



Universidad de  
Oviedo



# **ESCUELA POLITÉCNICA DE INGENIERÍA DE GIJÓN.**

## **GRADO EN INGENIERÍA ELECTRÓNICA INDUSTRIAL Y AUTOMÁTICA**

**ÁREA DE INGENIERÍA DE SISTEMAS Y AUTOMÁTICA**

**TRABAJO FIN DE GRADO N.º 17010426**

**SISTEMA DE DETECCIÓN DE PASOS DE PEATONES PARA  
AYUDA A INVIDENTES MEDIANTE VISIÓN ARTIFICIAL**

**D. ALEJANDRO DÍAZ ROSALES  
TUTOR: D. RAFAEL CORSINO GONZALEZ DE LOS REYES  
COTUTOR: D. JOSÉ ANTONIO CANCELAS CASO**

**FECHA: julio 2017**



# Índice.

MEMORIA.....	4
1.- Introducción.....	6
2.- Marco teórico.....	14
3.- Experimentos y resultados.....	30
4.- Discusión. ....	45
5.- Conclusión.....	48
6.- Referencias. ....	49
ANEXOS.....	53
1.- Parámetros para el procesamiento de las imágenes.....	55
2.- Tablas de resultados.....	56
3.- Normativa de los pasos peatonales.....	59
4.- Materiales adjuntados.....	61
PRESUPUESTO.....	63
1.- Coste de ejecución material.....	65
2.- Gastos generales y beneficio industrial.....	68
3.- Importe total. ....	69

**MEMORIA**

# Índice de la memoria.

1.-	Introducción.....	6
1.1.-	Objetivos.....	6
1.2.-	Estado del arte.....	6
2.-	Marco teórico.....	14
2.1.-	Detección de los pasos de peatones.....	14
2.1.1.-	Segmentación bipolar.....	14
2.1.2.-	Transformada de Hough probabilística.....	18
2.1.3.-	Invarianza proyectiva.....	19
2.2.-	Medida de distancias.....	23
2.2.1.-	Modelo de una cámara.....	23
2.2.2.-	Imágenes estéreo.....	26
3.-	Experimentos y resultados.....	30
3.1.-	Detección de los pasos de peatones.....	30
3.2.-	Medidas estéreo.....	38
4.-	Discusión.....	45
5.-	Conclusión.....	48
6.-	Referencias.....	49

# 1.- Introducción.

## 1.1.- OBJETIVOS.

Actualmente las intersecciones urbanas, específicamente los pasos de peatones tipo cebra son zonas en las que, ante la ausencia de un sistema de ayuda acústico para personas con problemas de visión y el limitado rango de detección que proporcionan los bastones blancos, se pueden generar situaciones de riesgo, causadas por la pérdida de orientación que pueden sufrir estas personas al querer cruzar por un paso de peatones.

La mayoría de estos peatones tienen ciertas facilidades para la orientación en función de referencias o marcas en el suelo y paredes, no obstante, es muy difícil para ellos alinearse precisamente con cada paso de peatones y saber cuánto les queda para terminarlo. Una colocación adecuada les permitirá entrar en él en la dirección correcta y evitar el peligro de extravío por la carretera, concediéndoles a su vez una mayor libertad y una menor dependencia de otras personas. Para lograr esto, el problema se abordará mediante la visión por computador, entre sus objetivos está modelar matemáticamente los procesos de percepción visual, dando así la posibilidad de obtener información de una serie de imágenes del entorno, mediante un programa realizado de manera específica al problema.

Este proyecto tiene como objetivo evaluar la viabilidad de un sistema basado en visión por computador que permita la localización de pasos de peatones en la escena y, eventualmente, guiar al usuario hacia los mismos.

## 1.2.- ESTADO DEL ARTE.

En este campo se han desarrollado varios métodos tanto para la ayuda de invidentes como para la conducción autónoma de vehículos. Algunos de estos se centran en detectar la forma y las bandas que conforman los pasos de peatones tipo cebra, Zhai, Cui, Gu y Kong en su publicación [1] proponen un método basado en MSER (*Maximally Stable Extremal Regions*), que consiste en encontrar qué puntos de una imagen corresponden con los puntos de otra tomada desde una perspectiva distinta, con lo que logran obtener

características del paso de peatones. A continuación, se haría uso de ERANSAC (*Extended Random Sample Consensus*) y ya se detectaría el paso de peatones. Antes de esto realizan una extracción temporal del fondo de la imagen (TMBE) logrando eliminar todos aquellos objetos en movimiento como coches y personas con el uso de varias imágenes consecutivas, esto solo es útil si la cámara está fija en una posición, algo incompatible a lo que se busca resolver en este proyecto, ya que el invidente busca desplazarse.

Otra opción es el uso de redes neuronales convolucionales (CNN) tal y como proponen Radványi, Varga y Karacs [2], quienes lo primero que buscan es detectar la superficie del camino, eliminando todos los objetos que puedan dificultar la posterior detección. Para la extracción del paso de peatones desarrollan el algoritmo de la red neuronal de forma manual, seleccionando ellos mismo los filtros o plantillas que se aplican. En otra de sus publicaciones [3] comparan este desarrollo manual con uno automático, el cual se realiza entrenando al sistema con distintas muestras y especificando cual debería ser el resultado. El algoritmo se crea buscando que plantillas hacen que a partir de una entrada se obtenga la salida solicitada, el sistema solo puede usar las plantillas [4] que defina el diseñador en una base de datos. Tanto el método manual como el automático muestran resultados similares.

Una solución más simple sería empezar con un binarizado de la imagen, consistente en tomar todos aquellos píxeles de la imagen con una intensidad mayor a un valor umbral y ponerlos a 1, el resto se ponen a 0, quedando una imagen binaria en blanco y negro. Después con una serie de plantillas predefinidas, se realiza el proceso conocido como *template matching* [5]. Las plantillas son imágenes binarias en las que hay bandas blancas sobre un fondo negro, emulando lo que se tiene en los pasos de peatones. El resultado es una imagen donde las zonas de la misma más parecidas a la plantilla son más claras, coincidiendo el punto de mayor valor en la imagen con un paso de peatones. Esta es la propuesta publicada por Asami y Ohnishi [6].

Para ayudar a personas con problemas de visibilidad también es útil el poder informarles sobre las características geométricas del paso de peatones. Esto es algo que abordaron Shioyama, Wu, Nakamura y Kitawaki [7], quienes realizaron un proyecto que

busca medir las longitudes reales de los pasos de peatones y detectar las luces de tráfico. Empiezan por un proceso de binarizado y un detector de contornos, obteniendo así las cuatro esquinas del paso de peatones, que junto con un modelo basado en las proyecciones y perspectivas estiman las diferentes longitudes de este (figura 1.1). El problema que surge con esta propuesta es que no funciona si la imagen no se captura de frente a un paso de peatones en la dirección en la que este se cruza.

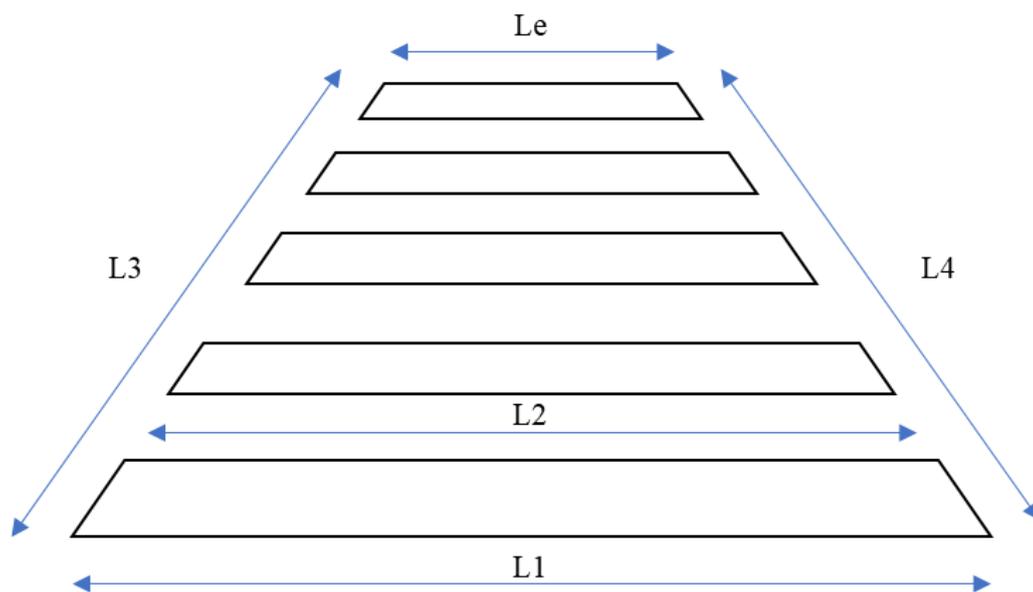


Figura 1.1. Modelo de un paso de peatones y las longitudes que obtienen de él en [7]. Basado en [7]

Para resolver el problema, en este documento se plantean dos etapas fundamentales, tal y como se muestra en la figura 1.2. La primera consiste en identificar el paso de peatones, del cual se espera obtener una imagen donde solo se muestren los mismos si es que existen. Mientras que en la última fase es en la que se espera determinar la posición del paso de peatones y la distancia a la que se encuentra.

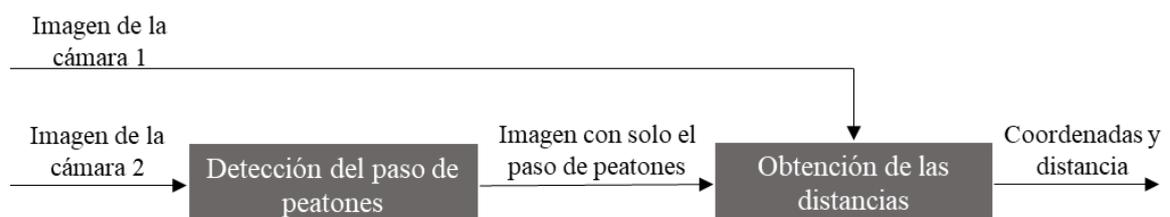


Figura 1.2. Esquematización global en etapas. El uso de dos cámaras para las imágenes de entrada y el que en una de ellas no se detecte el paso de peatones, se explica más a delante.

El primer paso va a consistir en detectar posibles candidatos a ser paso de peatones. Uddin y Shioyama [8] proponen una forma para detectarlos partiendo de una imagen en escala de grises. Para segmentar la imagen se basan en la bipolaridad que tiene la imagen en las zonas donde se encuentran los pasos, lo que genera candidatos a pasos de peatones. El método consiste en dividir la imagen en regiones con la misma distribución de intensidades y quedarse solo con aquellas que únicamente tienen píxeles muy oscuros y muy claros, ambos en la misma proporción.

Para verificar que los candidatos obtenidos son válidos, hacen uso de la invarianza geométrica basándose en lo desarrollado por Weiss [9], quien define la invarianza como una función derivada de unos descriptores de forma, ya sean locales o globales, los cuales no varían ante una transformación de coordenadas. En el caso de una imagen se mantendrían constantes ante cambios de perspectiva.

En España los pasos de peatones están normalizados, siendo el grosor de las bandas un valor que se mantiene siempre constante. La Fundación ONCE [10] [11] expone, que la señalización del paso de peatones consiste en bandas paralelas a la acera, pintadas sobre la calzada de 0,50 m de ancho y separados entre sí una distancia de 0,50 m. Por lo tanto, ante la presencia de pasos de peatones, como estos son bandas repetidas periódicamente de forma equidistante, si se conocen los puntos que se muestran en la figura 1.3, ante cualquier posición en el espacio en la que sea tomada una imagen con candidatos, sería posible verificar si estos pertenecen o no un paso de peatones.

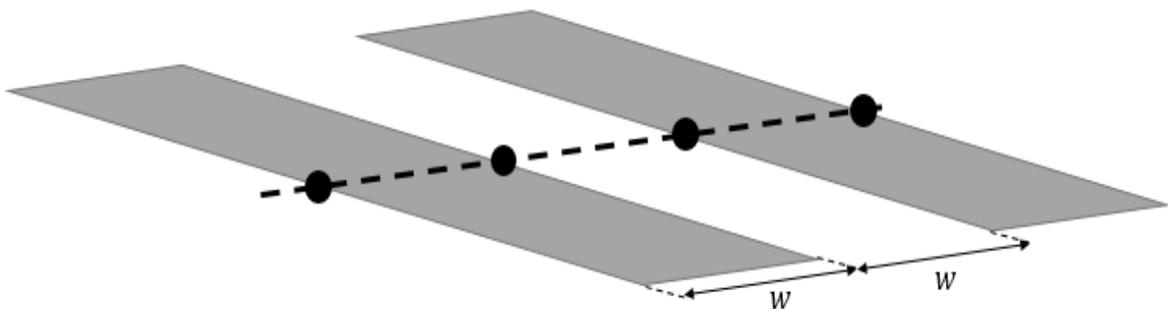


Figura 1.3. Cuatro puntos colineales en un paso de peatones, separadas entre sí una distancia  $w$ , Basado en [8].

Los pasos que siguen para obtener los puntos característicos empiezan por calcular la dirección del paso de peatones. El ángulo corresponde con la orientación de cada una de

las bandas que lo componen. Para calcularlo lo definen mediante una expresión matemática, la cual se transforma teniendo en cuenta que la señal es discreta y haciendo uso de la relación de Parseval. Como resultado queda una función dependiente de una transformada de Fourier de la imagen y un ángulo. El análisis de Fourier de una señal permite determinar sus frecuencias, en el caso de las imágenes las señales corresponden a los niveles de gris o intensidad de las distintas filas o columnas. Aquel ángulo que para una misma imagen maximice Fourier es el correspondiente al paso de peatones de esa imagen.

Una vez tienen el ángulo proponen eliminar como candidatos aquellos que lo tengan mayor a quince grados, para poder aplicar un detector de contornos local basado en el criterio de Fisher y obtener así los puntos característicos del paso de peatones. Se crea una ventana vertical, de 1 por un número impar mayor que 1, que se desplaza de arriba hacia abajo en la imagen. Para aplicar este criterio se definen dos clases o muestra, una es la intensidad de los píxeles en la parte superior de la ventana y la otra en la parte inferior, dejando el centro de la ventana fuera de las clases. El criterio de Fisher se define como la relación entre la varianza entre clases y la varianza dentro de las clases, ambas se pueden calcular mediante una fórmula. Esta relación será máxima cuando la clase de la parte superior de la ventana esté sobre una banda blanca y la clase de la parte inferior de la ventana esté sobre el asfalto, o viceversa, siendo en ese momento cuando la posición del punto central de la ventana coincide con un punto característico.

Para este proyecto interesa poder detectar un paso de peatones con cualquier orientación, por lo que no se eliminarían aquellos candidatos que tengan un ángulo mayor a quince grados, por otra parte, el realizar la transformada de Fourier muchas veces tiene un coste computacional muy alto. Por lo que para obtener los puntos característicos se usará una variante de la transformada de Hough desarrollada por Galambos, Matas y Kittler [12] llamada, transformada de Hough probabilística.

La transformada de Hough [13] es un algoritmo basado en una técnica de votación que permite identificar puntos que pertenecen a formas regulares básicas. En este caso una recta, que se puede describir usando coordenadas polares mediante dos parámetros, un ángulo y una distancia. Se inicia definiendo un acumulador de dimensión igual al número

de parámetros, es decir dos, a esto se le conoce como *Hough Acumulator Array*. Cada pixel de una imagen de contornos va a votar por una línea compatible, es decir votará por todas las líneas que puedan pasar a través de él. Varios puntos en una misma línea, alineados entre si, se traducen en un mayor número de votos para una recta en concreto. Por lo que después de que todos los píxeles de la imagen voten se buscan aquellas líneas que obtuviesen más votos en el acumulador.

La transformada de Hough probabilística maneja el acumulador de votos de una manera más eficiente, desde el punto de vista del coste computacional. Se seleccionan los píxeles de manera aleatoria y cuando el número de votos para una de las rectas supera un valor, esta pasa a considerarse real y se eliminan aquellos píxeles que votaron por ella. Este método se ve menos afectado por el ruido. Una vez se tengan las rectas se pueden obtener las coordenadas de los puntos característico del paso de peatones, los cuales son un punto por cada segmento, estando los puntos alienados entre sí.

Finalmente se determina si los candidatos son pasos de peatones por invarianza geométrica, obteniendo una imagen en la que solo se ve la zona en la que se encuentra el paso De esta forma la segunda etapa queda dividida como se muestra en la Figura 1.4.

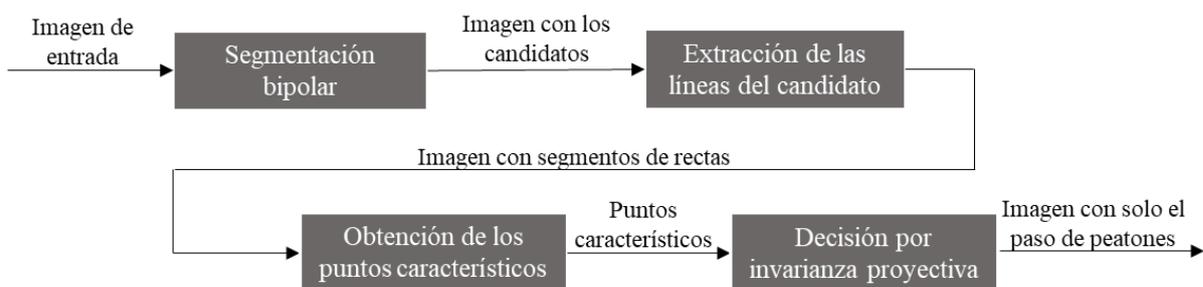


Figura 1.4. Esquematización global de la primera etapa

La segunda y última etapa consiste en determinar el punto más cercano del paso de peatones a la cámara y la distancia a la que este se encuentra. Para poder obtener la localización tridimensional de los objetos presentes en la escena es necesario el uso de dos cámaras en un montaje estéreo, o de una secuencia de imágenes hechas con la misma cámara, pero en posiciones distintas, es decir dos imágenes (cámaras) dispuestas de forma que la posición y orientación de una sobre la otra sea conocida o se pueda determinar.

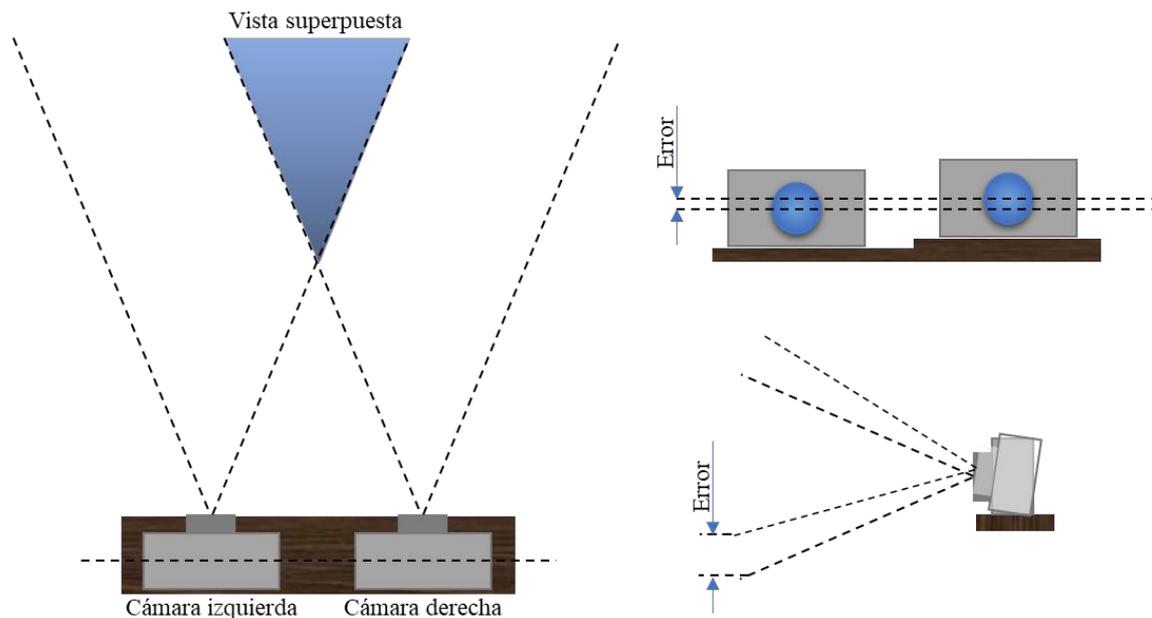


Figura 1.5. Cámaras estéreo alineadas en paralelo y los posibles errores verticales que pueden aparecer. Solo los objetos dentro de la zona de superposición pueden ser medidos. Basado en [15].

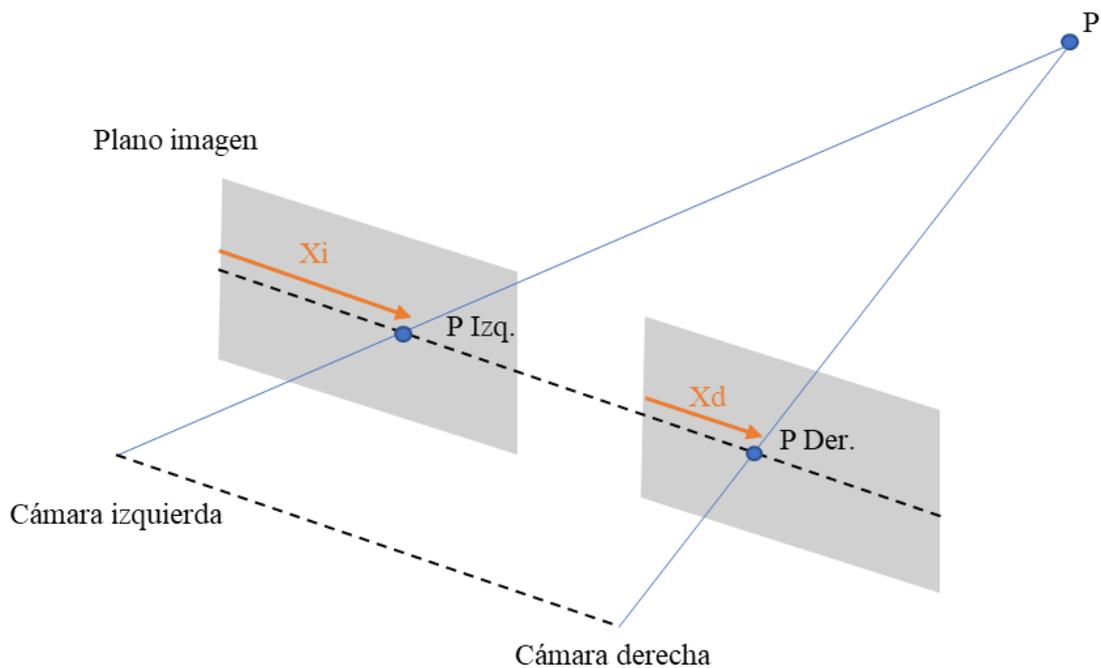


Figura 1.6. La diferencia entre la distancia  $X_i$  y  $X_d$  en píxeles es lo que se conoce como la disparidad para el punto  $P$

En su publicación [14] Yasir, Rahizall, Hasbullah y Amelia realizan un procesado de imágenes en tiempo real para calcular la distancia a la que se encuentra un objeto y medir su tamaño. Las imágenes estéreo las capturan con dos cámaras fijas y alineadas en paralelo sobre una base tal y como se muestra en la figura 1.5. Empiezan con la detección

del objeto en las dos imágenes, para después calcular su disparidad, la cual es el desplazamiento de un punto entre dos imágenes, como se puede apreciar en la figura 1.6. Conociendo la distancia focal, parámetro interno de las cámaras, y la separación entre las mismas se puede calcular esta distancia al objeto.

En este proyecto, debido a la aleatoriedad de la transformada de Hough probabilística que se usa en el proceso de detección del paso de peatones, cada vez que se procesa una imagen los resultados obtenidos pueden variar levemente; es decir que lo obtenido tras procesar la imagen de la derecha no tiene que ser análogo a lo obtenido en la de la izquierda. Por lo tanto, en vez de detectar el punto más cercano del paso de peatones en ambas imágenes, se hará solo en una y mediante el proceso de *template matching* [5] se obtendrán las coordenadas de este mismo punto en la otra imagen. Tal y como hicieron Jernej y Damir [15], quienes igualmente querían determinar la distancia a la que se encuentra un objeto con el uso de una cámara estéreo. El conjunto de estos pasos constituye la última etapa, los cuales se muestran en la figura 1.7.

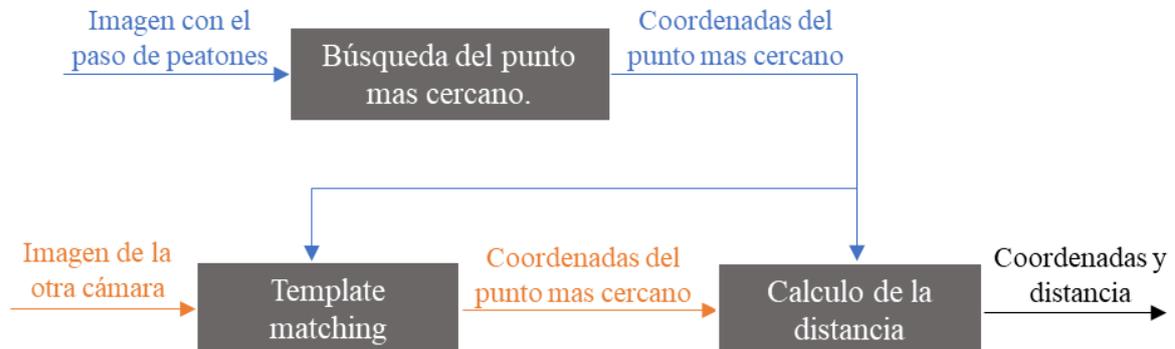


Figura 1.7. Esquematización global de la segunda etapa. En azul se muestran los datos correspondientes a las imágenes de la cámara izquierda y en anaranjado las de la derecha, o viceversa.

## 2.- Marco teórico.

### 2.1.- DETECCIÓN DE LOS PASOS DE PEATONES.

La primera etapa de esta propuesta (figura 1.4) busca identificar los diferentes pasos de peatones tipo cebrá que puedan existir en la escena. Como ya se ha mencionado, el método propuesto está basado en el trabajo realizado por Uddin y Shioyama [8]. Se mantienen los fundamentos, la segmentación bipolar y el uso de invarianzas geométricas; pero, se modifican los procesos intermedios teniendo en cuenta las características del problema a resolver.

#### 2.1.1.- Segmentación bipolar.

Lo que se busca con esta segmentación es obtener regiones de la imagen candidatas a contener pasos de peatones. En una situación ideal, este se caracterizaría por ser un conjunto de bandas blancas sobre un fondo de color negro, cubriendo cada uno el 50% de la superficie del plano. Sin embargo, En la mayoría de los casos reales, ya sea debido al desgaste de los materiales o cualquier otro factor externo, el pavimento en vez de ser negro, puede ser de cualquier otra tonalidad de gris oscuro y las bandas de gris claro.

Por lo tanto, lo que se busca con este método es identificar aquellas zonas de la imagen que poseen una cantidad equitativa de píxeles muy claros y muy oscuros. El primer paso consiste en separar la imagen en regiones, para después decidir que regiones cumplen las condiciones para ser paso de peatones. Independiente de que sean dos pasos distintos, ambos se basan en el análisis de la distribución de la intensidad en escala de grises, de una fracción de la imagen llamada bloque. Esta distribución se define en como:

$$p_0(x) = \alpha p_1(x) + (1 - \alpha)p_2(x) \quad (2.1)$$

Siendo  $p_1$  la distribución de la intensidad de los píxeles oscuros y  $p_2$  la de los píxeles claros, siendo  $\alpha$  el valor que define qué porcentaje de píxeles del bloque  $p_0$  corresponde tanto a  $p_1$  como  $p_2$ , puede tomar valores de cero a uno. Se sabe que la media y la varianza se definen como:

$$\mu_i = \int_{-\infty}^{\infty} x p_i(x) dx \tag{2.2}$$

$$\sigma_i^2 = \int_{-\infty}^{\infty} (x - \mu_i)^2 p_i(x) dx \tag{2.3}$$

El desarrollo de la expresión (2.2) junto con la (2.1) da como resultado que la media de un bloque es igual a:

$$\mu_0 = \alpha\mu_1 + (1 - \alpha)\mu_2 \tag{2.4}$$

Finalmente desarrollando la relación (2.3) con todas las anteriores se obtiene la siguiente expresión para definir la varianza de un bloque.

$$\sigma_0^2 = \alpha\sigma_1 + (1 - \alpha)\sigma_2 + \alpha(1 - \alpha)(\mu_1 - \mu_2)^2 \tag{2.5}$$

Una imagen bipolar es aquella en la que hay dos intensidades bien diferenciadas, específicamente valores claros y oscuros. Si  $\alpha$  es igual 0,5 una imagen completamente bipolar se da si  $\sigma_1$  y  $\sigma_2$  tienden a 0, tal y como se ve en la figura 2.1. Por lo tanto, la varianza del bloque siendo está prácticamente bipolar quedaría como:

$$\sigma_0^2 = \alpha(1 - \alpha)(\mu_1 - \mu_2)^2 \tag{2.6}$$

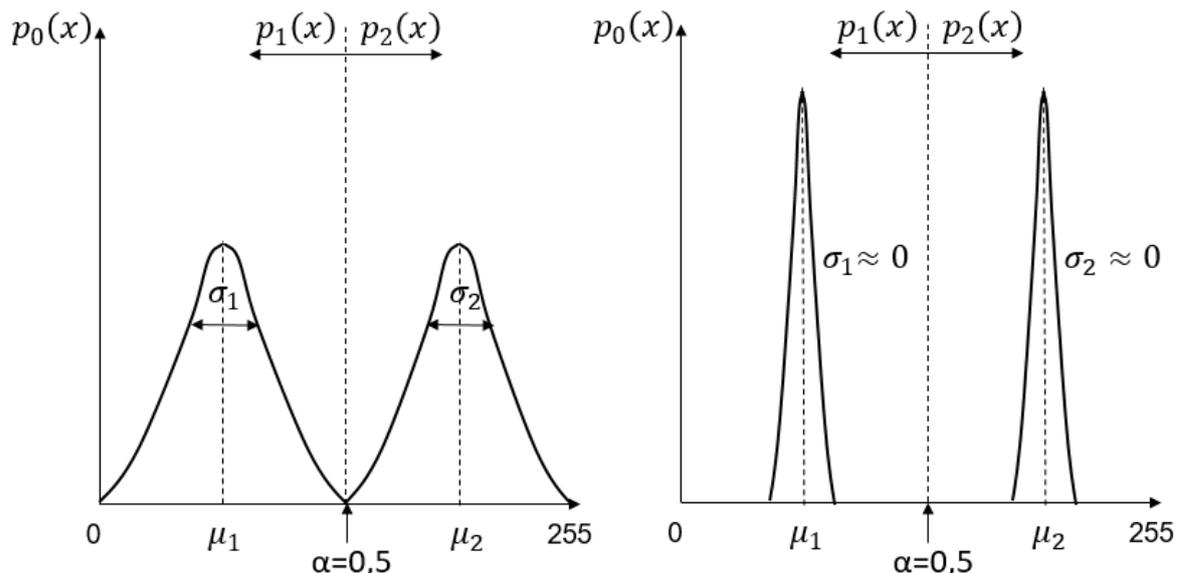


Figura 2.1. En la izquierda (Basada en una imagen de [8]) hay una imagen de una distribución bipolar cualquiera y en la de la derecha se tiene una imagen completamente bipolar.

En base a esto la bipolaridad  $\gamma$  se definirá como la relación entre la ecuación (2.6) y la varianza del bloque. Si una imagen es completamente bipolar,  $\gamma$  en la ecuación (2.7) es igual a 1, ya que el valor del denominador será igual al numerador. Pero en el caso de que  $\gamma$  sea 0 significara que la imagen no es bipolar (figura 2.2).

$$\gamma = \frac{\alpha(1-\alpha)(\mu_1-\mu_2)^2}{\sigma_0^2} \tag{2.7}$$

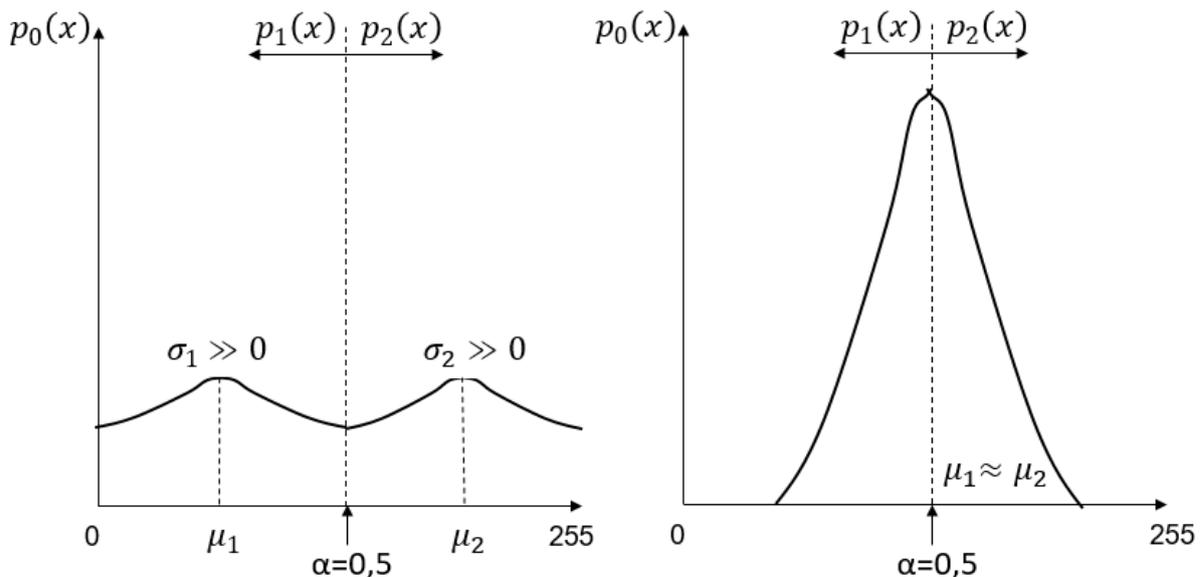


Figura 2.2. Ambas distribuciones son los casos en los cuales la bipolaridad es 0. En la de la izquierda se ve que mientras más grandes sean las varianzas más pequeñas son los picos y más distribuidos están los puntos de la imagen. En la distribución de la derecha todos los valores tienden al mismo valor de gris.

De cara a la aplicación de esta fórmula es conveniente poder modificar y ajustar  $\alpha$ , ya que en una imagen puede aparecer un paso de peatones en la cual, las bandas estén más cerca del valor intermedio de intensidad que del 0 o lo análogo en el caso del pavimento. Un escenario probable es el que las bandas sean aproximadamente grises (valor medio) y el pavimento muy oscuro, con un  $\alpha$  de 0,5 el resultado de la formula (2.7) puede estar muy lejano del 1 y considerarse la zona como no bipolar, pero si  $\alpha$  es mayor la bipolaridad daría como resultado un número próximo a 1. Este valor se puede definir experimentalmente de forma que se ajuste al tipo de paso de peatones que más se repita.

Como se dijo antes, este método empieza separando la imagen en regiones, lo cual se logra dividiendo la imagen en pequeños bloques, para después unir aquellos que son

similares y agruparlos en una misma región. Sus distribuciones de intensidades de grises se comparan definiendo la relación entre las distancias de sus medias (figura 2.3) como:

$$\text{Relación de distancias} = \frac{\text{Distancia interna}}{\text{Distancia externa}} = \frac{\min(\mu_2, \tilde{\mu}_2) - \max(\mu_1, \tilde{\mu}_1)}{\max(\mu_2, \tilde{\mu}_2) - \min(\mu_1, \tilde{\mu}_1)} \quad (2.8)$$

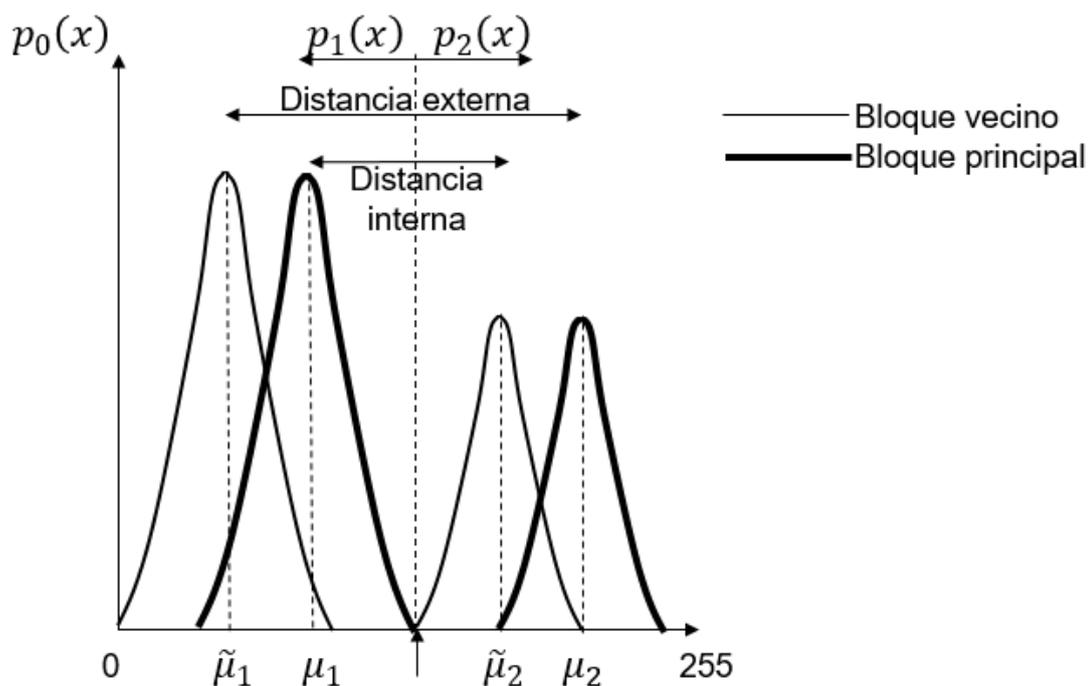


Figura 2.3. En la figura se muestra el significado de la fórmula que compara dos bloques vecinos. Basada en [8].

Siendo  $\mu_1$  y  $\mu_2$  las medias de un bloque y,  $\tilde{\mu}_1$  y  $\tilde{\mu}_2$  las medias de su bloque vecino con el que se está comparando. Si esta relación es mayor a un valor limite el cual se obtiene de forma de experimental, los bloques que se estén comparando se considerarán iguales, y por lo tanto de la misma región.

Una vez que se tienen todas estas regiones se evalúa cada una con la formula (2.7), convirtiéndose en candidatas a poseer un paso de peatones aquellas que tengan una bipolaridad mayor a un valor umbral. Este valor se define en función de las muestras recolectadas durante la fase experimental.

### 2.1.2.- Transformada de Hough probabilística.

Las regiones bipolares están formadas por pequeños bloques con la misma distribución de grises, por lo que aquellas que contengan un paso de peatones están compuestas en su mayoría por la superficie del cruce; es decir, solo tienen el pavimento y las bandas blancas. En estas regiones salvo figuras pintadas en el asfalto o aceras de un color parecido a este, las únicas rectas presentes son las del paso de peatones. La transformada de Hough probabilística [12] es un método que permitiría identificar estas rectas.

El objetivo de este método es minimizar el consumo computacional de la transformada de Hough estándar. El sistema de votación es el mismo que en el estándar, cada punto de los bordes vota por una línea compatible; es decir, vota por cada línea que pueda pasar a través de él. Estos votos se almacenan en un acumulador matricial, de dos dimensiones, relacionado con los parámetros que definen una recta en coordenadas polares. Las rectas principales o reales son aquellas que reciben un mayor número de votos, ya que hay más puntos en la imagen alineados con ella. Teniendo una imagen con  $N$  píxeles los pasos que componen este proceso son los siguientes:

- I. Se verifica si la imagen de entrada está vacía, si es así se finaliza el algoritmo.
- II. Se escoge un píxel de manera aleatoria en la imagen y se almacena su voto en el contenedor o acumulador.
- III. Se elimina el píxel de la imagen.
- IV. Se busca el pico más alto del acumulador de votos, si éste no supera un valor umbral se vuelve al paso I. Este valor define el número de votos mínimos para considerar a una recta como real.
- V. El pico del acumulador representa un camino o línea infinita, en el cual se busca el segmento de píxeles más largo que contenga. Puede ser que los píxeles estén uno detrás de otro o que estén separados una brecha máxima definida por un valor límite (figura 2.4).

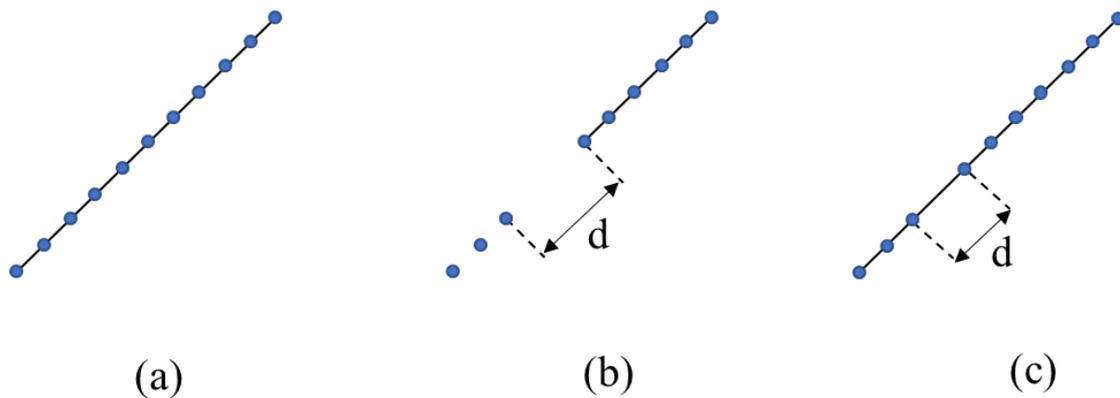


Figura 2.4. Comportamiento del algoritmo al definir un segmento encontrado: (a) los píxeles están uno detrás de otro por lo que el segmento se define sobre todos ellos, (b) la distancia  $d$  entre dos de los píxeles es mayor que el valor máximo permitido para considerar dos puntos de un mismo segmento, por esta razón se define como segmento el más grande, en este caso el superior, (c) en este caso la distancia máxima no es superada por lo que todos los puntos se consideran dentro del segmento.

- VI. Se guardan y eliminan todos los píxeles pertenecientes al segmento, tanto en la imagen como sus votos del acumulador.
- VII. Si la línea que se obtuvo es mayor que la longitud mínima admisible, se define a ese conjunto de píxeles como una línea.
- VIII. Se vuelve al paso I.

Los candidatos que salen de la segmentación bipolar no pueden pasar directamente por este proceso. Se ha de realizar un umbralizado [16] para tener una imagen binaria en la cual las bandas del paso de cebras estén a 1 y el resto de la imagen a 0. Para finalizar con un detector de contornos y obtener una imagen que solo tenga los bordes de las bandas, a la que ya puede aplicársele la transformada de Hough probabilística. Se usa el detector de contornos de Sobel [17].

### 2.1.3.- Invarianza proyectiva.

El proceso para obtener los puntos característicos a partir de las rectas se explica más adelante al no requerir fundamentos teóricos. El último paso consiste en comprobar si se está ante un paso de peatones o no, para eso se hace uso de la publicación hecha por Weiss [9]. En ella realiza un estudio del concepto general de la invarianza para reconocer objetos.

A veces, dos imágenes de un mismo objeto pueden diferir entre sí, debido a los diferentes puntos en el espacio en que estas pueden ser tomadas, dificultando su detección. Las invarianzas geométricas son una buena solución a este problema, ya que son descriptores de forma que se mantienen constantes bajo cualquier transformación geométrica como un cambio de perspectiva. Estos descriptores de forma se obtienen a partir de la geometría del objeto a reconocer.

Existen dos maneras de representar formas, una es la aproximación implícita en la cual se definen como una relación entre coordenadas  $X_i$  (Descriptor local), donde  $a_k$  son los coeficientes característicos de la forma, los cuales son descriptores globales. En la aproximación explícita las coordenadas de los puntos de una forma son funciones del parámetro de una curva local  $t$  o superficie, las cuales al derivarla dan los descriptores de forma locales.

$$f(a_k, X_i) = 0 \quad , \quad X_i = X_i(t) \quad (2.9) \quad (2.10)$$

Una invarianza es la derivada de una función dependiente de descriptores globales o locales, los cuales no varían ante transformaciones de las coordenadas  $X_i$  y el parámetro  $t$ . Se define una invarianza relativa  $I$  de peso  $w$  y en función de los descriptores de forma como:

$$\tilde{I} = J^{-w} I \quad (2.11)$$

Donde  $\tilde{I}$  es el nuevo sistema y  $J$  el jacobiano de la transformación,  $w$  depende del tipo de transformación. Se consigue un cambio similar de sistemas al multiplicar las coordenadas homogéneas por un factor.

$$\tilde{I} = \lambda^d I \quad (2.12)$$

Siendo  $d$  el grado de invarianza. Tanto el jacobiano como  $\lambda$  de un punto a otro, dependen de  $X_i$  y  $t$ , pero no dependen de los descriptores de forma local  $a_i$  y global  $\partial X_i^n / \partial t^n$ .

Una transformación proyectiva sobre un plano se puede escribir de la siguiente manera donde  $T$  es una matriz 3x3 con ocho parámetros significativos, en la cual  $trans_x$  y

$trans_y$ , son elementos de traslación,  $proj_1$  y  $proj_2$  representan la inclinación y el punto de vista, los cuales no son transformaciones líneas, y por ultimo  $aff_i$  representa la rotación den las direcciones x e y.

$$\begin{pmatrix} \tilde{x} \\ \tilde{y} \\ 1 \end{pmatrix} = \frac{1}{xT_{31}+yT_{32}+T_{33}} T \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \tag{2.13}$$

$$T = \begin{pmatrix} aff_1 & aff_2 & trans_x \\ aff_3 & aff_4 & trans_y \\ proj_1 & proj_2 & 1 \end{pmatrix} \tag{2.14}$$

La proyectividad se puede traducir o descomponer en una perspectiva más una traslación. Por lo tanto, si una proyección tiene un punto invariante esta se puede representar como una perspectiva geométrica. Un ejemplo seria la intersección de dos rectas siendo este el punto invariante (figura 2.5).

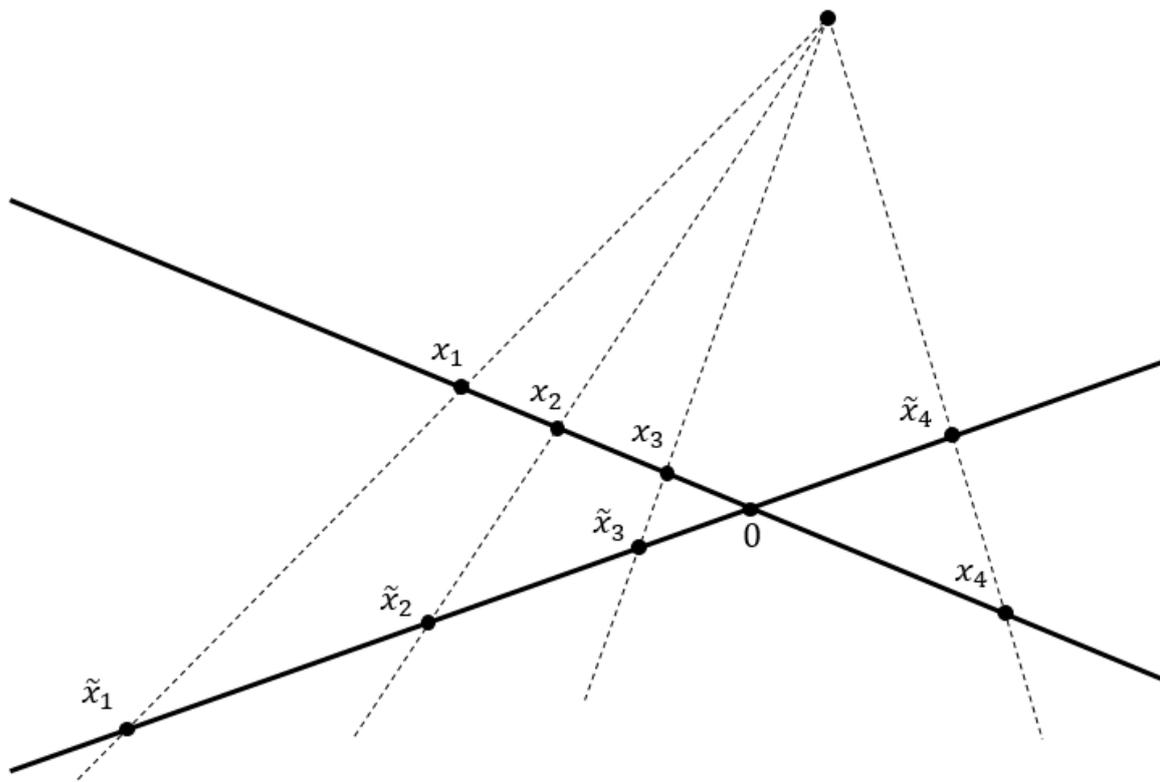


Figura 2.5. Perspectiva, la relación cruzada de 4 puntos se mantiene. El punto 0 es un punto invariante [9].

En el caso de que se tengan líneas paralelas el punto invariante pasa a estar en el infinito, a lo que se le conoce como subgrupo afín. Ante una afinidad los elementos  $proj_i$  de la matriz T desaparecen por lo que ahora la transformación proyectiva es lineal.

$$\begin{pmatrix} \tilde{x} \\ \tilde{y} \end{pmatrix} = \begin{pmatrix} aff_1 & aff_2 \\ aff_3 & aff_4 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} trans_x \\ trans_y \end{pmatrix} \quad (2.15)$$

Con el uso de coordenadas homogéneas se puede arreglar la no linealidad presente en el denominador de la ecuación (2.13), quedando la transformación proyectiva en estas coordenadas como se muestra a continuación, donde  $\lambda$  es un factor arbitrario.

$$\tilde{x} = \lambda_x T x \quad (2.16)$$

En la publicación definen muchos métodos para definir invarianzas, aquí se usará el método determinista. Muchas entidades geométricas pueden ser definidas en forma de determinante, en una dimensión la distancia euclídea entre 2 puntos se define

$$l_{12} = X_1 - X_2 = \begin{vmatrix} X_1 & 1 \\ X_2 & 1 \end{vmatrix} \quad (2.17)$$

En coordenadas homogéneas las coordenadas se multiplican por un factor arbitrario  $\lambda$ , por lo que en dos dimensiones el determinante es multiplicado por  $\lambda_1$  y  $\lambda_2$ . Como ya se vio antes una transformación proyectiva en coordenadas homogéneas es lineal (2.16), por lo que solo se multiplica el determinante por  $|T|$  y  $\lambda_i$ .

$$X_i = \lambda_i(X_i, Y_i, 1) \rightarrow |\widetilde{X}_1, \widetilde{X}_2| = \lambda_1 \lambda_2 |T| |X_1, X_2| \quad (2.18)$$

Por lo tanto, se puede decir que determinante de puntos y líneas en coordenadas homogéneas es una invarianza proyectiva invariante de peso -1 en T y con un grado de invarianza de 1 en cada  $\lambda$ . Este método consiste en buscar relaciones de varios determinantes en los cuales todos los factores  $\lambda_i$  y  $|T|$  se anulen, convirtiendo así los invariantes relativos en absolutos. Para cuatro puntos colineales, se puede definir la relación cruzada de sus distancias euclídeas como la relación cruzada de los determinantes en coordenadas homogéneas.

$$\frac{l_{12}l_{34}}{l_{13}l_{24}} = \frac{(\lambda_1 \lambda_2 |X_1, X_2|)(\lambda_3 \lambda_4 |X_3, X_4|)}{(\lambda_1 \lambda_3 |X_1, X_3|)(\lambda_2 \lambda_4 |X_2, X_4|)} = \frac{|X_1, X_2| |X_3, X_4|}{|X_1, X_3| |X_2, X_4|} \quad (2.19)$$

Como se aprecia la matriz de transformación T se anuló al igual que las  $\lambda_i$ , por lo que esta relación de distancias entre 4 puntos coloniales se cumplirá ante cualquier cambio de perspectiva. En el anexo 3 se muestran las normas y principales que rigen los pasos de peatones en España, estos están compuestos por bandas de 50 cm de ancho y separadas entre sí 50 cm. Al ser las bandas equidistantes, conociendo solo 4 puntos, uno por cada segmento longitudinal de las bandas y estando los puntos alienados entre sí, se puede definir la invarianza proyectiva de un paso de peatones con la ecuación (2.19). Como se muestra en la Figura 1.3 se sabe también que la distancia entre todos los puntos es la misma por lo que:

$$I = \frac{l_{12}l_{34}}{l_{13}l_{24}} = \frac{w w}{2w 2w} = 0.25 \quad (2.20)$$

Donde  $w$  es el grosor y la separación de las bandas. Si se tienen  $n$  puntos se pueden obtener  $(n - 3)$  proyecciones invariantes  $I$ .

Para comprobar que en la imagen hay un paso de peatones se toman las coordenadas de los puntos de cada banda (puntos característicos) y si sus distancias cumplen la relación (2.20) más menos un porcentaje que se obtiene experimentalmente, se considerara un paso de peatones

$$\left| \frac{I(k)}{0.25} - 1 \right| < \frac{\text{Porcentaje}}{100} \quad k = 1, 2, \dots, n - 3 \quad (2.21)$$

## 2.2.- MEDIDA DE DISTANCIAS.

Las cámaras se encargan de mapear el mundo tridimensional en un plano bidimensional, llamado imagen. Es necesario conocer cómo se forma una imagen para poder obtener información 3D a partir de ellas. Por lo que primero, se ha de hablar del modelo de que se usara para definir una cámara.

### 2.2.1.- Modelo de una cámara.

El modelo *pinhole* [18] es una aproximación muy usada para una gran cantidad de aplicaciones de la visión artificial debida a su similitud con cámaras de lente delgada. En los todos los puntos del espacio tridimensional son proyectados sobre un plano de dos

dimensiones, llamado plano imagen, siguiendo cada uno una línea recta que pasa por el centro de la cámara (también conocido como centro óptico o centro de la proyección).

En la figura 2.6 se muestra el modelo en el cual el plano imagen se encuentra separado del centro óptico una distancia  $f$ , denominada distancia focal. Este parámetro lo determina la lente de la cámara.

La transformación proyectiva puede ser factorizada en tres transformaciones simples, las cuales corresponden a transiciones entre los cuatro sistemas de coordenadas que se muestran en la figura 2.6 y se explican a continuación.

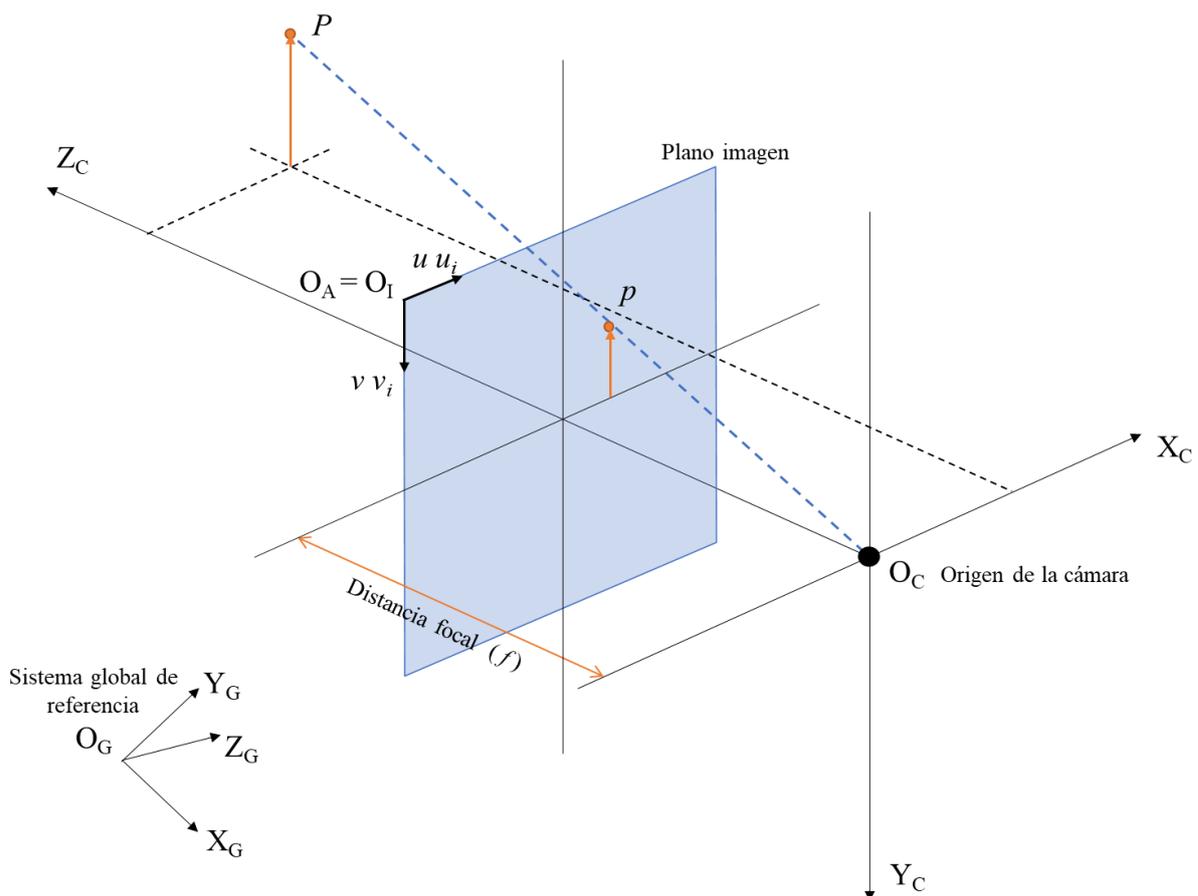


Figura 2.6. Modelo de proyección central. El punto  $P$  se proyecta en el plano imagen como el punto  $p$  únicamente con dos coordenadas, siendo la  $Z_C$  del plano igual a la distancia focal de la cámara.

- I. Sistema de referencia global en coordenadas euclideas, tiene su origen en el punto  $O_G$ .
- II. Sistema de referencia de la cámara en coordenadas euclideas, su origen coincide con el punto focal y se denota con la letra y subíndice  $O_C$ . El eje  $Z_C$  está alineado con el eje óptico y pasa perpendicularmente por el plano imagen.
- III. Sistema de referencia de la imagen en coordenadas euclideas, sus ejes están alineados con el sistema coordenadas de la cámara. Siendo la coordenada  $u_i$  análoga a la  $X_C$  y la  $v_i$  con la  $Y_C$ . La coordenada  $Z$  no aporta información en la imagen.
- IV. Sistema de referencia afín de la imagen, su origen coincide con el del sistema de coordenadas euclídeo de la imagen, se denota como  $O_A$ . Los ejes coinciden a excepción de  $v$ , el cual puede presentar un giro con respecto a  $v_i$ .

La primera transformación consiste en el paso de un sistema arbitrario como el sistema de coordenadas I al de la cámara, es decir el sistema de referencia II. Esto se puede hacer con dos movimientos, una traslación definida por el vector de traslación  $t$  y una rotación con la matriz  $R$ , la cual expresa las tres rotaciones elementales en los ejes de coordenadas  $(X_G, Y_G, Z_G)$ . Estos dos elementos son los conocidos como parámetros extrínsecos de la cámara. Siendo  $P$  un punto en el espacio esta transformación expresada en coordenadas homogéneas queda así:

$$P_C = \begin{pmatrix} R & -Rt \\ 0_{1 \times 3} & 1 \end{pmatrix} P_G \quad (2.22)$$

La segunda transformación representa el cambio de coordenadas del sistema II al III, proyectando el punto  $P_C$  sobre el plano imagen y expresándolo en el sistema de coordenadas euclídeo de la cámara. Este cambio de  $R^3$  a  $R^2$  genera dos ecuaciones (2.23) (2.24) no lineales, donde  $f$  es la distancia focal de la cámara. Escribiendo estas ecuaciones en coordenadas homogéneas se genera la siguiente relación.

$$u_i = \frac{X_{cf}}{z_c} \quad , \quad v_i = \frac{Y_{cf}}{z_c} \quad (2.23) \quad (2.24)$$

$$p_i = \begin{pmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} P_c \quad (2.25)$$

La última transformación es la que tiene en cuenta el paso de coordenadas reales a píxeles y la forma que estos puedan tener, es decir se va del sistema III al IV. La matriz que define esta relación se llama la matriz de calibración intrínseca ( $K$ ) y como su nombre lo indica posee los parámetros intrínsecos de la cámara.

$$p = K u_i = \begin{pmatrix} g_u & s & -u_0 \\ 0 & g_v & -v_0 \\ 0 & 0 & 1 \end{pmatrix} u_i \quad (2.26)$$

Los parámetros que componen la matriz  $K$  son:  $g_u$  y  $g_v$  factores de escalado a lo largo del eje  $u$  y  $v$  respectivamente, normalmente ambos valores son iguales a la distancia focal. El *skew* ( $s$ ) se debe a un error en la alineación de los ejes horizontal y vertical del sensor, y de aquí viene la posibilidad de que  $v$  y  $v_i$  no coincidan. Todos estos valores pueden ser extraídos mediante un proceso de calibración de cámara [19], [20].

La unión de estos tres movimientos, es decir la multiplicación de las tres matrices es lo que define el modelo completo de una cámara.

### 2.2.2.- Imágenes estéreo.

Partiendo de una sola cámara es muy complicado extraer información 3D de la escena, conociendo únicamente el modelo descrito anteriormente. Un punto en el plano imagen corresponde a una línea en el espacio, ya que no se puede determinar la profundidad a la que este se encuentra. El sistema visual del ser humano muestra una solución a dicho problema, al igual que este posee dos ojos el uso de dos cámaras (figura 1.5) da dos vistas diferentes, las cuales permiten obtener todo tipo de información 3D de los puntos que se ven en ambas. Es necesario conocer como están dispuestas las cámaras entre sí para poder hacer esto.

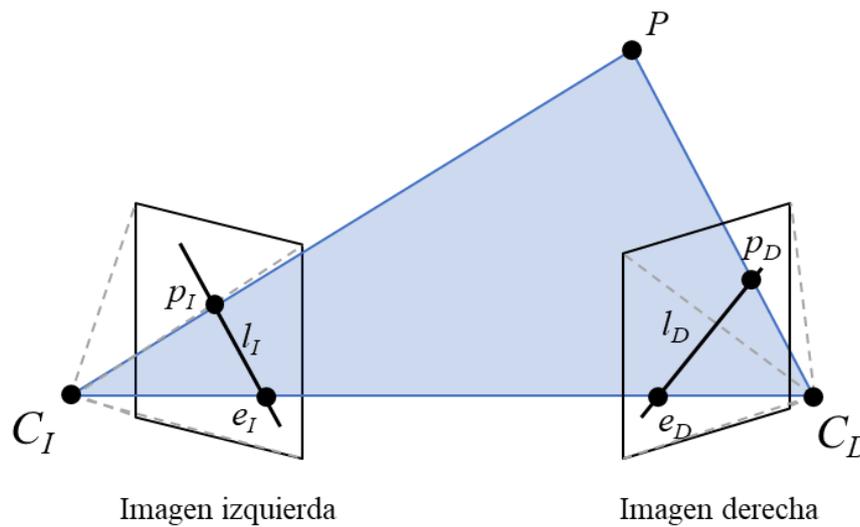


Figura 2.7. Geometría de dos cámaras en estéreo.  $l_I$  y  $l_D$  son las líneas epipolares, mientras que  $e_I$  y  $e_D$  son los epipolos. Basada en [21].

La geometría de un sistema general de dos cámaras se muestra en la Figura 2.7, es conocida como geometría epipolar. La línea que conecta los puntos  $C_I$  y  $C_D$  es la línea base, la cual intercepta los planos imagen en los epipolos  $e_I$  y  $e_D$ . Cualquier punto  $P$  visto desde las dos cámaras, forma un plano con los centros ópticos de ambas  $C_I$  y  $C_D$ , llamado plano epipolar. Las líneas epipolares son aquellas que se forman con la intersección de los planos imagen con el plano epipolar.

Siendo  $p_I$  y  $p_D$  las proyecciones del punto  $P$  en ambas cámaras. La línea  $C_I P$  muestra todas las posibles posiciones de  $P$  en la imagen de la izquierda y se proyectan como la línea epipolar  $l_D$  en la de la derecha. Esto se conoce como correspondencia epipolar.

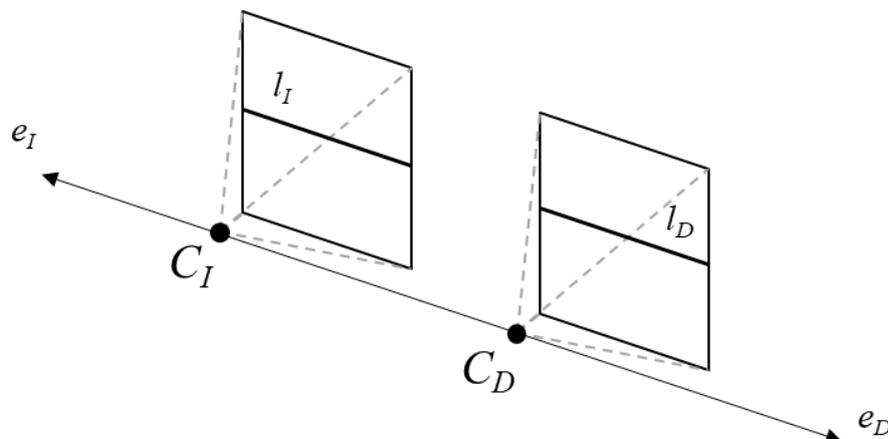


Figura 2.8. Configuración estéreo rectificada. Las líneas epipolares son paralelas y los epipolos se sitúan en el infinito. Basado en [21].

Para medir las distancias a la que se encuentran los objetos de la cámara Yasir, Rahizall, Hasbullah y Amelia [14] usaron una configuración estéreo en la cual las cámaras están en paralelo. A esta disposición se le conoce como configuración rectificada o canónica [21], donde los planos imagen coinciden entre sí y la línea  $C_I C_D$  es paralela a ellos. Los epipolos se sitúan en el infinito y las líneas epipolares coinciden con las filas de las imágenes, tal y como se muestra en la figura 2.8. Esta configuración facilita todas las operaciones posteriores, como la búsqueda de correspondencias de un punto en a las imágenes, ya que al estar en la misma horizontal se reducen los posibles candidatos y por ende los errores.

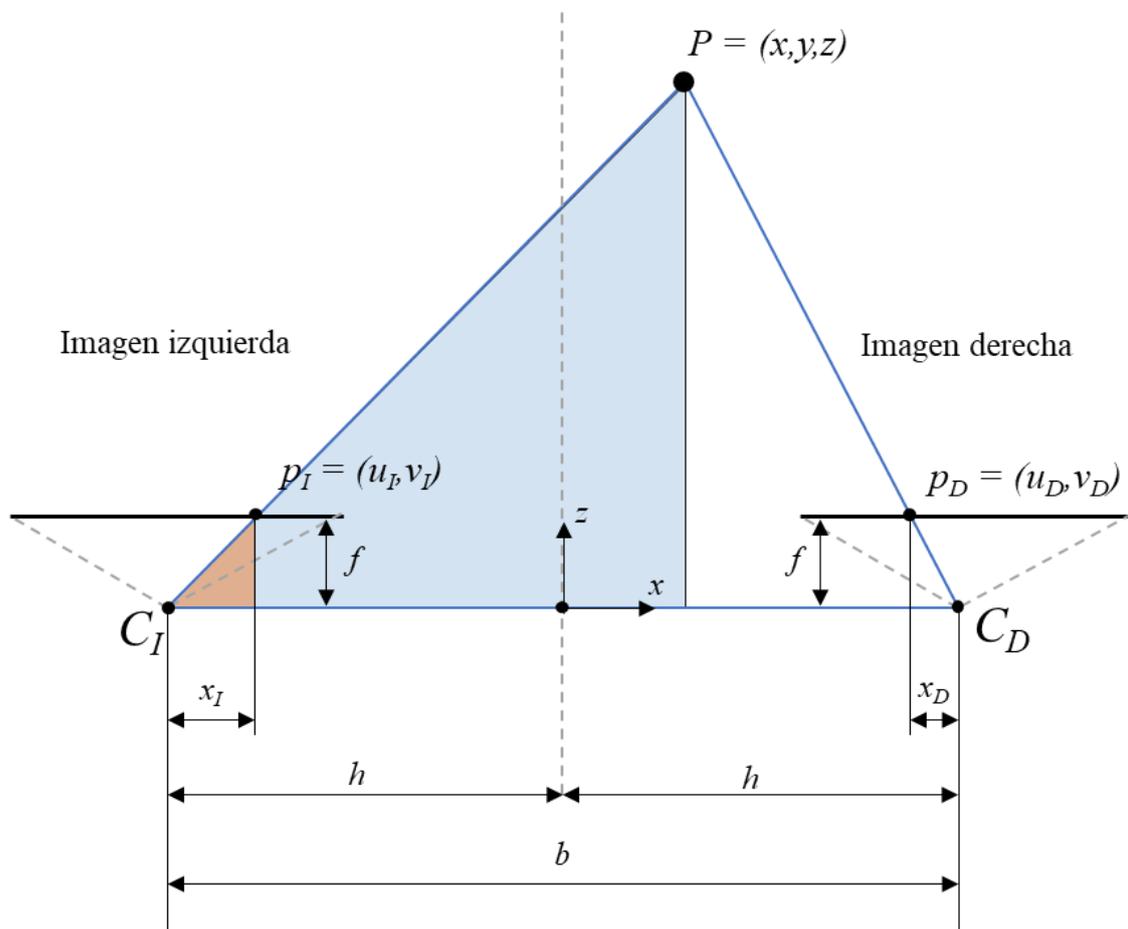


Figura 2.9. Variables y medidas básicas presentes en una configuración estéreo rectificada, donde se proyecta el punto  $P$  sobre los dos planos imagen. En este caso  $x_I$  es positivo, está a la derecha del centro óptico y  $x_D$  es negativo. Basada en [21].

En la Figura 2.9 se muestran dos cámaras con la configuración canónica, la separación entre sus centros ópticos se denomina línea base y se denota con la letra  $b$ . El centro de esta línea se usa como sistema de coordenadas para definir la posición de todos

los puntos en el espacio. El punto  $P$  se encuentra en el campo de visión de las dos cámaras y este es proyectado sobre sus planos imagen, quedando definido únicamente por dos coordenadas. La diferencia que existe entre las coordenadas  $u$  de estos puntos proyectados se conoce como disparidad ( $d$ ); es decir la diferencia entre  $x_I$  y  $x_D$ .

La coordenada  $z$  del punto  $P$  representa la distancia a la que se encuentra este punto de la cámara, la cual se puede calcular mediante semejanza de triángulos. Las rectas  $C_I P_I$  y  $p_I P$  tienen la misma inclinación y son las hipotenusas de dos triángulos, el primero es el de color naranja y el segundo el de azul. En el caso de la cámara de la derecha se da el mismo caso, pero al revés, dando como resultado las siguientes relaciones:

$$\frac{x_I}{f} = -\frac{h+x}{z} \quad , \quad \frac{x_D}{f} = -\frac{h-x}{z} \quad (2.27) \quad (2.28)$$

Combinando las dos expresiones anteriores y eliminando la coordenada  $x$  se puede definir  $z$  como:

$$z = \frac{2hf}{x_I - x_D} = \frac{bf}{d} \quad (2.29)$$

Como ya se dijo antes, la disparidad se calcula como la diferencia entre  $x_I$  y  $x_D$ , teniendo en cuenta sus signos. Pero también se puede calcular como se mostró en la Figura 1.6, como la diferencia entre las coordenadas  $u_I$  y  $u_D$  de las proyecciones del punto. Con esto ya se conoce la distancia a la que se encuentra un punto u objeto de la cámara, para las otras dos coordenadas se usarían estas las ecuaciones:

$$x = \frac{-b(x_I + x_D)}{2d} \quad , \quad y = \frac{bv}{d} \quad (2.30) \quad (2.31)$$

## 3.- Experimentos y resultados.

Los medios con los que se va a hacer el proyecto son, dos cámaras industriales USB 2 uEye XS del fabricante Imaging Development Systems (IDS), estas pueden capturar 15 fotogramas por segundo de 5 megapíxeles. Ambas cámaras se disponen sobre una misma base, de forma que estén paralelas y alineadas entre sí, quedando separadas a una distancia de 125 mm. Se empleará también la cámara trasera de un iPhone 6 de la marca Apple, para recolectar un mayor número de muestras, destinadas a la etapa de detección de pasos de peatones. Esta cámara es de 8Mpx.

Se creó una base de datos con distintas imágenes, tanto con las cámaras en estero como con la individual, las cuales en su mayoría poseen pasos de peatones. Las imágenes son tomadas desde una altura de 1,7 metros desde el nivel del suelo aproximadamente. Para probar el algoritmo se hace uso tanto de las librerías de OpenCV 3.2.0 para C++ sobre el sistema operativo Ubuntu 16.04 como el programa MATLAB de MathWorks. Se uso un computador con un procesador Intel Core i7 7700HQ de 2,80 GHz y 16 Gb de memoria RAM.

### 3.1.- DETECCIÓN DE LOS PASOS DE PEATONES.

En este apartado se especifica la forma en que se llevó a cabo el algoritmo para la detección de los pasos de peatones, cada uno de sus pasos y resultados. Algunos de los valores usados en este algoritmo se obtuvieron de forma experimental, buscando aquellos que diesen mejores resultados. Cada cámara tiene diferentes parámetros (anexo 1). El procedimiento comienza pasando la imagen de entrada a escala de grises, y continua con los siguientes pasos son:

- I. Segmentación en regiones homogéneas, se divide la imagen en bloques de 80x80 píxeles<sup>1</sup>, y se asigna una etiqueta al primer bloque, el de arriba a la izquierda. Con el histograma de este bloque se calculan sus medias ( $\mu_1$  y  $\mu_2$ ), se selecciona su bloque vecino y se calculan también las medias ( $\tilde{\mu}_1$  y  $\tilde{\mu}_2$ ). Con estos valores se

---

<sup>1</sup> El valor de cada uno de estos parámetros corresponde a las cámaras estéreo. En el anexo X se encuentran los correspondientes a la cámara individual, pueden tener el mismo valor.

calcula la relación de distancias (2.8) y si está es mayor a  $0.7^1$  se considerará que ambos bloques tienen la misma distribución, perteneciendo ambos a la misma región. Por lo tanto, si es mayor al valor límite, al bloque vecino se le pone la misma etiqueta, pero si es menor se le pone otra etiqueta. Esto se repite con todos los bloques de la imagen hasta que esta esté completamente etiquetada (figura 3.2).



Figura 3.1. Ejemplo de una imagen de entrada para el algoritmo de detección.



Figura 3.2. Imagen etiquetada, obtenida tras el paso I. Se muestran las diferentes regiones obtenidas de la imagen de entrada. Tiene el brillo aumentado con respecto a la original para que se pueda apreciar.

- II. Regiones bipolares, se calcula cuál es la región con el área más grande, y a aquellas regiones que tengan un área mayor al 50% de esta se les calcula su bipolaridad. El grado de bipolaridad de una región se calcula con la ecuación (2.7) siendo el

parámetro  $\alpha$  igual a  $0.6^1$ , si devuelve un resultado mayor a  $0.65^1$  se considera que la región de la imagen es lo suficientemente bipolar como para ser un candidato a paso de peatones.

- III. Eliminación de los candidatos mal posicionados. Si los candidatos están en la zona superior de la imagen, por perspectiva las bandas del paso de peatones estarán muy juntas y complicarán la extracción de los puntos característicos, además de que, si está muy lejos, el usuario no requiere aun que se le informe de la posición. Si un candidato está solo en los laterales de la imagen es posible que esté cortado, por lo que no se consideraran candidatos, pero se le informara al usuario que más adelante a su derecha o a su izquierda tiene un posible paso de peatones. Por lo tanto, aquellos candidatos que solo estén dentro de la zona correspondiente al 40% superior de la imagen serán eliminados, al igual que aquellos que estén en el 20% de la imagen desde alguno de los laterales (figura 3.3).



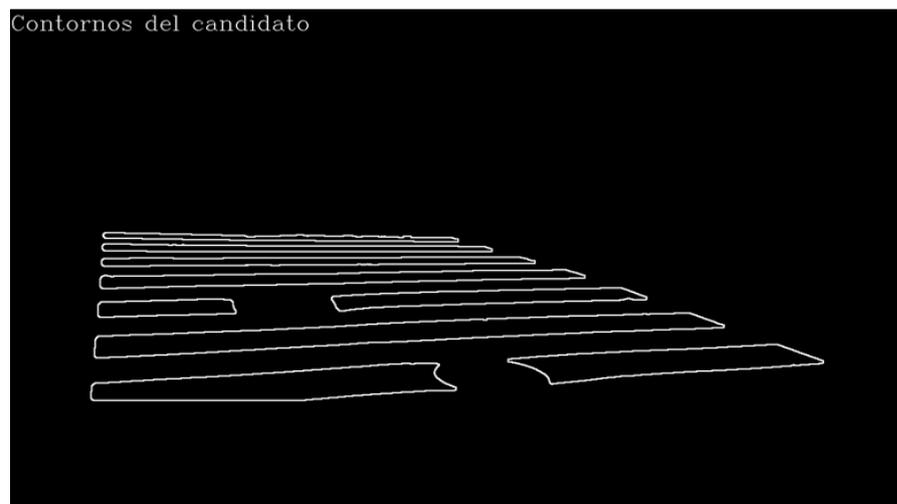
Figura 3.3. Tras los pasos II y III se definen las regiones en la imagen que pasan a considerarse como candidatas a poseer pasos de peatones.

- IV. Creación de una imagen binaria del candidato. Lo primero que se hace es aplicar un filtro de mediana [22] para eliminar el ruido y suavizar las texturas tanto del asfalto como de la pintura del paso de peatones. Aplicando un umbralizado [16] definido por el mismo valor  $\alpha$  que se usó anteriormente se obtendría una imagen binaria en la que solo estarían presentes las bandas blancas. El siguiente paso es realizar un

proceso morfológico, específicamente el cierre [23], para eliminar posibles huecos en las bandas debidos al desgaste de la pintura (figura 3.4(a)).



(a)



(b)

Figura 3.4. Primeros pasos por los que pasa un candidato: (a) imagen tras haber aplicado el filtro de mediana, umbralizado y cierre, y (b) contornos de la región

- V. Detección de las rectas de las bandas. Aplicando el detector de contornos Sobel [17] tanto en la dirección x como en la dirección y, se unen ambas imágenes para tener una de contornos (figura 3.4(b)). Finalmente se aplica la transformada de Hough probabilística, la cual devuelve cada línea como sus coordenadas del punto inicial y del final (figura 3.5). Si no se detecta ninguna línea se elimina el candidato y se pasa al siguiente si todavía quedan.

- VI. Extracción de los puntos característicos. Lo primero es determinar cuál es el ángulo medio de todas las rectas detectadas por Hough, en función de este ángulo se procederá. Si la orientación de las bandas es aproximadamente horizontal (menor a  $20^\circ$  y mayor a  $160^\circ$ , solo se trabaja en el primer y segundo cuadrante), se recorre la imagen variando la coordenada Y, pero manteniendo constante la X, siendo esta igual al valor medio de las coordenadas X de los puntos finales e iniciales de todas las rectas. Cada vez que se pase por una línea se guardarán las coordenadas para definir con estas los puntos característicos. En el caso de que las bandas sean verticales (entre  $20^\circ$  y  $340^\circ$ ) la coordenada Y se mantiene constante en el valor medio y se va incrementando la X, análogo a si son horizontales (Figura 3.6).



Figura 3.5. Rectas detectadas en el candidato mediante la transformada de Hough probabilística.



Figura 3.6. Puntos característicos encontrados en la región que se estudia.

VII. Invarianza proyectiva. El último paso es verificar que la ecuación (2.21) se cumple, siendo el rango admisible de un 10%. Se empieza con las coordenadas de los 4 primeros puntos característicos para después repetirlo con las coordenadas del segundo punto hasta el quinto y así sucesivamente con todos. Si la ecuación se verifica, aunque sea una vez ya se considera el candidato como paso de peatones. En el caso de que todavía queden candidatos por revisar se va con uno nuevo al paso 4 y así hasta terminar con todos.

Para comprobar la efectividad del algoritmo se procesaron una total de 56 fotos de las bases de datos creadas. En la Tabla 3.1 se muestran los resultados de todas las imágenes capturadas por la cámara individual, mientras que en la Tabla 3.2 se ven los resultados de las fotos realizadas con las cámaras en estéreo. La efectividad del algoritmo resulto en un 57,14%.

Decisión	Imagen con un paso de peatones	Imagen sin pasos de peatones
Hay un paso de peatones	19	3
Sin pasos de peatones	14	5

Tabla 3.1. Resultados de la detección, imágenes de la cámara individual

Decisión	Imagen con un paso de peatones	Imagen sin pasos de peatones
Hay un paso de peatones	8	0
Sin pasos de peatones	7	0

Tabla 3.2. Resultados de la detección, imágenes de las cámaras en estéreo

En la figura 3.7 se ven los resultados de algunos de los pasos que sufre otra de las imágenes y uno de sus candidatos hasta que se detecta el paso de peatones, el cual no es detectado por completo ya que en la zona más alejada las bandas están difusas y se definen mal las regiones. En este caso no se detectó nada en los otros dos candidatos.

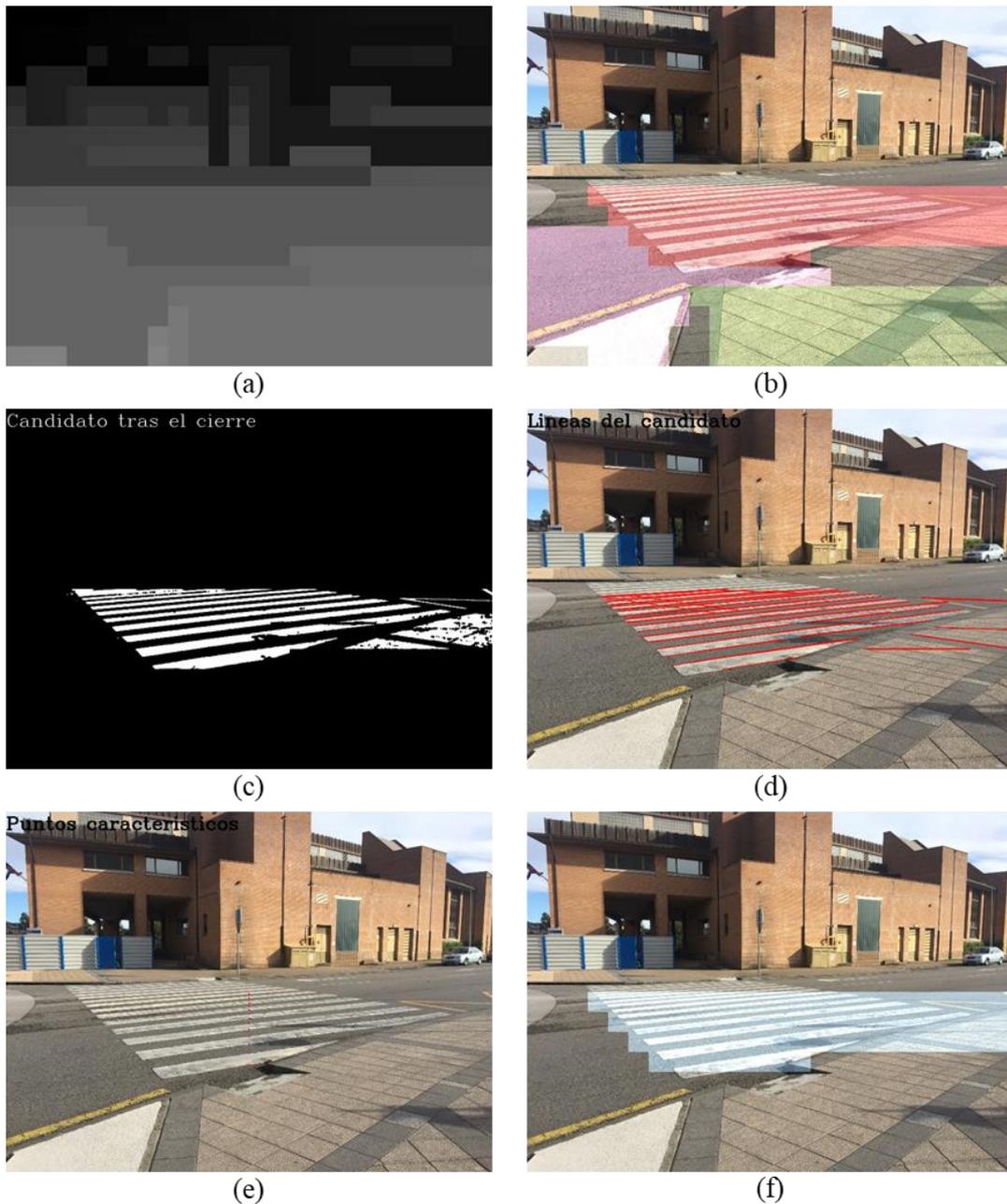


Figura 3.7. Etapas del algoritmo de detección de pasos de cebr: (a) imagen segmentada en regiones homogéneas, (b) en color están las regiones bipolares, (c) uno de los candidatos después del umbralizado y el cierre, (d) rectas detectadas, (e) extracción de los puntos característicos, imagen aumentada de tamaño y (f) resultado final luego de que se comprobasen todos los candidatos, a color la región donde se detectó el paso de peatones.

Hay una gran cantidad de pasos de cebr que no fueron detectados, aquellos que tenían la pintura muy desgastada no eran lo suficientemente bipolares y por lo tanto no pasaban a ser candidatos, también ocurre lo mismo cuando el asfalto no es lo suficientemente oscuro o hay una sombra sobre las bandas, en la figura 3.8(a) esto es apreciable. Otra ocasión en la que no se detectan es cuando los pasos de peatones son

cortos, no son reconocidos debido a que no se extraen adecuadamente los puntos característicos, en la figura 3.8(b) en la cual no se detectó nada solo se ve una banda completa mientras que las otras dos están cortadas. También se da la situación en la cual la segmentación divide el paso de cebra en regiones más pequeñas y al analizar cada una por separado ocurre lo mismo que en el caso anterior.

Como se había explicado anteriormente la transformada de Hough probabilística escoge píxeles para la votación de forma aleatoria, lo que repercute en los resultados de la misma forma, es decir que para una misma imagen pueden generarse resultados distintos, esto se aprecia en la figura 3.8(c), en la cual el paso de peatones es muy largo y la pintura está muy agrietada, generando muchas líneas que pueden variar levemente su posición cada vez que se aplica el algoritmo, lo que unas veces resultaría en que la ecuación (2.21) se cumpliera y otras en las que no.

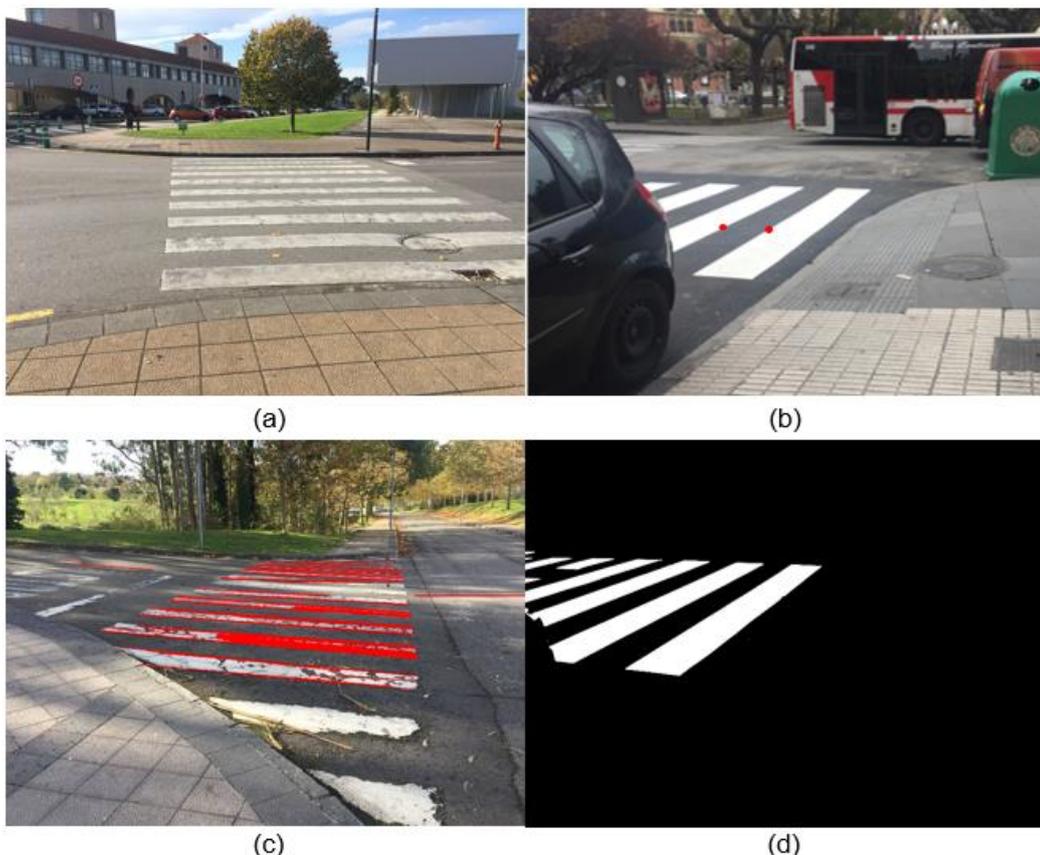


Figura 3.8. Algunos resultados experimentales: (a) imagen con un paso de peatones muy desgastado y un asfalto muy claro, (b) imagen en la que no se detectó el paso de peatones, solo se identificaron dos puntos característicos, (c) imagen que presenta una paso peatones con la pintura agrietada por lo que se detecta una gran cantidad de líneas y (d) imagen umbralizada de un candidato que posteriormente se define como paso de peatones.

Una imagen que presenta resultados óptimos es la que se muestra en la Figura 3.8(d), la cual ya ha pasado por el umbralizado, hay suficientes bandas para obtener los puntos característicos y en la imagen original estas están bien contrastadas. De las imágenes que no poseen pasos de peatones en 3 se dio un falso positivo, algo que es peligroso de cara al usuario, aunque ninguna de estas detectó un paso de peatones en la carretera

### 3.2.- MEDIDAS ESTÉREO.

A partir de esta etapa solo se pueden usar imágenes obtenidas con las cámaras en estéreo. El primer paso consiste en realizar una calibración de la cámara, para saber la distancia focal de las cámaras y así poder determinar la profundidad de un punto con la ecuación (2.29).

Para esta tarea se usa el programa MATLAB de MathWorks, específicamente su aplicación para la calibración de una cámara simple (*Single Camera Calibration App*), basada en el método de Heikkila [19] y Zhang [20]. Se tomaron 15 imágenes de un tablero de ajedrez con la retícula de 80x80mm y 16x13 intersecciones. De cada cámara se obtienen un valor de la distancia focal, al ser las cámaras iguales sus valores se parecen, por en los pasos posteriores se usará la media de estos valores. La herramienta de calibración da la distancia focal en píxeles, para pasarla a milímetros se tendría que multiplicar por un valor que relaciona estas dos unidades. Independientemente de esto el valor la distancia focal se necesita en píxeles, dando como resultado un valor de 1339,4 píxeles.

Para comprobar el funcionamiento sin la influencia de perturbaciones provocadas por la escena, se determinará la distancia a la que se encuentra un objeto en un entorno controlado. Se parte de tres imágenes, las dos obtenidas por la cámara estéreo y una del objeto a estudiar. El algoritmo se inicia realizando un proceso de *template matching* [5] con la imagen del objeto en las dos tomadas por las cámaras, esto devuelve dos imágenes en escala de grises. Se busca en ambas el pixel más claro; es decir, el de mayor valor, siendo su posición la misma en la que se encuentra el objeto en las imágenes. Conocidas las coordenadas en ambas fotografías se pasa a calcular la disparidad tomando las

coordenadas  $x$  y restándolas entre sí. Finalmente, junto con la distancia focal y la línea base (125mm) se calcula la distancia a la que se encuentra el objeto con la formula (2.29).

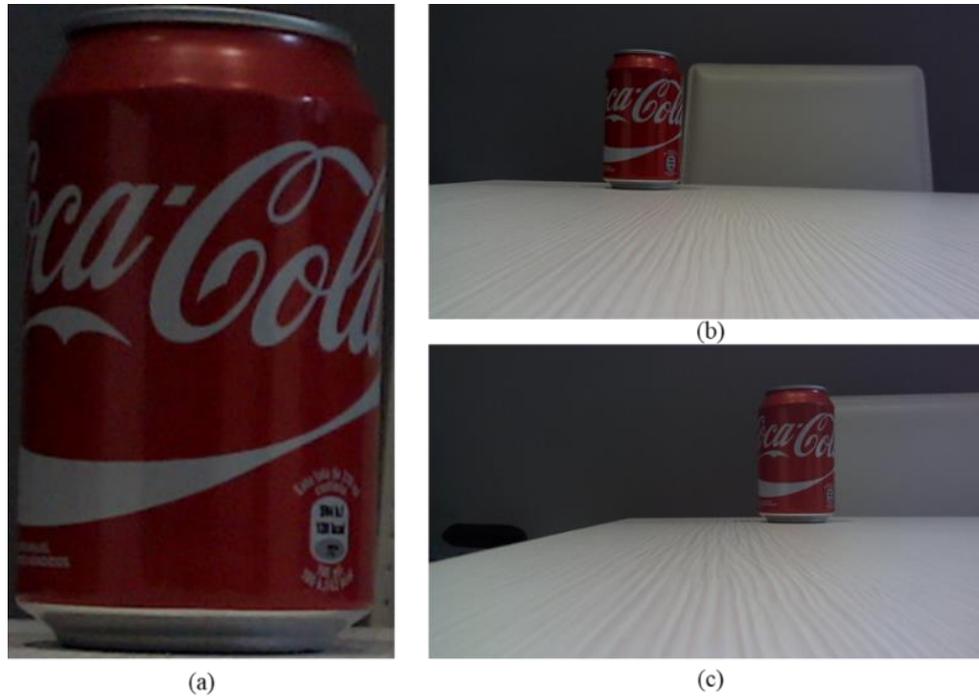


Figura 3.9. Imágenes de entrada para determinar la distancia a la que se encuentra un objeto. Son: (a) el objeto, (b) imagen de la cámara derecha, (c) imagen de la cámara izquierda.

Imagen	Distancia (m)		Disparidad (píxeles)		Error (m)	Error (píxeles)
	Real	Detectada	Real	Detectada		
1	0,900	0,792	186	211	0,108	25
2	0,480	0,459	349	365	0,021	16
3	0,270	0,267	620	627	0,003	7
4	1,500	1,204	112	139	0,296	27

Tabla 3.3. Resultados del cálculo de la distancia sobre entorno controlado.

En la Tabla 3.3 se muestran los resultados obtenidos a partir de las cuatro imágenes hechas. Suponiendo que no existe error alguno en la calibración, ni en la base donde se dispusieron las cámaras, se puede calcular la disparidad real, la cual es la que debería de haber salido para que la distancia real y la calculada coincidiesen. La diferencia entre los valores nos determina el error tanto en metros como en píxeles. El error en píxeles de la imagen 1 y 4 es aproximadamente el mismo, pero en metros la imagen 4 tiene un error casi

tres veces mayor al de la 1. Se puede ver en la Figura 3.10 que esto se debe a que mientras más lejos esté el objeto, el error cometido al calcular la disparidad afecta más a la distancia.

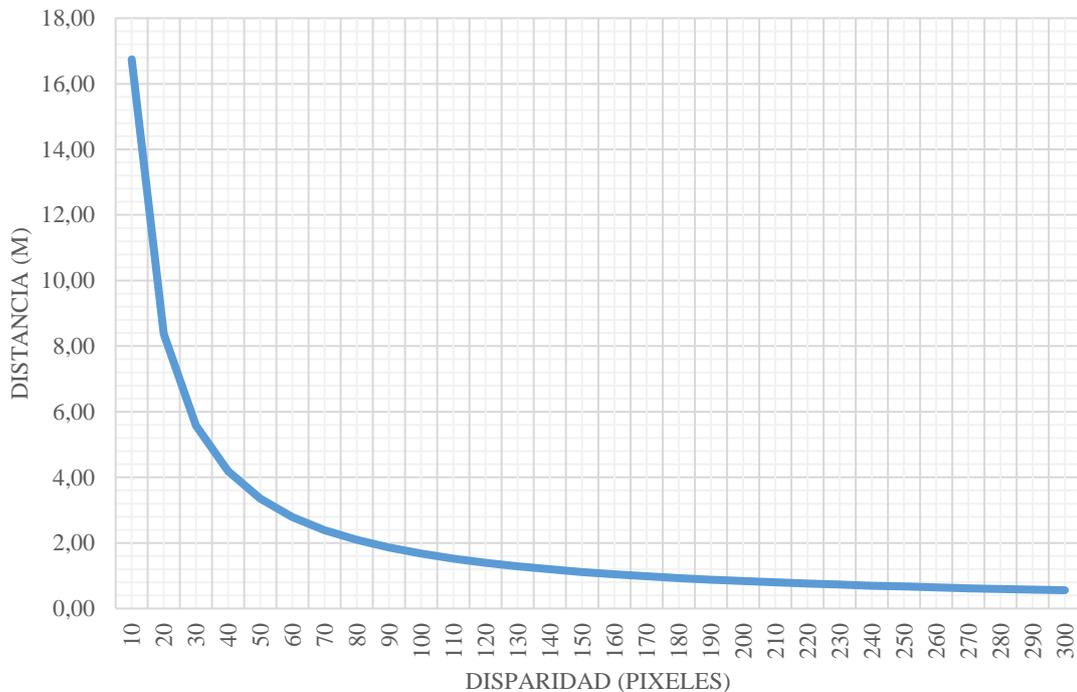


Figura 3.10. Gráfico que relaciona la disparidad con la distancia a la que se encuentra un objeto. Para la cámara estéreo usada.

Una vez hecho esto se pasa a implementar el algoritmo sobre lo obtenido en la etapa de detección, tras procesar la imagen de la cámara derecha, esto se realizará para todas las regiones definidas como pasos de peatones. Los pasos a seguir (figura 1.7) son los siguientes:

- I. Punto del paso de peatones más cercano a la cámara, en función del ángulo medio de las rectas detectadas por la transformada de Hough probabilística se escogerá un punto característico como el más cercano a la cámara. Los ángulos pueden tener un valor de entre  $0^\circ$  y  $180^\circ$ . Si este es menor a  $20^\circ$  o mayor a  $160^\circ$  las bandas están horizontales (figura 3.11(a)), por lo que el punto más cercano es el que tenga la coordenada y de mayor valor. En el caso de que esté entre  $90^\circ$  y  $160^\circ$  (figura 3.11(b)), como la coordenada y es igual para todos los puntos se selecciona aquel

que tenga la  $x$  más pequeña. Finalmente, si está entre  $20^\circ$  y  $90^\circ$  (figura 3.11(c)) se escoge aquel que tenga la coordenada  $y$  de mayor tamaño

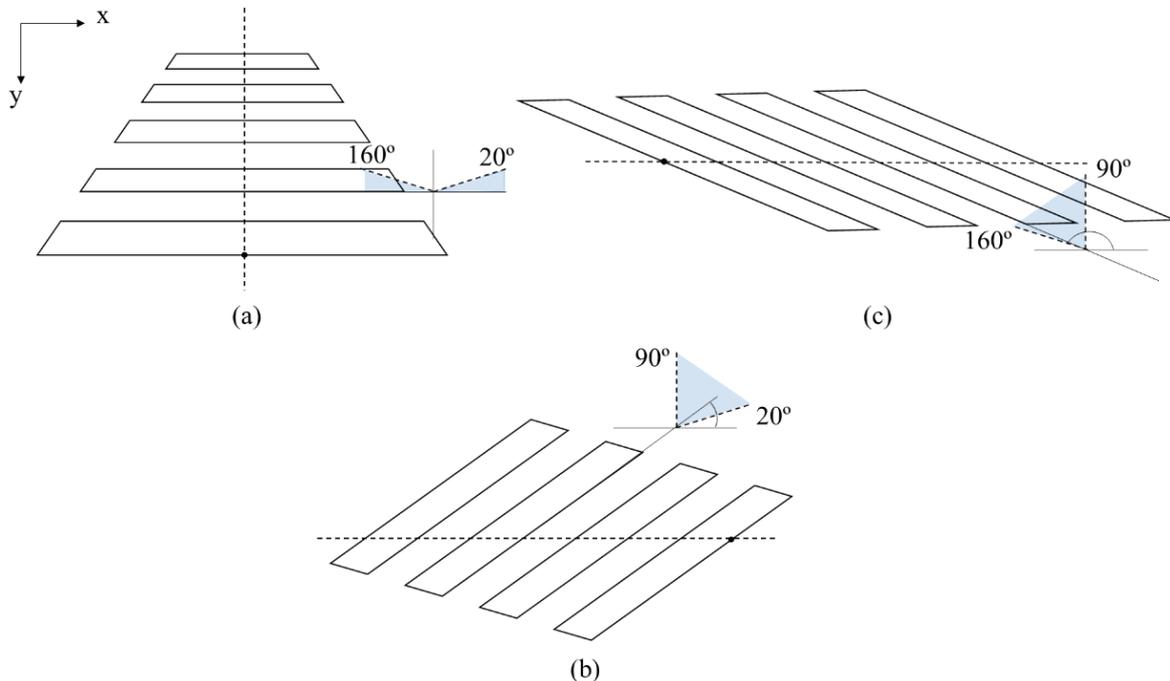


Figura 3.11. Se muestran las posibles orientaciones que puede tener el paso de peatones en la imagen, y como en función de este se decide cual es el punto más cercano a la cámara. La línea rayada representa la recta en la cual se encuentran todos los puntos característicos.

- II. Se recorta un recuadro (Figura 3.12(a)) en la imagen de la derecha de  $100 \times 100$  píxeles dejando en el medio el punto más cercano. En el caso de que el punto se encuentre a menos de 100 píxeles del borde inferior de la imagen, el recuadro pasara a ser igual a el número de filas menos la coordenada  $y$  del punto. Este cuadrado es la plantilla (*template*) del punto a buscar.
- III. Se recorta la zona de búsqueda de la imagen de la izquierda (Figura 3.12(b)), al estar las líneas epipolares en la misma horizontal se escoge la coordenada  $y$  del punto más cercano y se pone como la fila central de la zona de búsqueda. Se deja como margen superior e inferior un numero de filas igual al tamaño de la plantilla (*template*).
- IV. Con la zona de búsqueda y la plantilla se realiza el proceso de *template matching* [5], el resultado es una imagen en la cual el pixel de mayor valor es en el cual se

encuentra el punto que más se parece a la plantilla (Figura 3.12(c)). Para obtener el punto más cercano en la imagen de la izquierda hay que ajustar las coordenadas de este punto obtenido antes, sumando a ambas coordenadas la mitad de la distancia de las filas de la plantilla.

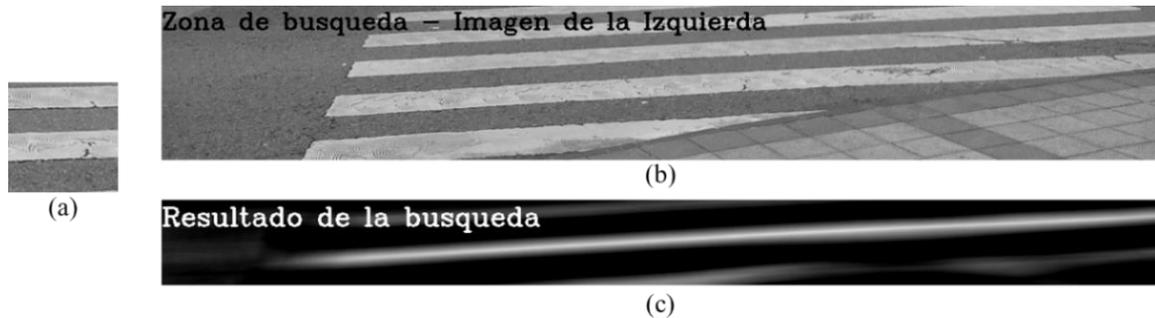


Figura 3.12. Imágenes usadas y obtenidas en los pasos II, III y IV. Se tiene (a) la plantilla donde está el punto que se busca, (b) la zona de búsqueda en la imagen de la izquierda y (c) el resultado del proceso de *template matching*.

- V. Se calcula la dispersión restando la coordenada  $x$  del punto en la imagen de la derecha con el de la izquierda. Para obtener la distancia a la que se encuentra este punto con la ecuación (2.29).

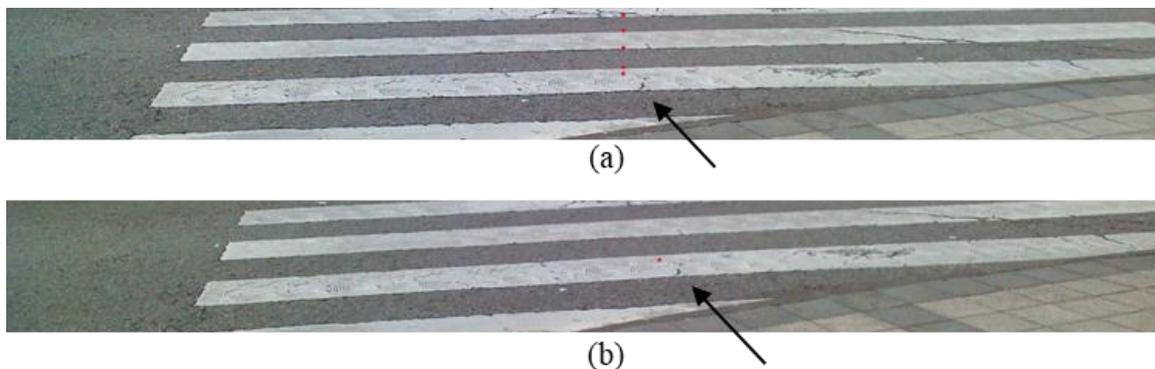


Figura 3.13. Resultado del *template matching*, en (a) se ve el punto en la imagen de la derecha y en (b) el punto encontrado en la de la izquierda. Gracias a la marca de la pintura señalada por las flechas, se ve como la posición del punto es la correcta.

Como se mostró en la Tabla 3.2, en 8 de las imágenes en estéreo se detectó un paso de peatones, de las cuales en 6 se tiene registro sobre su distancia. En la Tabla 3.4 se muestran tanto las reales como las obtenidas por el programa y al igual que antes, se calculó el error tanto en metros como en píxeles.

Imagen	Distancia (m)		Disparidad (píxeles)		Error (m)	Error (píxeles)
	Real	Detectada	Real	Detectada		
1	6,120	4,069	27,357	41,146	2,051	13,789
2	7,080	2,777	23,648	60,290	4,303	36,642
3	4,170	3,365	40,150	49,755	0,805	9,605
4	6,500	3,723	25,758	44,970	2,777	19,213
5	4,810	1,966	34,808	85,160	2,844	50,353
6	3,570	0,625	46,898	267,880	2,945	220,982

Tabla 3.4. Distancia calculada en las imágenes donde había paso peatonal y el error cometido respecto a la distancia real.

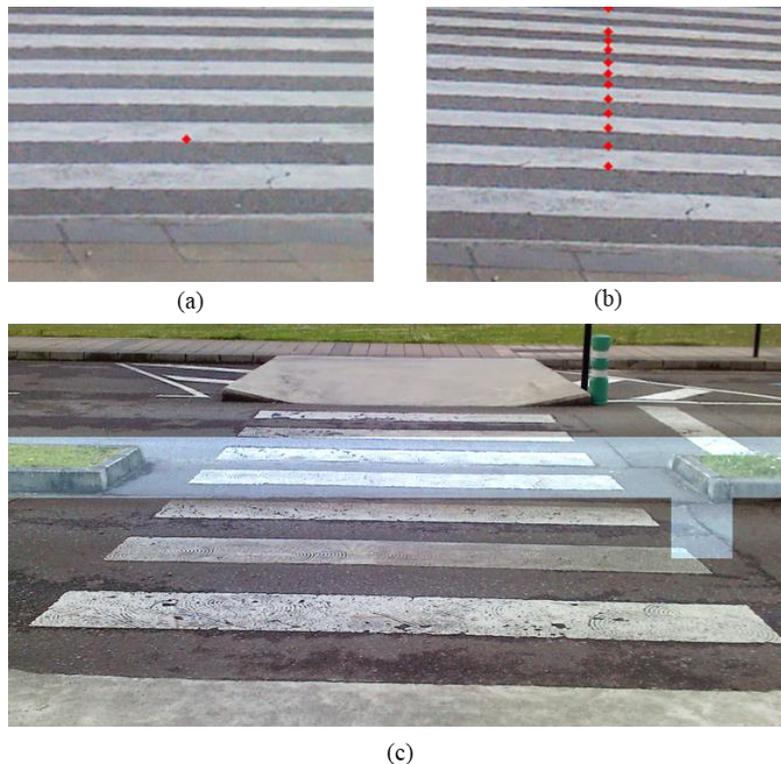


Figura 3.14. Alguno resultado experimentales: (a) y (b) vista izquierda y derecha respectivamente de un mismo paso peatonal con su punto más cercano definido en ambas y la imagen (c), zona en la que se encontró un paso de cebrado, pero la cual no le comprende en su totalidad.

Existe un gran error en una gran parte de las imágenes. En el caso de la segunda, se ve en la figura 3.14(a)(b) que el punto está situado en el mismo sitio en ambas vistas, pero el error en metros es de 4 y en píxeles de 36. Si este valor se compara con los errores obtenidos en las pruebas de entorno controlado el error en píxeles es ligeramente superior.

En la imagen 3 aparece un error que no es debido a esta etapa del algoritmo de decisión, si no a la de detección. El candidato definido como paso de peatones si lo es, pero la región que se supone lo comprende no lo hace por completo, tal y como se muestra en la figura 3.14(c). Esto se deriva en que el punto respecto al cual el algoritmo calcula la distancia no es el mismo del que se sacó al momento de realizar la base de datos.

Otro caso relevante es el que se muestra en la imagen 6, en donde el error es de 220,982 píxeles. Al momento de realizar el *template matching* y buscar el pixel de mayor valor no se obtuvo el punto correcto. La banda en donde este se encuentra es muy regular a lo largo de toda la imagen (figura 3.15), por lo que para el código casi todos los puntos de la parte baja de esa banda son iguales a la plantilla.

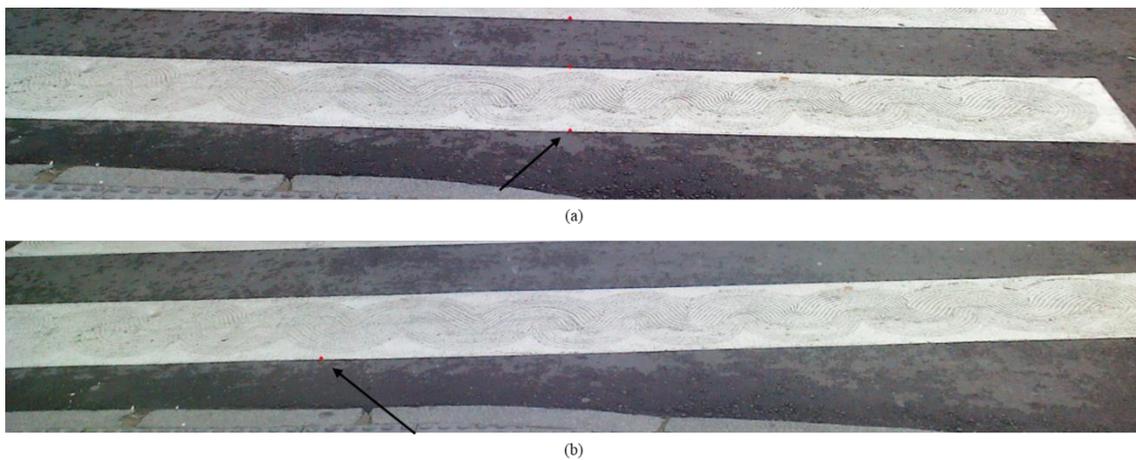


Figura 3.15. Vista derecha (a) e izquierda (b) de un paso de peatones, en el cual el proceso de *template matching* falló al buscar el punto que se indica con la flecha en la imagen (a) en la vista (b).

Como la idea de este programa es que lo puedan usar personas invidentes cuando estén cerca de pasos de peatones eso implica que ha de funcionar a tiempo real. La media del tiempo que tardaron en procesarse cada una de los 15 pares de imágenes estéreo fue de 0,178 segundos, permitiendo así un funcionamiento aproximado de 5 fotogramas por segundo.

## 4.- Discusión.

Se alcanzó el objetivo, aunque los resultados no son lo suficientemente satisfactorios como para que este sistema sea aprovechado por algún usuario con dificultades en la visión, la detección de solo un 57,14% de pasos de peatones es un valor muy bajo, y no es admisible la obtención de falsos positivos, no se debería detectar un paso de peatones en sitios donde no existen, es algo que podría afectar la seguridad del usuario.

Los valores que se determinan de forma experimental en la primera etapa se obtuvieron analizando cómo cambian los resultados de 5 imágenes distintas al variar los valores de estos parámetros, si se consideraran más imágenes al momento establecer estos valores, sería posible acercarlos más a su valor óptimo y por lo tanto aumentar el rendimiento del sistema. También se podría variar el algoritmo en función de las características del candidato que se está analizando, por ejemplo, si este está en la mitad superior de la imagen sería conveniente no aplicar el proceso morfológico, el cierre, ya que en estos casos las bandas del paso de cebra están muy próximas entre sí y el realizar el cierre sobre ellas podría unir las, perdiéndose información.



Figura 4.1. Imagen en el cual los candidatos dividen al paso de cebra en tres partes, dificultando los procesos posteriores

En ninguno de los casos se considera que una persona o un vehículo pueda estar sobre el paso de peatones, durante ese tiempo es muy probable que no se detecte nada, lo que sería un problema si el invidente se desplaza por una zona muy transitada. Otra mejora que se podría estudiar es el unir dos candidatos si estos están en contacto, y analizarlos con

el programa tanto separadas como juntas y decidir en función de los resultados. La figura 4.1 es una imagen en la que esto podría dar buenos resultados.

A veces, en las regiones candidatas, puede no haber pasos peatonales, sino que t hay otras marcas en el pavimento, como pueden ser flechas, líneas que delimitan zonas de estacionamiento o cualquier otro tipo de símbolo vial. Estos elementos pueden alterar el comportamiento del código. Una solución, es que después del umbralizado se realice un estudio del factor de forma de las figuras presentes en la imagen y que solo se mantengan aquellas que tengan forma de rectángulo.

En un principio se iba a hacer una tercera etapa antes de las dos propuestas en este proyecto. Con esta los que se buscaba era mejorar la imagen de entrada, de forma que la detección funcionase mejor ante perturbaciones. Como todas las imágenes son tomadas en el exterior es muy probable que muchos pasos de peatones tengan sombras encima de ellos, afectando el resultado de los procesos posteriores. Un método que busca solucionar esto es el propuesto por Drew y Finlayson [24] [25] en el cual logran obtener un imagen invariante a la iluminación, suponiendo que la imagen con la que parten es tomada por una cámara de tres sensores (R,G,B) con presencia de luz Planckiana y compuesta de superficies Lambert. Bajo esta luz y ciertas temperaturas los cuerpos negros emiten ondas en el espectro de lo visible. Se sabe de Álvarez y López [26], [27] quienes se basaron en estos trabajos que los rayos del sol son aproximadamente este tipo de luz y que la superficie de las carreteras son mayoritariamente lambertianas, las cuales son superficies ideales que reflejan la energía incidente en ellas desde un punto hacia todas las direcciones de igual manera, es decir que si se varía el punto de vista desde el cual se ve esta superficie su luminancia no cambia.

Se empieza generando un mapa de cromaticidad 2D de la imagen original, el cual consiste en tomar dos de los canales de la imagen (R, G o B) y normalizarlos por el tercero, para después aplicar logaritmos a ambos y graficar todos los valores obtenidos. A medida que la temperatura de la luz Planckiana cambia, hay vector de cromaticidad que se mueve aproximadamente en una línea recta, la cual es independiente de la magnitud y dirección de la luz, esta recta está definida por el ángulo de invarianza a la iluminación, parámetro

que depende de cada cámara. Si se proyectan todos los puntos de la imagen sobre esta recta se obtiene una imagen en escala de grises invariante a la iluminación. El problema surge al analizar las imágenes que se obtienen con este método, no es posible usarlas en las siguientes dos etapas, ya que las bandas del paso de peatones se oscurecen y se perdía esta característica de bipolaridad, en la cual se basa la fase de detección.

Por esta razón no se incluyó esta etapa en la propuesta final. Como trabajo futuro se podría buscar una solución al problema de las sombras, pero sin que afectase negativamente a las demás etapas.

En cuanto al cálculo de la distancia a la que se encuentra el paso de peatones se registró un error máximo de 4 metros. Este error se puede deber a una mala calibración de la cámara, al momento de realizar las fotografías del tablero o algún error en el montaje de las mismas, quedando mal alineadas en la base donde se dispusieron. En cualquier caso, hay que tener en cuenta que, al momento de aplicar el programa en una situación real, mientras más se acerca el usuario al paso de peatones, menor será el error en metros de la distancia. Se podría también probar una configuración estereó de las cámaras diferente, que de mejores resultados para este tipo de aplicación.

El proceso de *template matching* no fue exacto en algunos casos, como el que se describió en el apartado de resultados, donde la banda blanca es uniforme a lo largo de toda su extensión. En vez de esta herramienta se podría usar un método basado en SIFT [28], que es un proceso que extrae características o puntos relevantes de la imagen, pero por la misma razón que falló la técnica propuesta esta también podría fallar. Si en la primera etapa se detectara siempre el mismo punto, como la esquina de una de las bandas, este error se eliminaría, ya que al procesar las imágenes de las dos vistas se obtendría el mismo punto en ambas.

## 5.- Conclusión.

La creación de un sistema para detectar pasos de peatones, basado en la visión artificial, es viable. En el método aquí planteado el sistema acierta en la mitad de los casos, mientras que métodos publicados por otras personas obtienen mejores resultados, aunque se desconoce las características de todas las imágenes que usaron. Se logró desarrollar el objetivo planteado, con una propuesta para obtener una imagen en la cual se muestra únicamente la región que comprende un paso de peatones siempre y cuando exista uno en la imagen, y se adquiriera información acerca de la posición en la que este se encuentra, específicamente la distancia a la que está.

Se dejan planteados posibles caminos a tomar para ampliar y mejorar lo aquí propuesto. Finalmente, se destaca el procedimiento que se usó para llegar a una solución, el cual consistió en dividir el problema general en unos más sencillos, hasta solo tener tareas simples. Para lograr esto, se investigaron distintos documentos publicados con el propósito de resolver problemas parecidos y aprovecharlos para resolver el objetivo de trabajo.

En lo que corresponde al desarrollo personal y las competencias adquiridas, se ha tenido que recurrir al aprendizaje autónomo ya que temas como la visión artificial en tres dimensiones o el uso de las invarianzas geométricas, no son contenidos que se expliquen durante la carrera. Además de la programación en C++, el uso de herramientas como las librerías de OpenCV o trabajar en un sistema operativo Linux.

## 6.- Referencias.

- [1] Y. Zhai, G. Cui, Q. Gu, y L. Kong, «Crosswalk Detection Based on MSER and ERANSAC», en *2015 IEEE 18th International Conference on Intelligent Transportation Systems*, 2015, pp. 2770-2775.
- [2] M. Radványi, B. Varga, y K. Karacs, «Advanced crosswalk detection for the Bionic Eyeglass», en *2010 12th International Workshop on Cellular Nanoscale Networks and their Applications (CNNA 2010)*, 2010, pp. 1-5.
- [3] M. Radvanyi, G. E. Paziienza, y K. Karacs, «Crosswalks recognition through CNNs for the bionic camera: Manual vs. automatic design», en *European Conference on Circuit Theory and Design, 2009. ECCTD 2009*, 2009, pp. 315-318.
- [4] L. Kék, K. Karacs, y T. Roska, *Cellular Wave Computing Libray*. Budapest, Hungría: CELLULAR SENSORY WAVE COMPUTERS LABORATORY, HUNGARIAN ACADEMY OF SCIENCES, 2007.
- [5] S. Milan, H. Vaclav, y B. Roger, «Template matching», en *Image Processing, Analysis, and Machine Vision*, Cuarta., CENGAGE Learning, pp. 233-235.
- [6] T. Asami y K. Ohnishi, «Crosswalk location, direction and pedestrian signal state extraction system for assisting the expedition of person with impaired vision», en *2014 10th France-Japan/ 8th Europe-Asia Congress on Mechatronics (MECATRONICS)*, 2014, pp. 285-290.
- [7] S. Tadayoshi, W. Haiyuan, N. Naoki, y K. Suguru, «Measurement of the length of pedestrian crossings and detection of traffic lights from image data», p. 9, de agosto de 2002.
- [8] M. S. Uddin y T. Shioyama, «Bipolarity and Projective Invariant-Based Zebra-Crossing Detection for the Visually Impaired», en *2005 IEEE Computer Society*

*Conference on Computer Vision and Pattern Recognition (CVPR'05) - Workshops*, 2005, pp. 22-22.

- [9] I. Weiss, «Geometric invariants and object recognition», p. 43, Agosto 1992.
- [10] «Manual de vados y pasos peatonales». 26-oct-2016.
- [11] «Manual Accesibilidad para técnicos municipales». 26-oct-2016.
- [12] C. Galamhos, J. Matas, y J. Kittler, «Progressive probabilistic Hough transform for line detection», en *Proceedings. 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No PR00149)*, 1999, vol. 1, p. 560 Vol. 1.
- [13] S. Milan, H. Vaclav, y B. Roger, «Hough transform», en *Image Processing, Analysis, and Machine Vision*, Cuarta., CENGAGE Learning, pp. 210-217.
- [14] Y. M. Mustafah, R. Noor, H. Hasbi, y A. W. Azma, «Stereo vision images processing for real-time object distance and size measurements», en *2012 International Conference on Computer and Communication Engineering (ICCCE)*, 2012, pp. 659-663.
- [15] M. Jernej y V. Damir, «Distance measuring based on stereoscopic pictures», presentado en 9th International PhD Workshop on Systems and Control: Young Generation Viewpoint, Izola, Eslovenia, 2008, p. 6.
- [16] S. Milan, H. Vaclav, y B. Roger, «Thresholding», en *Image Processing, Analysis, and Machine Vision*, Cuarta., CENGAGE Learning, pp. 179-187.
- [17] S. Milan, H. Vaclav, y B. Roger, «Edge detectors», en *Image Processing, Analysis, and Machine Vision*, Cuarta., CENGAGE Learning, pp. 133-139.

- [18] S. Milan, H. Vaclav, y B. Roger, «Camera model», en *Image Processing, Analysis, and Machine Vision*, Cuarta., CENGAGE Learning, pp. 598-601.
- [19] J. Heikkila y O. Silven, «A four-step camera calibration procedure with implicit image correction», en *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1997, pp. 1106-1112.
- [20] Z. Zhang, «A flexible new technique for camera calibration», *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, n.º 11, pp. 1330-1334, nov. 2000.
- [21] S. Milan, H. Vaclav, y B. Roger, «Two cameras, stereopsis», en *Image Processing, Analysis, and Machine Vision*, Cuarta., CENGAGE Learning, pp. 609-619.
- [22] S. Milan, H. Vaclav, y B. Roger, «Median filtering», en *Image Processing, Analysis, and Machine Vision*, Cuarta., CENGAGE Learning, pp. 130-133.
- [23] S. Milan, H. Vaclav, y B. Roger, «Mathematical morphology», en *Image Processing, Analysis, and Machine Vision*, Cuarta., CENGAGE Learning, pp. 684-721.
- [24] M. S. Drew y G. D. Finlayson, «Recovery of Chromaticity Image Free from Shadows via Illumination Invariance», presentado en ICCV 03 Workshop on Color and Photometric Methods in Computer Vision, Nice, Francia, 2003, pp. 32-39.
- [25] G. D. Finlayson, M. S. Drew, y L. Cheng, «Intrinsic Images by Entropy Minimization». Springer-Verlag Berlin Heidelberg 2004, 2004.
- [26] J. M. Á. Alvarez y A. M. Lopez, «Road Detection Based on Illuminant Invariance», *IEEE Trans. Intell. Transp. Syst.*, vol. 12, n.º 1, pp. 184-193, mar. 2011.

- [27] J. M. Alvarez, A. Lopez, y R. Baldrich, «Illuminant-invariant model-based road segmentation», en *2008 IEEE Intelligent Vehicles Symposium*, 2008, pp. 1175-1180.
- [28] S. Milan, H. Vaclav, y B. Roger, «SIFT», en *Image Processing, Analysis, and Machine Vision*, Cuarta., CENGAGE Learning, pp. 474-477.

# **ANEXOS**

# Índice de los anexos.

1.-	Parámetros para el procesamiento de las imágenes.....	55
2.-	Tablas de resultados.....	56
3.-	Normativa de los pasos peatonales .....	59
4.-	Materiales adjuntados.....	61
4.1.-	Base de datos – Cámara estéreo.....	61
4.2.-	Base de Datos – Cámara individual.....	61
4.3.-	Base de Datos de prueba – Cámara estéreo.....	61
4.4.-	Calibración.....	61
4.5.-	Códigos.....	62
4.5.1.-	Detección y medición de pasos de peatones.....	62
4.5.2.-	Detección y medición de pasos de peatones – Paso a paso.....	62
4.5.3.-	Distancia a un objeto.....	62
4.6.-	Video.....	62

# 1.- Parámetros para el procesamiento de las imágenes.

En la siguiente tabla se muestra el valor de los parámetros usados en el algoritmo, tanto para la cámara individual como la estéreo. Estos valores pueden coincidir.

Concepto	Cámara individual	Cámara estéreo
Escalado de la imagen	1/4	1
Alpha ( $\alpha$ )	0,6	0,6
Dimensión del bloque	136	80
Relación de distancia mínima	0,7	0,7
Bipolaridad mínima	0,65	0,65
THP <sup>2</sup> – Número de votos mínimos	500	250
THP – Longitud mínima	500	250
THP – Distancia máxima para conectar puntos	90	35

---

<sup>2</sup> Siglas para referirse a la transformada de Hough probabilística.

## 2.- Tablas de resultados.

A continuación, se muestran dos tablas detalladas con todos los resultados obtenidos tras procesar las bases de datos. En las dos siguientes tablas se muestran los resultados de la etapa de detección, tanto con las imágenes tomadas por las cámaras individual como con la estéreo. En verde están los resultados correctos y en rojo los incorrectos.

Individual				
Imagen	Pasos de peatones bien detectados	Paso de peatones no detectado	Paso de peatones detectado donde no había	No se detecta y no hay paso de peatones
1	1	0	0	0
2	1	0	0	0
3	1	0	0	0
4	1	0	0	0
5	0	1	0	0
6	1	0	0	0
7	1	0	0	0
8	1	0	0	0
9	1	0	0	0
10	1	0	0	0
11	0	1	0	0
12	0	0	1	0
13	0	1	0	0
14	1	0	0	0
15	0	1	0	0
16	0	0	0	1
17	0	0	0	1
18	0	0	0	1
19	0	1	0	0
20	0	1	0	0
21	0	0	1	0
22	0	0	1	0
23	0	0	0	1
24	0	0	0	1
25	0	1	0	0
26	1	0	0	0
27	0	1	0	0

Imagen	Pasos de peatones bien detectados	Paso de peatones no detectado	Paso de peatones detectado donde no había	No se detecta y no hay paso de peatones
28	1	0	0	0
29	1	0	0	0
30	1	0	0	0
31	0	1	0	0
32	0	1	0	0
33	0	1	0	0
34	1	0	0	0
35	1	0	0	0
36	0	1	0	0
37	1	0	0	0
38	1	0	0	0
39	1	0	0	0
40	0	1	0	0
41	0	1	0	0
<b>Total</b>	<b>19</b>	<b>14</b>	<b>3</b>	<b>5</b>

Estéreo				
Imagen	Pasos de peatones bien detectados	Paso de peatones no detectado	Paso de peatones detectado donde no había	No se detecta y no hay paso de peatones
1	1	0	0	0
2	1	0	0	0
3	0	1	0	0
4	1	0	0	0
5	1	0	0	0
6	0	1	0	0
7	0	1	0	0
8	0	1	0	0
9	1	0	0	0
10	1	0	0	0
11	0	1	0	0
12	1	0	0	0
13	1	0	0	0
14	0	1	0	0
15	0	1	0	0
<b>Total</b>	<b>8</b>	<b>7</b>	<b>0</b>	<b>0</b>

En esta tabla se muestran los resultados de la segunda etapa tras procesar las imágenes estéreo, al igual que el tiempo total de ejecución del programa para cada una de ellas.

Decisión Imagen	Distancia real (m)	Disparidad real (píxeles)	Distancia detectada (m)	Disparidad detectada (píxeles)	Error (m)	Error (píxeles)	Tiempo de ejecución (s)
1			2,215	75,587			0,17052
2			3,977	42,098			0,15322
3							0,22374
4	6,120	27,357	4,069	41,146	2,051	13,789	0,20709
5	7,080	23,648	2,777	60,290	4,303	36,642	0,18953
6	4,400	38,051					0,21076
7	5,200	32,197					0,17956
8	3,720	45,007					0,16252
9	4,170	40,150	3,365	49,755	0,805	9,605	0,22754
10	6,500	25,758	3,723	44,970	2,777	19,213	0,23488
11	3,620	46,250					0,16672
12	4,810	34,808	1,966	85,160	2,844	50,353	0,17720
13	3,570	46,898	0,625	267,880	2,945	220,982	0,12058
14	6,000	27,904					0,14470
15	3,420	48,955					0,09856
<b>Total</b>							<b>0,17781</b>

### 3.- Normativa de los pasos peatonales

Los pasos de peatones se definen como la zona de intersección entre la circulación rodada y el tránsito peatonal, es decir, la parte del itinerario peatonal que cruza la calzada de circulación de vehículos al mismo o a diferente nivel. En el Reino de España, independientemente de las normas de ámbito estatal o autonómico que estén en vigor, los elementos que lo constituyen deben cumplir una serie de características y objetivos básicos.

El ancho de los pasos de peatones se debe corresponder con el de los vados<sup>3</sup> definidos en las aceras, o en su ausencia, con la señalización de la acera con pavimento diferenciado. El tercio central debe tener una anchura igual a la parte central del vado correspondiente, siendo la mínima admisible de 1,80m. La directriz del paso de peatones se define como la línea que une los puntos medios de sus bandas de señalización, la cual se unirá a su vez con los puntos centrales de los vados existentes a ambos lados. Esta directriz será preferiblemente perpendicular la acera, a menos que la situación no lo permita, ambos casos se muestran en la figura A-3.1.

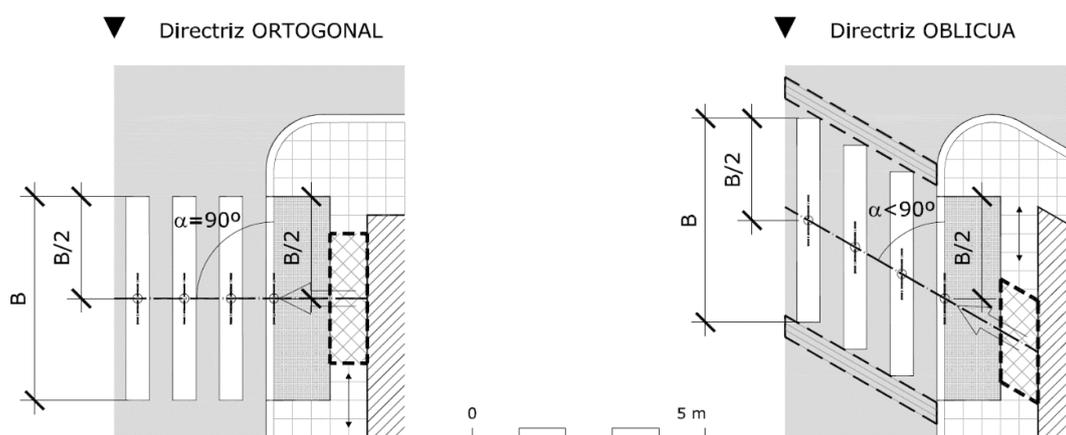


Figura A-3.1. En la imagen de la izquierda el paso de peatones tiene la directriz perpendicular a la acera (Ortogonal), en la de la derecha la directriz esta inclinada (oblicua). La anchura se define con la letra B. Imagen de [11].

<sup>3</sup> Los vados peatonales son las modificaciones de las zonas de un itinerario peatonal, mediante planos inclinados que comunican niveles diferentes, facilitando a los peatones el cruce de las calzadas destinadas a la circulación de vehículos.

Si la directriz es oblicua, el paso se localiza muy al extremo coincidiendo con curvas en las aceras o si está elevado respecto a la carretera, se debe incorporar una banda longitudinal con textura hecha de pavimento. Lo que se busca con esta banda es acotar el espacio de cruce, impidiendo que personas con discapacidad visual abandonen el paso de cebra por sus laterales.

La señalización de los pasos peatonales consiste en bandas paralelas a la acera pintadas sobre la calzada, de 0,50m de ancho y separadas entre sí una distancia de 0,50m. En la figura A-3.2 se ven los componentes y medidas más relevantes de un paso de cebra. La pintura deberá resistir a la intemperie, mantener el color (generalmente blanco), no ser deslizante con la lluvia y ser resistente al desgaste producto del tráfico rodado.

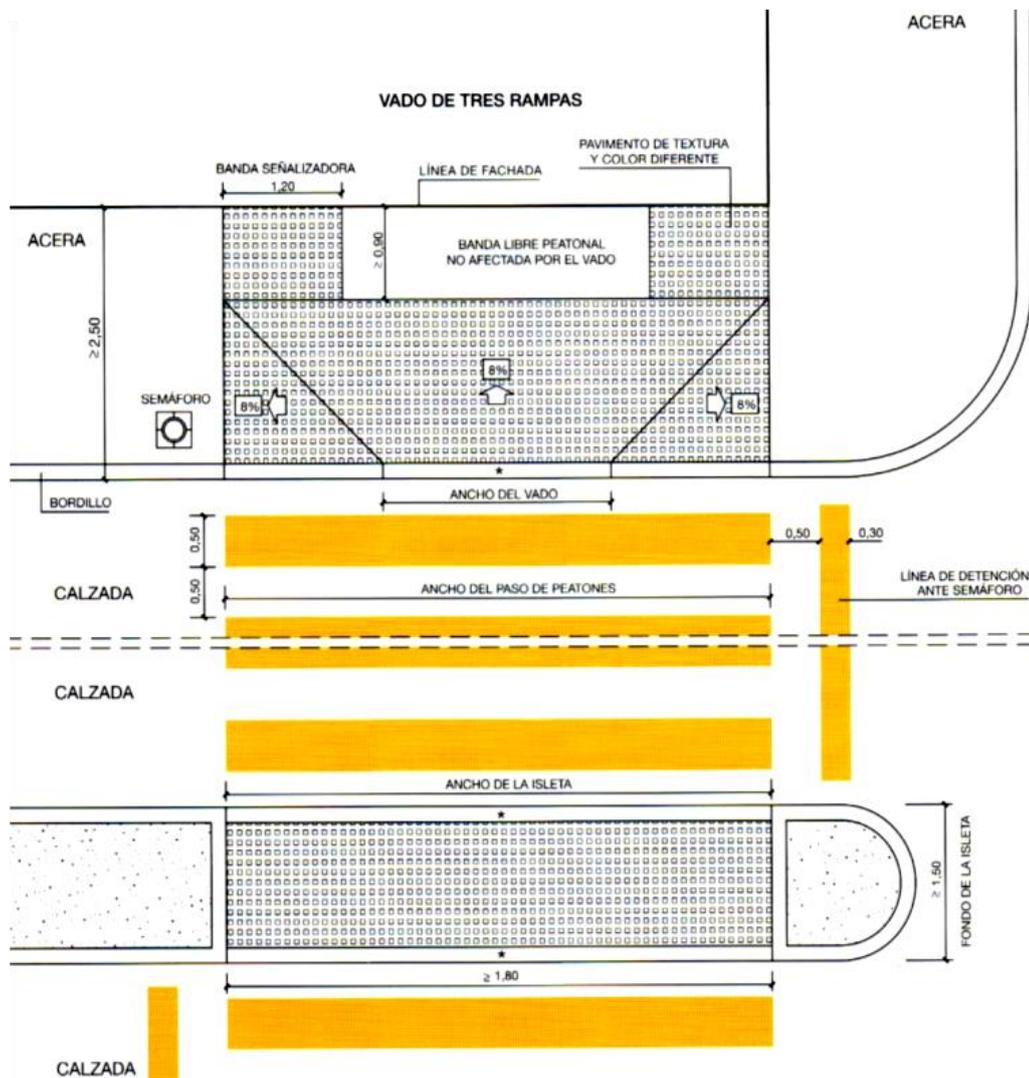


Figura A-3.2. Cruce con semáforo en isleta. Medidas y componentes básicos de los pasos peatonales. Imagen de [10].

## **4.- Materiales adjuntados.**

Junto a este documento se adjuntan diferentes carpetas en las cuales se encuentran aquellos archivos que se usaron para el desarrollo del proyecto. Su contenido se describe a continuación:

### **4.1.- BASE DE DATOS – CÁMARA ESTÉREO.**

Contiene 15 pares de imágenes (derecha e izquierda), en las cuales aparecen diferentes pasos de peatones. Se conoce la distancia a la que 12 de estos se encuentran. Los resultados obtenidos para cada imagen se encuentran en el anexo, coincidiendo la numeración de la tabla con la de los nombres de las imágenes.

### **4.2.- BASE DE DATOS – CÁMARA INDIVIDUAL.**

Está formada por 41 imágenes con diferentes escenas, tanto con pasos de cebra como sin ellos. Los resultados obtenidos para cada imagen se encuentran en el anexo, coincidiendo la numeración de la tabla con la de los nombres de las imágenes.

### **4.3.- BASE DE DATOS DE PRUEBA – CÁMARA ESTÉREO.**

Posee las imágenes usadas en el proceso de obtención de distancias sin perturbaciones. Siendo un total 4 tríos de imágenes, donde dos corresponden a las que se capturan con las cámaras estéreo y la otra es la que actúa como plantilla en el proceso de template matching. Los resultados obtenidos para cada imagen se encuentran en la tabla, coincidiendo la numeración de la tabla con la de los nombres de las imágenes.

### **4.4.- CALIBRACIÓN.**

En esta carpeta se almacenan las imágenes usadas para las calibraciones de las cámaras, y los archivos generados por MATLAB en donde se guardan los resultados de las mismas.

## 4.5.- CÓDIGOS.

Se adjuntan los diferentes algoritmos desarrollados durante el proyecto. Tanto en archivos de texto con la extensión .cpp o .h como en formato PDF para facilitar su lectura en caso de consulta.

### 4.5.1.- Detección y medición de pasos de peatones.

En esta carpeta está el algoritmo propuesto como solución. En él se muestran directamente los resultados y el tiempo que tarda en analizar un par de imágenes.

### 4.5.2.- Detección y medición de pasos de peatones – Paso a paso.

Código de la solución propuesta, el cual muestra por pantalla las diferentes etapas por las que se pasa hasta llegar al resultado final.

### 4.5.3.- Distancia a un objeto.

Código en el cual se comprueba, el funcionamiento del método para obtener la distancia a la que se encuentra un objeto.

## 4.6.- VIDEO.

Se realizó un video en formato MP4, que ilustra las diferentes etapas del método propuesto. Las dos imágenes usadas corresponden con las número 4 de la base de datos de las cámaras estéreo.

# **PRESUPUESTO**

# Índice del presupuesto.

1.-	Coste de ejecución material.....	65
1.1.-	Coste de mano de obra.....	65
1.2.-	Costes de equipos.....	66
1.3.-	Costes de software.....	66
1.4.-	Coste total de presupuesto de ejecución material.....	67
2.-	Gastos generales y beneficio industrial.....	68
3.-	Importe total.....	69

# 1.- Coste de ejecución material.

El coste de ejecución material incluye tres categorías, coste de equipos, coste de software y coste de mano de obra por el tiempo empleado en el proyecto. Se considera una amortización de los equipos de 3 años.

## 1.1.- COSTE DE MANO DE OBRA.

A la realización del proyecto se destina un empleado con un salario bruto de 1500 € al mes. La jornada laboral es de 8 horas al día y 5 días a la semana. En la tabla mostrada a continuación se estima la duración del proyecto

Concepto	Cantidad
Estudio del problema	10 días
Estudio de las librerías de OpenCV	5 días
Creación de las bases de datos	1 día
Desarrollo del algoritmo de la primera etapa	12 días
Desarrollo del algoritmo de la segunda etapa	5 días
Pruebas de funcionamiento	2 días
Creación de la documentación	5 días
<b>Total</b>	<b>40 días</b>
	<b>2 meses</b>

Concepto	Precio unitario	Cantidad	Costo
Salario mensual	1500 €	2	3000 €
Seguridad social a cargo de la empresa	31,25 %		937,5 €
<b>Subtotal</b>			<b>3937,5 €</b>

En la siguiente tabla se muestra el desglose del gasto asociado a la seguridad social por parte de la empresa. Este gasto tiene un porcentaje que oscila entre 31% y 35% del salario bruto mensual.

Concepto	Porcentaje
Contingencias generales	23,60 %
Contingencias profesionales	
Accidentes de trabajo y enfermedades profesionales	1,35%
Desempleo	5,50%
Formación profesional	0,60%
Fondo garantía profesional	0,20%
<b>Total de seguridad social</b>	<b>31,25%</b>

### 1.2.- COSTES DE EQUIPOS.

El proyecto tiene una duración estimada de 40 días lo que corresponde a un 5,56% de la amortización global a 3 años de los equipos. Al final de su vida útil no se espera obtener ningún beneficio con su venta.

Concepto	Precio unitario	Cantidad	Costo
Dell XPS 15 9560 - Intel i7 7700HQ	1.827,68 €	0,056	102,35 €
USB 2 uEye XS	300 €	0,111	33,3 €
Base para cámaras en estéreo	15 €	0,056	0,84 €
<b>Subtotal</b>			<b>136,49 €</b>

### 1.3.- COSTES DE SOFTWARE.

Concepto	Precio unitario	Cantidad	Costo
Sistema operativo Linux – Ubuntu 16.04	0 €	1	0 €
Librerías OpenCV 3.2.0	0 €	1	0 €
MATLAB	2000 €/año	0,167	334 €
Microsoft Office 365	57 €/año	0,167	9,52 €
<b>Subtotal</b>			<b>343,52 €</b>

#### 1.4.- COSTE TOTAL DE PRESUPUESTO DE EJECUCIÓN MATERIAL.

Concepto	Subtotal
Costes de mano de obra	3937,5 €
Costes de equipos	136,49 €
Costes de software	343,52 €
<b>Subtotal</b>	<b>4417,51 €</b>

## 2.- Gastos generales y beneficio industrial.

Los gastos generales son los gastos obligatorios que surgen de la utilización de las instalaciones de trabajo. El beneficio industrial son las ganancias que percibe la empresa por realizar el proyecto.

Concepto	Porcentaje	Subtotal
Gastos generales	13%	574,28 €
Beneficio industrial	6%	265,05 €
<b>Subtotal</b>		<b>839,33 €</b>

### 3.- Importe total.

Concepto	Subtotal
Coste total del presupuesto de ejecución material	4417,51 €
Gastos generales y beneficio industrial	839,33 €
Subtotal	5256,84 €
IVA 21%	1103,94 €
Total	6360,78 €

El coste total del proyecto suma la cantidad de:

**SEIS MIL TRESCIENTOS SESENTA EUROS CON SETENTA Y OCHO  
CENTIMOS**