muy próximos entre sí. Esta situación sería una de las más críticas a analizar debido a la dificultad de separar las partes de la imagen (3.6) que corresponderían a cada individuo.

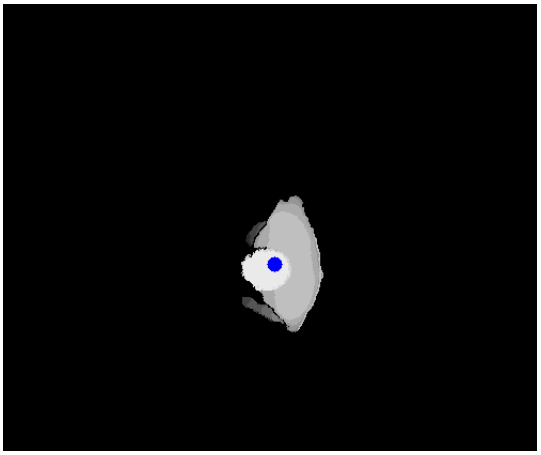
   - La figura 4.3b muestra nuevamente el resultado empleando el filtro sombrero mejicano. Este tipo se secuecias obtienen peores resultados debido principalmente a que la existencia en la imagen de personas muy próximas entre sí, supone la existencia de un único contorno, y por lo tanto una única ROI a analizar, de esta forma, la variación de alturas entre los individuos existentes en la escena es crítica. Como se puede observar, no todas las cabezas de los individuos se situarían, tras el reescalado de la ROI, a un mismo nivel, por lo tanto no a todos los individuos se les entregará la misma importancia, debido a esto los picos de intensidad resultantes tienen diferente valor, llegando incluso a juntarse
   - La figura 4.3c muestra la detección mediante el algoritmo desarrollado donde, al encontrarse todos los individuos dentro de la zona efectiva del sensor, son localizados correctamente. En este tipo de situaciones, donde es crítica la correcta separación de las regiones pertenecientes a cada individuo, el algoritmo implementado obtiene mejores resultados que el basado en sombrero mejicano.

   El desarrollo de este estudio [1], está basado en pruebas con sujetos varones y de alturas muy similares en todos los casos, por lo tanto, no sería un algoritmo de alta garantía de acierto en diferentes escenarios. Como vemos en esta última secuencia, donde los sujetos se encuentran muy cercanos entre sí,y tienen alturas muy diferentes el filtrado a través del sombrero mexicano no consigue una separación exacta de todos los contornos. De la misma forma la localización por contornos es demasiado inexacta a la hora de saber que píxeles de la imagen corresponden a un cadidato a persona y cuales a otro, debido a que sería necesario un cojunto de las operaciones morfológicas de erosión y dilatación por lo que se perdería información importante del contorno.

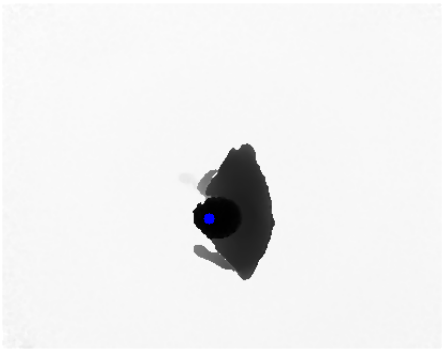
   Una vez analizado el trabajo se puede afirmar que el algoritmo propuesto únicamente supone una detección de bordes (uso más común del filtro LoG), añadiendo a este una detección de maximos, que proporciona un alto número de aciertos, exceptuando situaciones como la mostrada en la figura 4.3b, donde la existencia de dos máximos muy ceranos con valores de intensidad muy diferentes supone la detección de uno solo, sin incluir ninguna etapa de clasificación que diferencie personas de cualquier otro objeto que aparezca en la escena, por lo tanto detectará cualquier otro elemento presente como si de una persona se tratara.



(a) Imagen de profundidad

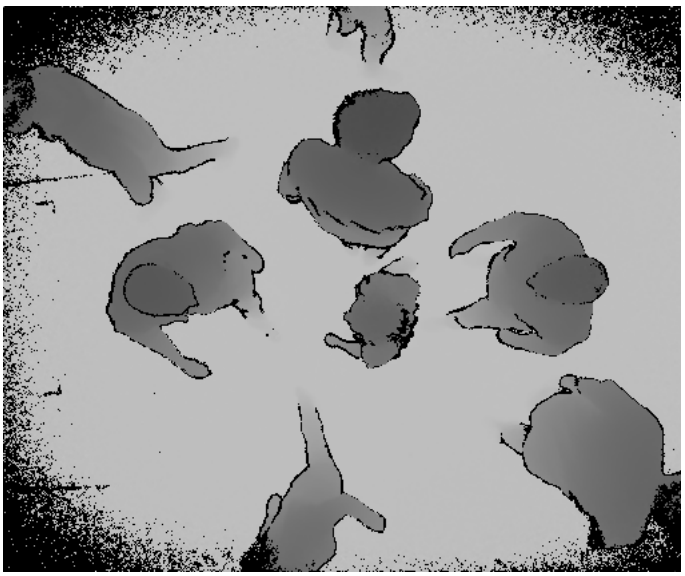


(b) Detección mediante LoG

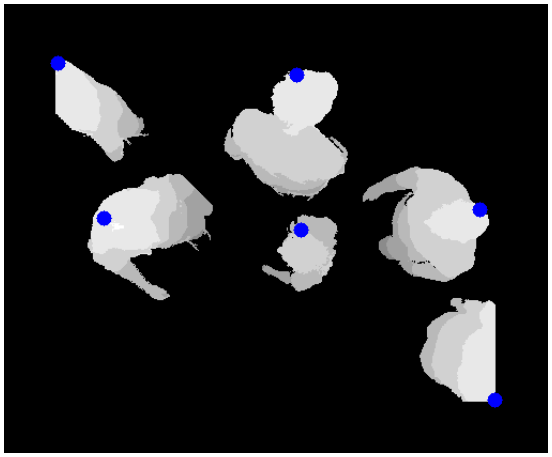


(c) Detección mediante PCA

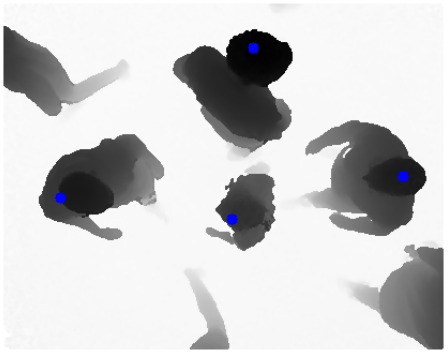
Figura 4.1: Comparación en la detección de una única persona en la escena mediante filtro LoG y algoritmo basado en clasificador PCA



(a) Imagen de profundidad

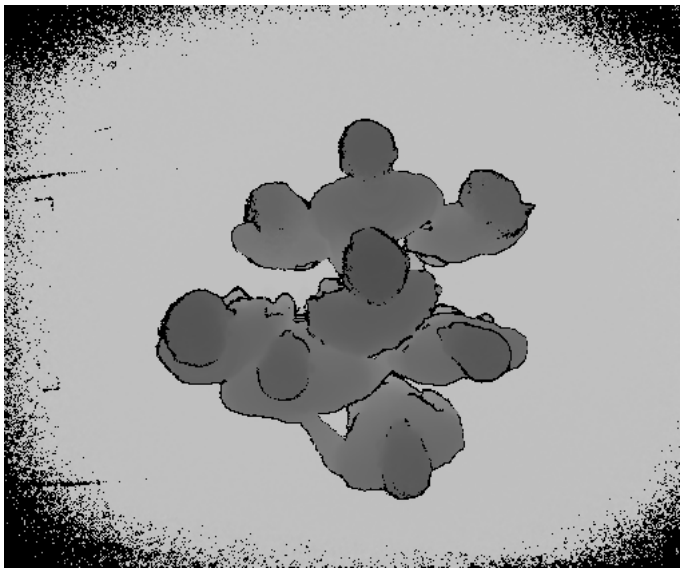


(b) Detección mediante LoG

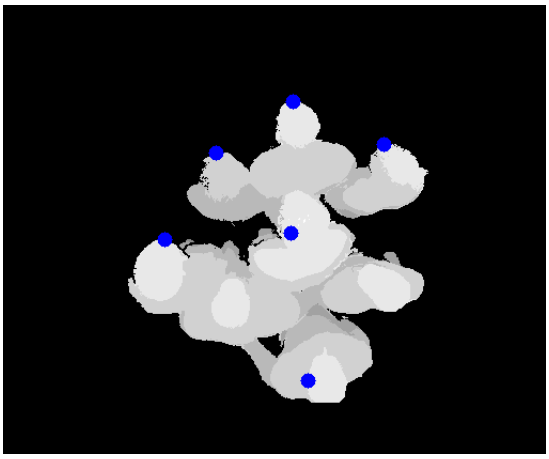


(c) Detección mediante PCA

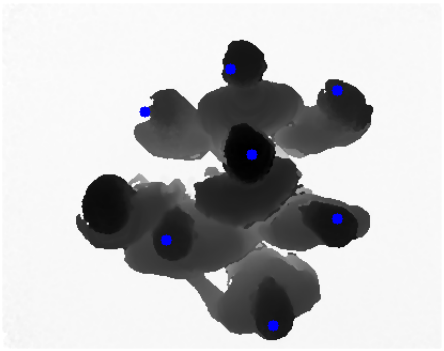
Figura 4.2: Comparación en la detección de varias personas separadas en la escena mediante filtro LoG y algoritmo basado en clasificador PCA



(a) Imagen de profundidad



(b) Detección mediante LoG



(c) Detección mediante PCA

Figura 4.3: Comparación en la detección de varias personas separadas en la escena mediante filtro LoG y algoritmo basado en clasificador PCA

4.2 Estudio del algoritmo implementado

En este apartado se expodrán las diferentes métricas de evalución y la efectividad del algoritmo basándose en el análisis de diferentes secuencias pertenecientes a la base de datos implementada 3.2, previamente etiquetadas, así como imágenes ejemplo de las diferentes secuencias y su detección mediante el algoritmo implementado.

4.2.1 Métricas de evaluación

Para evaluar el algoritmo implementado se han utilizado métricas de calidad ampliamente utilizadas cuando se trabaja en sistemas de detección, es decir, cuando la respuesta del sistema es un valor negativo o positivo. Dichas métricas se exponen a continuación.

- **Verdaderos positivos (VP):**
  Número de elementos etiquetados como personas y que nuestro algoritmo detecta como personas.
- **Verdaderos negativos (VN):**
  Número de elementos etiquetados como no personas y que nuestro algoritmo detecta como no personas.
- **Falsos positivos (FP):**
  Número de elementos etiquetados como no personas y que nuestro algoritmo detecta como personas.
- **Falsos negativos (FN):**
  Número de elementos etiquetados como personas y que nuestro algoritmo detecta como no personas.

4.2.2 Resultados de la detección en secuencias

En la siguiente tabla se muestran los valores métricos calculados expresados en porcentaje de VP, FP y FN sobre el número total de personas, entendiendo por número total de personas la suma de las personas que aparecen en cada imagen de una secuencia completa. Las secuecias a analizar, diferentes a las utilizadas en el entrenamiento, pertenecerán a la base de datos grabada (3.2) y han sido etiquetadas previamete, contrastando entonces, los resultados de la detección con los ficheros *"GroundTruth"* de cada secuencia.

Detección de personas					
Tipo secuencia	nºFrames	Total Personas	VP	FP	FN
Un único individuo en el centro de la escena	400	400	96.931 %	2.301 %	3.069 %
	374	374	97.771 %	4.178 %	2.228 %
	377	377	96.791 %	0.802 %	3.208 %
Un único individuo andando aleatoriamente	334	123	91.869 %	0 %	8.130 %
	332	90	95.555 %	1.111 %	4.444 %
	380	148	93.243 %	0 %	6.756 %
	277	91	98.901 %	0 %	1.098 %
Mútiples individuos andando aleatoriamente	1911	1482	87.314 %	0 %	12.685 %
Mútiples individuos andando en fila	621	248	94.354 %	0 %	5.645 %
Mútiples individuos andando juntos	897	386	74.352 %	0 %	25.647 %

Tabla 4.1: Valores métricos de detección calculados para diferentes secuencias

A continuación se mostrarán una serie de imágenes de profundidad correspondientes a las secuencias analizadas en 4.1 donde aparecen varios individuos. El análisis de estas secuencias es el más crítico debido a la oclusión de unos individuos en otros, que impide al sensor una visibilidad completa del sujeto, de la misma forma el análisis de las SR pertenecientes a cada individuo se vuelve más compleja ya que los contornos de cada persona se unen formando uno solo.

- **Secuencia múltiples individuos andando aleatoriamente:**
  En esta secuencia, mostrada en la figura 4.4, se aprecian un conjunto de personas distribuidas aleatoriamente por toda la escena. Como se ha mencionado anteriormente únicamente serán detectadas como personas aquellas cuyo contorno entero, o prácticamente entero (cabeza y hombros), se encuentre dentro de la zona de interés, evitando así posibles confusiones de contornos parciales de personas con otros elementos. Esta secuencia cuenta con un 87.314 % de detecciones correctas (VP), segun los datos analizados en 4.1, siendo el número de no detecciones (FN) situaciones creadas por oclusiones de unas personas con otras, donde el algoritmo no es capaz de diferenciar correctamente la ROI perteneciente a cada máximo.

- **Secuencia múltiples individuos andando en fila:**
  En esta secuencia, mostrada en la figura 4.5, aparecen un conjunto de individuos andando en linea. Este tipo de situaciones no suponen un gran conflicto para la detección debido a que la cabeza y

hombros de la personas se aprecian en su totalidad y al correcto funcionamiento del agoritmo de localización de mínimos entre personas (explicado en 3.13 y 3.15). Esta secuencia cuenta con un 94.354 % de detecciones correctas (VP), segun los datos analizados en 4.1.

- **Secuencia múltiples individuos andando juntos:**
  En esta secuecia, mostrada en la figura 4.6, aparecen un conjunto de individuos moviéndose por la escena formando un grupo. Este tipo de secuencias son las más críticas debido a la difultad de separación de los contornos pertenecientes a cada persona. Esta secuencia cuenta con un 74.352 % de detecciones correctas (VP), segun los datos analizados en 4.1. Las pricipales causas de la no detección en este tipo de secuencias es causada por la diferencia de alturas entre los individuos, donde puede aparecer confusión entre la detección de la cabeza de una persona y los hombros de la persona siguiente, cuando ambos se encuentran a la misma altura.

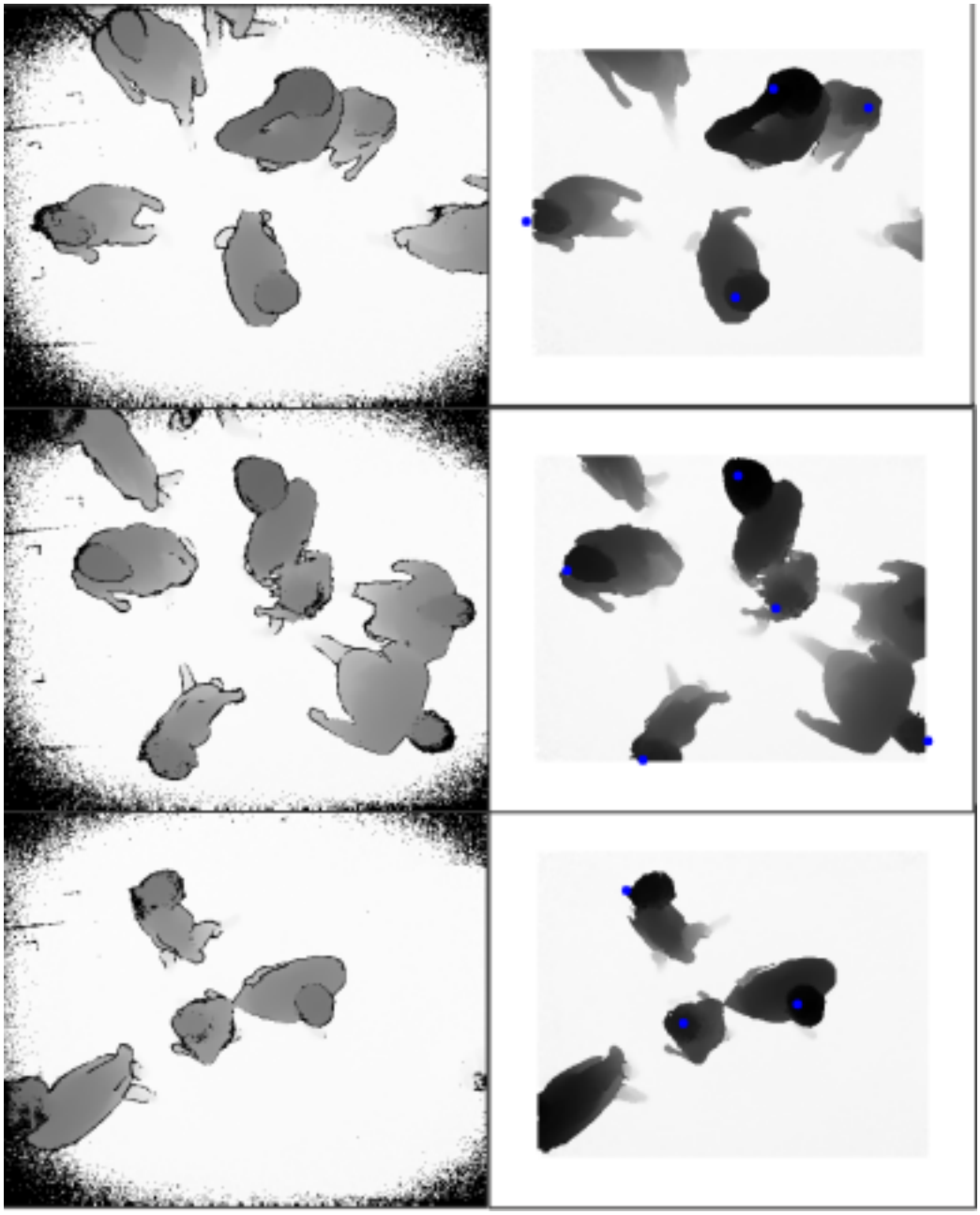


Figura 4.4: Secuencia analizada múltiples individuos andando aleatoriamente

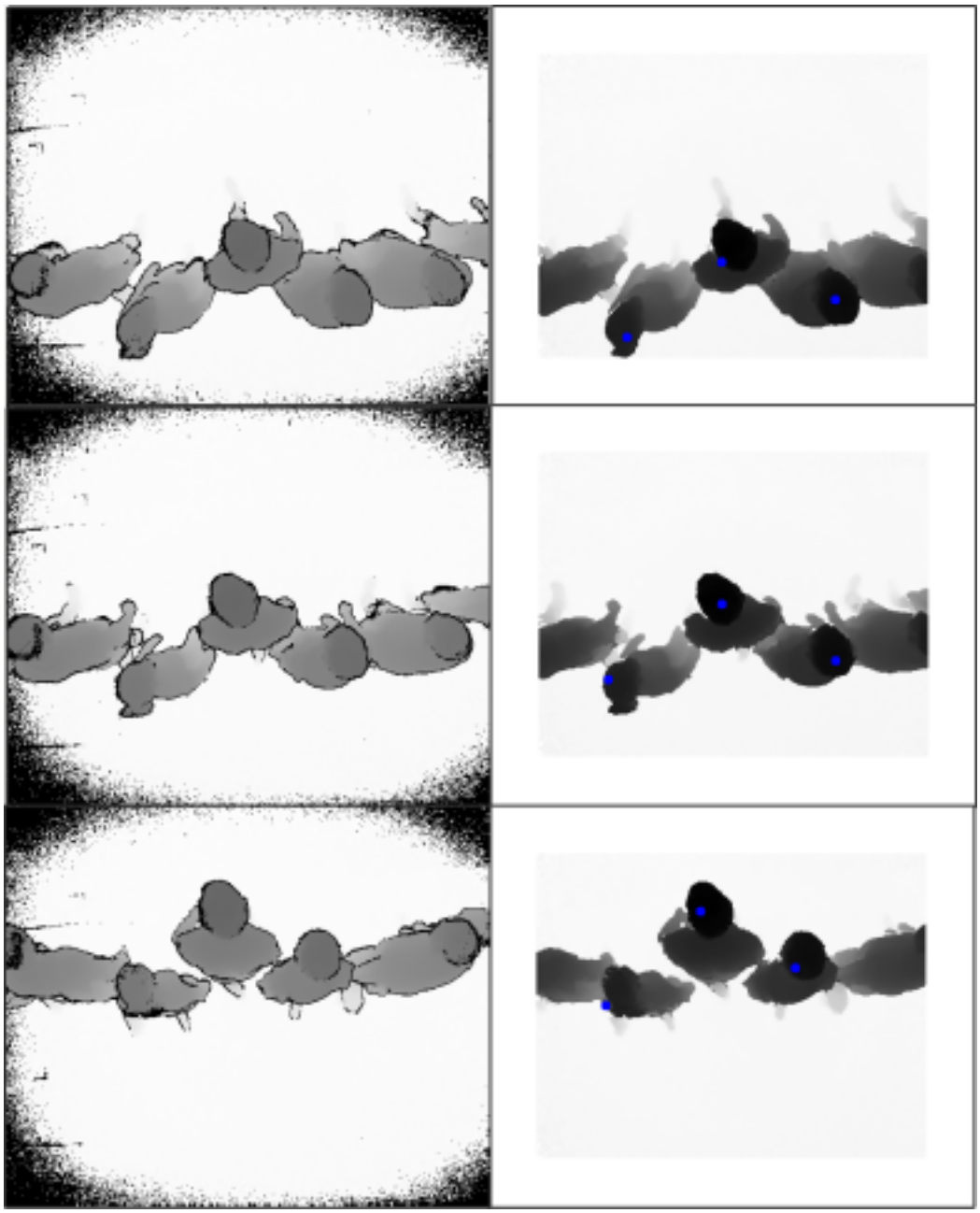


Figura 4.5: Secuencia analizada múltiples individuos andando en linea

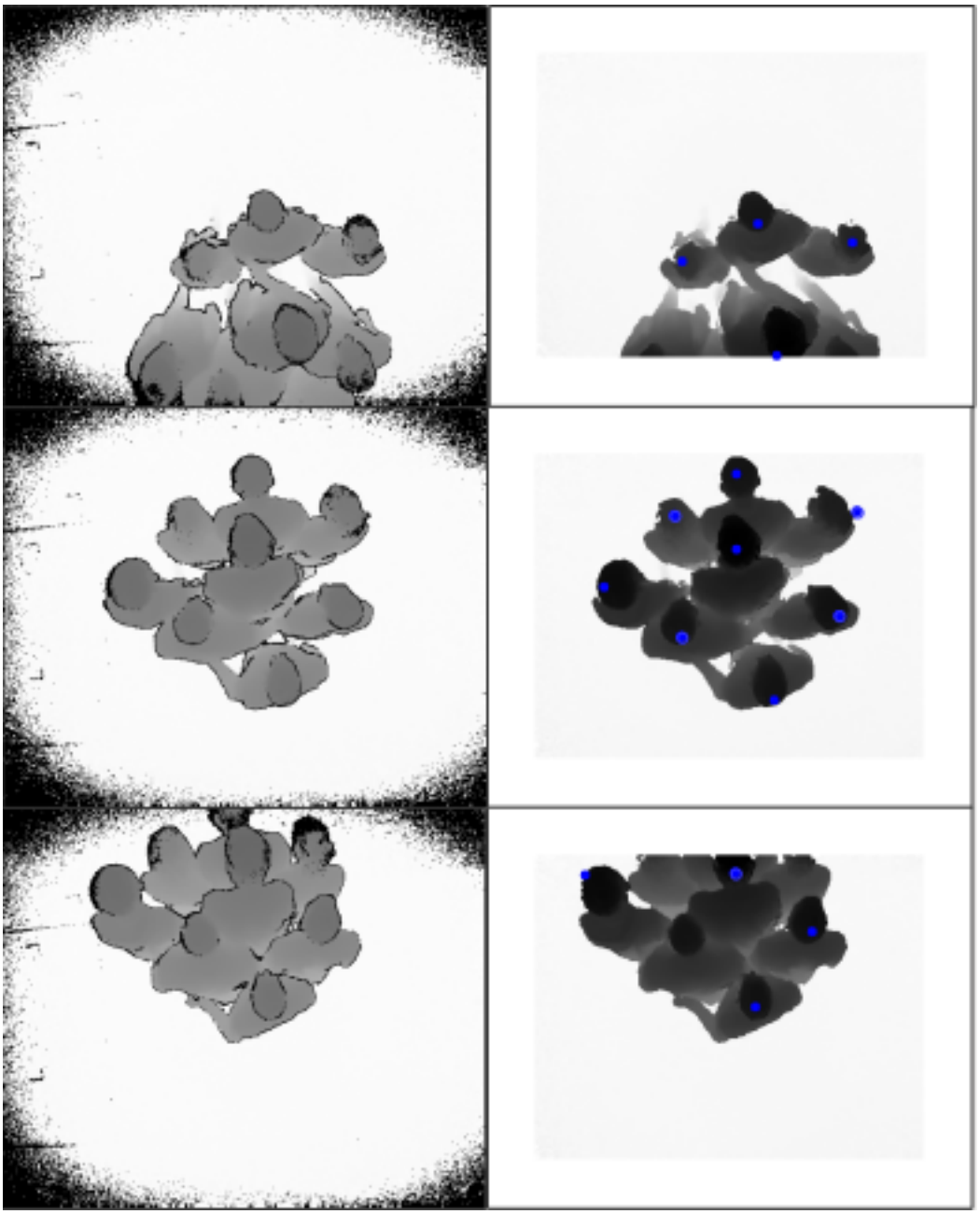


Figura 4.6: Secuencia analizada múltiples individuos andando juntos

4.2.3 Evaluación de tiempos de ejecución

En la figura 4.7, se muestra una gráfica con los tiempos de ejecución de las diferentes etapas del sistema. Los tiempos han sido extraidos a partir del análisis de una secuencia donde aparecen un gran número de personas aleatoriamente por toda la escena (figura 4.4) y han sido calculados como la media del tiempo empleado en cada frame durante una secuencia completa. Como se puede observar, el mayor porcentaje de tiempo se emplea en la extracción del vector de características de cada máximo.

$$T_{total} = 0{,}0329s \tag{4.1}$$

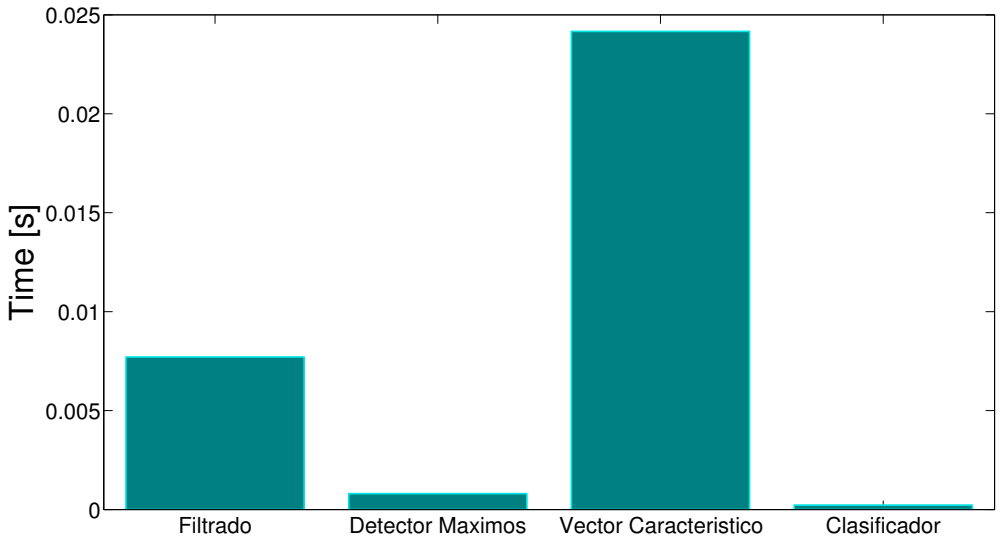


Figura 4.7: Tiempo de ejecución de una secuecia con múltiples individuos andado por la escena

Como se puede observar, a pesar de que la secuencia de entrada es compleja (incluye múltiples personas en la escena) el tiempo medio de procesamiento es de unos 0.0329 segundos/imagen. Esto implica que el algoritmo implementado podría analizar unas 30 imágenes por segundo.

Capítulo 5

Conclusiones y líneas futuras

En este capítulo se incluirán las conclusiones y los logros obtenidos en este trabajo, así como futuras lineas de desarrallo que puedan mejorar los resultados aquí obtenidos.

5.1 Conclusiones

Este trabajo se ha centrado en el análisis de los diferentes métodos de detección y conteo de personas mediante la utilización de sensores TOF .

En primer lugar se ha analizado el algoritmo descrito en [1] como posible detector, descartando su uso al comprobar su inefectividad en secuencias diferentes a las utilizas de muestra en dicho artículo. La necesaria similitud entre las alturas de los individuos presentes en la escena, y la detección de cualquier elemento presente en la escena como persona, hacen de este algoritmo un sistema no válido para la detección de personas.

El algoritmo desarrollado en este trabajo propone una solución al problema basándose en la utilización de un detector de máximos y un analizador de ROI pertenecietes a cada máximo, con el fin de separar los contornos pertenecientes a cada persona, ya que, el principal problema que aparece en escenas donde los usuarios se encuentran muy cerca unos de otros. Una vez obtenida la ROI perteneciente a cada máximo se extraerá el vector de características de dicho elemento. El clasificador utilizado, basado en el algoritmo de análisis de componentes principales PCA, utilizando vectores característicos que describen la forma de la cabeza y hombros de una persona vista desde el sensor y la circularidad de la cabeza, se encargará de asignar el elemento a alguna de las clases seleccionas o a ninguna, evaluando si se trata o no de una persona.

Una vez analizados los datos de las diferentes secuecias pertenecientes a la base de datos grabada (3.2), podemos afirmar que los resultados obtenidos empleando el algoritmo desarrollado superan a los resultados pertenecietes al estudio que se tomó como punto de partida [1].

Para decidir si el sistema implementado es de buena calidad nos hemos basado en las métricas obtenidas en 4.1, donde los resultados se sitúan en valores en torno al 95 % de detecciones en secuencias de una única persona, y e torno al 85 % en secuencias donde aparecen múltitud de personas en diferetes situaciones.

Estos datos obtenidos demuestran que el sistema implementado es capaz de detectar el número de personase en una escena con un porcentaje de fiabilidad bastate elevado.

5.2 Líneas de trabajo futuras

En este apartadoo se propondrán algunas de las líneas futuras de investigación y mejora del diseño expuesto en este trabajo.

- **Mejoras en la detección de máximos.** El empleo de una SR de tamaño fijo puede suponer variaciones en la detección de máximos con diferentes cámaras cuya resolución sea distinta a la resolución de la cámara empleada. Un sistema que emplee un tamaño de SR adaptativa en función de la altura a la que se sitúe la cámara y la resolución de la misma supondría un algoritmo exportable a diferentes escenas.
- **Empleo de descriptores 3D.** En este trabajo únicamente se extrae información acerca de la forma de la persona mediante la proyección de su altura en 2D. Empleando descriptores 3D se podría conseguir una mayor precisión en la detección de diferentes características de la persona.
- **Estudio del comportamiento con otras bases de datos.** En el desarrollo de este proyecto se ha empleado una única base datos sobre un entorno conocido y controlado. En futuros estudios se podrían realizar pruebas sobre diferentes bases de datos, incluyendo variaciones en el entorno y en la ubicación de la cámara.

Bibliografía

[1] C. Stahlschmidt, A. Gavriilidis, J. Velten, and A. Kummert, "Applications for a people detection and tracking algorithm using a time-of-flight camera," *Multimedia Tools and Applications*, pp. 1–18, 2014.

[2] J. Sell and P. O'Connor, "The xbox one system on a chip and kinect sensor," *IEEE Micro*, no. 2, pp. 44–53, 2014.

[3] D. Ramanan, D. Forsyth, A. Zisserman *et al.*, "Tracking people by learning their appearance," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 29, no. 1, pp. 65–81, 2007.

[4] D. Lefloch, F. A. Cheikh, J. Y. Hardeberg, P. Gouton, and R. Picot-Clemente, "Real-time people counting system using a single video camera," in *Electronic Imaging 2008.* International Society for Optics and Photonics, 2008, pp. 681 109–681 109.

[5] T.-Y. Chen, C.-H. Chen, D.-J. Wang, and Y.-L. Kuo, "A people counting system based on face-detection," in *Genetic and Evolutionary Computing (ICGEC), 2010 Fourth International Conference on.* IEEE, 2010, pp. 699–702.

[6] C. Y. Jeong, S. Choi, and S. W. Han, "A method for counting moving and stationary people by interest point classification," in *Image Processing (ICIP), 2013 20th IEEE International Conference on.* IEEE, 2013, pp. 4545–4548.

[7] A. B. Chan, Z.-S. J. Liang, and N. Vasconcelos, "Privacy preserving crowd monitoring: Counting people without people models or tracking," in *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on.* IEEE, 2008, pp. 1–7.

[8] A. Bevilacqua, L. D. Stefano, and P. Azzari, "People tracking using a time-of-flight depth sensor," in *Video and Signal Based Surveillance, 2006. AVSS'06. IEEE International Conference on.* IEEE, 2006, pp. 89–89.

[9] C. Stahlschmidt, A. Gavriilidis, J. Velten, and A. Kummert, "People detection and tracking from a top-view position using a time-of-flight camera," in *Multimedia Communications, Services and Security.* Springer, 2013, pp. 213–223.

[10] L. Jia and R. J. Radke, "Using time-of-flight measurements for privacy-preserving tracking in a smart room," *Industrial Informatics, IEEE Transactions on*, vol. 10, no. 1, pp. 689–696, 2014.

[11] R. Lange and P. Seitz, "Solid-state time-of-flight range camera," *Quantum Electronics, IEEE Journal of*, vol. 37, no. 3, pp. 390–397, 2001.

[12] D. Jiménez-Cabello, "Correction of errors in time of flight cameras," *Tesis Doctoral, Universidad de Alcalá de Henares*, 2015.

[13] A. Mohan, C. Papageorgiou, and T. Poggio, "Example-based object detection in images by components," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 23, no. 4, pp. 349–361, 2001.

[14] P. O. Gislason, J. A. Benediktsson, and J. R. Sveinsson, "Random forests for land cover classification," *Pattern Recognition Letters*, vol. 27, no. 4, pp. 294–300, 2006.

[15] "Información sobre gnu/linux en wikipedia," http://es.wikipedia.org/wiki/GNU/Linux [Último acceso 1/noviembre/2013].

[16] "Página de la aplicación emacs," http://savannah.gnu.org/projects/emacs/ [Último acceso 1/noviembre/2013].

[17] "Página de la aplicación kdevelop," http://www.kdevelop.org [Último acceso 1/noviembre/2013].

[18] L. Lamport, *LaTeX: A Document Preparation System, 2nd edition.* Addison Wesley Professional, 1994.

[19] "Página de la aplicación octave," http://www.octave.org [Último acceso 1/noviembre/2013].

[20] "Página de la aplicación cvs," http://savannah.nongnu.org/projects/cvs/ [Último acceso 1/noviembre/2013].

[21] "Página de la aplicación gcc," http://savannah.gnu.org/projects/gcc/ [Último acceso 1/noviembre/2013].

[22] "Página de la aplicación make," http://savannah.gnu.org/projects/make/ [Último acceso 1/noviembre/2013].

Apéndice A

Manual de usuario

Este manual de usuario presenta la interfaz de programación del software desarrollado en este proyecto. Todas las funciones aquí implementadas, han sido programadas en C/C++ sobre las bibliotecas *OpenCV 2.4.6.1*. Esto hace necesario una instalación previa de esta versión de OpenCV o de cualquier otra versión psterior compatible.

A.1 Direcctorios necesarios

Durante la ejecución del programa será necesaria la existencia de los directorios mostrados a continuación:

- Graficas: en este directorio se generarán las diferentes imágenes en formato *.png relativas a la representación de los vectores de cada clase.
- Resultados: en este directorio se generarán los ficheros de resultados en formato *.result de cada frame analizado. Dichos ficheros contendrán una estructura por frame del siguiente tipo.

  ```
  typedef struct
  {   int Frame;
      int PCAresult[maxDETECTIONS];
      Point position[maxDETECTIONS];
      int numMax;
  }st_resultSave;
  ```

  Los campos integrados en la estructura serán los siguientes:

  - `Frame`: índice del frame analizado dentro de la secuencia.
  - `PCAresult`: array binario con los resultados de la detección para cada uno de los máximos localizados. El número de máximas detecciones posibles queda definido por `maxDETECTIONS`.
  - `position`: array de puntos con posiciones de cada uno de los máximos localizados dentro de la escena.
  - `numMax`: número de maximos reales localizados
- ClasesPCA: dentro de este directorio se almacenarán los datos correspondientes a cada una de las clases implementadas en formato *.data. Dichos ficheros estarán formados por los vectores característicos extraidos de los sujetos de entrenamiento pertenecientes a cada clase.

- GroundTruth: dentro de este directorio se almacenarán los resultados de las secuecias etiquetadas en formato *.gt. Estos ficheros se leeran mediante la libreria de lectura de ficheros de resultados localizada dentro del repositorio GEINTRA *"far-field/sourceLocationResultsLib/sourcelocationresultslib.h"*

A.2 Algoritmo detección de personas

Los archivos fuentes necesarios para el funcionamiento del sistema se encuentran dentro del directorio **TOFCounting**. El sistema a implementar consta de una única función raíz implementada en el archivo **DetectorTOF** la cual analizará el número de personas existentes en una única imagen. Para el análisis de una secuencia entera, será necesaria un etapa previa que desglose cada una de las secuencias de la base de datos en frames individuales. La función raiz consta de la siguiente declaración:

```
int TOFdetector(int16_t *z2D16, int cont_frames, char *sequenceName)
```

La función `TOFdetector` recibe como parámetros:

- `z2D16`: array de tipo $int16_t$ correspondiente a la imagen de pronfundidad extraida del sensor. El tamaño del dicho array vendrá determinado por la resolución de la cámara a utilizar, en nuestro caso dado que se tratan de imagenes con resolución 512X424 el array tendrá un tamaño $size = 217088$.
- $cont\_frames$: Para establecer una relación entre frames el caso de introducir una secuencia completa, esta variable indicará el índice del frame a analizar dentro de la misma.
- `sequenceName`: nombre de la secuencia a analizar.

Dicha función msotrará por pantalla, si así se quisiese, activando el campo $SHOW\_RESULT$, una imagen indicativa de las personas detectadas mediante un punto azul sobre cada una de ellas.

A.3 Algoritmo evaluación de resultados

Los archivos fuentes necesarios para el funcionamiento del sistema de evaluación de resultados se encuentran dentro del directorio **TOFScoring**. Este programa se arranca mediante la orden, sobre la terminal de comandos de linux, **./tofCountingScoring**, siendo necesarios el uso de dos argumentos

- Primer argumento: fichero *.gt correspondiente al etiquetado de la secuencia a evaluar, en el caso de no existir dicho fichero incluyendo el comando '0' se realizará el analisis estadístico en función de los máximos localizados y los máximos evaluados como personas.
- Segundo argumento: fichero *.result situado en la carpeta Resultados correspondiente a la secuencia a analizar.

Dicha función imprimirá por pantalla el número de frames analizados, número de personas total analizadas, porcentaje de verdaderos positivos, falsos pisitivos y falsos negativos de la secuencia completa.

Apéndice B

Pliego de condiciones

Para la correcta utilizació del sistema desarrollado en este trabajo se debe disponer de un hardware y un software que cumpla unos requisitos mínimos.

B.1 Requisitos de Hardware

- Procesador de 32/64 bits
- 1GB de memoria RAM o superior
- Al menos 100MB de memoria libres en el disco duro para funciones y datos.

B.2 Requisitos de Software

- Sistema operativo Linux Ubuntu 14.04.1 LTS
- Libreria *OpenCV 2.4.6.1*
- Al menos 100MB de memoria libres en el disco duro para funciones y datos.
- Compilador GNU GCC

Apéndice C

Presupuesto

C.1 Costes de equipamiento

- Equipamento Hardware utilizado:

Concepto	Cantidad	Coste Unitario	Subtotal(€)
PC Acer Core I3	1	500€	500€
Kinect 2	1	200€	200€
Coste total HW			500€

Tabla C.1: Coste equipamiento Hardware utilizado

- Recursos Software utilizados:

Concepto	Cantidad	Coste Unitario	Subtotal(€)
Ubuntu 14.04.1 LTS	1	0€	0€
Libreria OpenCv 2.4.6.1	1	0€	0€
Software LaTeX	1	0€	0€
Coste total SW			0€

Tabla C.2: Coste equipamiento Software utilizado

C.2 Costes de mano de obra

Concepto	Cantidad	Coste Unitario	Subtotal(€)
Desarrollo SW	200	65€/hora	13000€
Mecanografiado del documento	100	15€/hora	1500€
Coste total mano de obra			14500€

Tabla C.3: Coste debido a mano de obra

C.3 Costes total del presupuesto

Concepto	Subtotal(€)
Equipamiento Hardware	700€
Recursos Software	0€
Mano de obra	14500€
Coste total presupuesto	15200€

Tabla C.4: Coste total del presupuesto

El importe total del presupuesto asciende a la cantidad de: QUINCEMIL DOSCIENTOS EUROS

En Alcalá de Henares a ___ de _____ de 20___

Raquel García Jiménez.

Universidad de Alcalá
Escuela Politécnica Superior

ESCUELA POLITECNICA
SUPERIOR

Universidad
de Alcalá