



UNIVERSIDAD MICHOACANA DE
SAN NICOLÁS DE HIDALGO

FACULTAD DE INGENIERÍA ELÉCTRICA
DIVISIÓN DE ESTUDIOS DE POSGRADO

**“Estimación de la Profundidad Utilizando
Imágenes Estereoscópicas”**

Por:
Iván González López

Una tesis presentada en cumplimiento de los requisitos
para la obtención de grado de
Maestro en Ciencias de la Ingeniería Eléctrica

Director de Tesis
Dr. Félix Calderón Solorio

MORELIA, MICHOACÁN

Agosto 2016



Este trabajo está dedicado a mis padres y hermanos, por su apoyo incondicional y por estar conmigo en todo momento.

A mis compañeros de posgrado, por que fue toda una experiencia cursar este trayecto a su lado, y por que nunca faltó su apoyo.

A mi asesor Félix Calderón, por todo el conocimiento y el tiempo que invirtió en mi.

Y por último, a la DEP FIE, por haberme ofrecido tal oportunidad de inmesurable valor.

Contenido

| | |
|--|----|
| Dedicatoria | 1 |
| Contenido | 3 |
| Resumen | 5 |
| Abstract | 7 |
| Lista de Figuras | 9 |
| Lista de Tablas | 13 |
| Lista de Símbolos | 15 |
| | |
| 1. Introducción | 1 |
| 1.1. Visión Estereoscópica Artificial | 6 |
| 1.2. Propuesta de Solución | 12 |
| 1.3. Objetivos de la Tesis | 12 |
| 1.3.1. Objetivo general | 12 |
| 1.3.2. Objetivos particulares | 12 |
| 1.4. Descripción de Capítulos | 13 |
| | |
| 2. Modelos de visión estereoscópica | 15 |
| 2.1. Modelo pinhole | 16 |
| 2.2. Modelo pinhole para cámaras estéreo | 20 |
| 2.2.1. Modelo en coordenadas de la imágenes | 24 |
| 2.2.2. Parámetros extrínsecos del sistema | 26 |
| 2.3. Cálculo de la disparidad | 28 |
| 2.4. Conclusión del capítulo | 31 |
| | |
| 3. Construcción y Calibración de un Sistema Estéreo | 33 |
| 3.1. Características del sistema estereoscópico | 33 |
| 3.1.1. Características de las cámaras | 34 |
| 3.1.2. Características de la estructura del sistema | 35 |
| 3.1.3. Ensamblado final del sistema | 39 |
| 3.2. Calibración del sistema estéreo | 40 |
| 3.2.1. Patrón de calibración para el sistema estéreo | 40 |
| 3.2.2. Calibración con OpenCV | 44 |
| 3.3. Imágenes anaglifo como marco de referencia para estimar profundidad | 52 |
| 3.3.1. Imágenes anaglifo a partir de un sistema estéreo | 53 |

| | | |
|-------------|--|-----|
| 3.3.2. | Ejemplo de imágenes anaglifo creadas con el sistema estéreo | 54 |
| 3.4. | Marco de referencia para la estimación de la profundidad con el sistema estéreo | 56 |
| 3.5. | Reducir la distancia entre los centros de las cámaras de manera artificial . . | 59 |
| 3.5.1. | Relación disparidad-profundidad con α como nuevo parámetro del modelo de las cámaras | 66 |
| 3.6. | Conclusiones del capítulo | 71 |
| 4. | Estimación de la Profundidad Basado en Imágenes-Estéreo | 73 |
| 4.1. | Correspondencia y mapa de disparidad | 74 |
| 4.1.1. | Métodos Basados en Puntos Característicos | 75 |
| 4.1.2. | Métodos de Búsqueda Basados en Correlación | 77 |
| 4.1.3. | Obtención de la correspondencia utilizando métodos de búsqueda ba- sados en correlación | 79 |
| 4.1.4. | Cálculo de los mapas de disparidad | 82 |
| 4.1.5. | Geometría Epipolar | 83 |
| 4.1.6. | Obtención del Mapa de Disparidad, Implementando Geometría Epipolar | 86 |
| 4.2. | Estimación de la profundidad. | 87 |
| 4.3. | Conclusiones del Capítulo | 91 |
| 5. | Resultados | 93 |
| 5.1. | Resultado para la generación de Imágenes Anaglifo. | 93 |
| 5.2. | Resultados y comparativas entre los diferentes algoritmos para el cálculo de los mapas de disparidad. | 94 |
| 5.3. | Resultados para la Estimación de la Profundidad | 104 |
| 5.4. | Conclusiones del Capítulo | 110 |
| 6. | Conclusiones y Trabajos Futuros | 113 |
| 6.1. | Conclusiones | 113 |
| 6.2. | Trabajos Futuros | 114 |
| Referencias | | 117 |

Resumen

Con el objetivo de estimar la profundidad de los objetos en una escena a partir de un par de imágenes tomadas con un sistema estereoscópico de cámaras, en esta tesis se muestran las etapas de desarrollo de un algoritmo para la estimación de la profundidad, utilizando como herramienta básica, un sistema estereoscópico construido a partir de un par de cámaras web, tratando de asemejar las características del sistema estereoscópico humano. Partiendo de esto, se desarrolla el proceso de construcción de el sistema estéreo de cámaras utilizado para esta tesis, y se mencionan sus características principales.

Como parte de las etapas de desarrollo de este algoritmo de estimación, se analizan las características del modelo matemático de una cámara convencional, conocido como modelo Pin-Hole, y como dicho modelo es aplicado en un sistema de cámaras estéreo, a través de un proceso de calibración, con la finalidad de obtener información característica del sistema, y con la intención de mejorar los resultados de estimación de la distancia.

Con el sistema de cámaras calibrado, se propone la implementación de una técnica conocida como imágenes anaglifo, con el objetivo de evaluar visualmente los resultados del proceso de calibración, y para llevar a cabo algunos ajustes al modelo de relación Disparidad - Profundidad, para crear un marco de referencia que facilite la estimación de la profundidad a partir de los pares de imágenes tomadas con el sistema.

Se presenta también un análisis y comparación entre diferentes medidas de distancia basadas en la correlación para la búsqueda de correspondencia, y el cálculo de mapas de disparidad. Se analizan además, diferentes técnicas para mejorar los resultados en el cálculo de mapas de disparidad, tales como la implementación de líneas epipolares, y la aplicación de esquemas de color como RGB, HSV, y Escala de grises.

Finalmente, con los mapas de disparidad, y el marco de referencia establecido a partir de los ajustes al modelo de relación Disparidad - Profundidad, se presenta la propuesta de un algoritmo para la estimación de la profundidad, utilizando un mapa de colores conocido como Hot Colormap (mapa de colores caliente), para asignar una tonalidad de color en particular a los píxeles de una imagen, para resaltar a que distancia se encuentran en base a su valor de disparidad, logrando de esta forma estimar la profundidad de los objetos, con respecto de las cámaras de forma muy clara.

Palabras clave: estereoscópico, estéreo, anaglifo, disparidad, profundidad.

Abstract

In order to estimate the depth of the different objects in a scene, captured with a stereo camera system, in this thesis are presented the different stages of development of an algorithm for estimating the depth, using as a basic tool, a stereo system built from a pair of usb webcams, trying to resemble the characteristics of human stereoscopic system. Therefore, the process of building the stereo camera system used for this thesis is developed, and the main features of this, are mentioned.

As part of the development stages of this estimation algorithm, the characteristics of the mathematical model of a conventional camera, known as Pin-Hole model, are analyzed. Also, this model is applied to a stereo system through a calibration process, in order to obtain characteristic information of this system, and with the intention to improve results in the depth estimation. With the calibration and with the information related to the stereo system's model, the implementation of a technique known as anaglyph images is proposed, with the intention of carrying out some adjustments to the Disparity - Depth model, in order to create a framework reference to facilitate the estimation of the depth, with the stereo system.

It is also presented, an analysis and comparison between different methods to obtain correspondences between pixels based on the correlation, in order to obtain dense disparity maps. Different techniques are analyzed, with the intention of getting such disparity maps, techniques like the application of different color schemes such as RGB, HSV, or Grayscale, and the calculation and implementation of epipolar lines.

Finally, with the disparity-depth model's framework established, and the calculation of disparity maps, an algorithm for estimating depth is presented, using a color scale that identifies a range of distance in specific, from a shade of color in particular, with the intention of highlighting how far the objects are, taking as reference the system's cameras.

Finally, with the depth disparity model and with the generated disparity maps, an algorithm for estimating the depth is presented, right through a color scheme that defines for a distance range of a particular color, to estimate how far is an object from the cameras, in an approximate way.

Lista de Figuras

| | | |
|-------|--|----|
| 1.1. | Interacción entre un haz de luz y el ojo humano. | 2 |
| 1.2. | Como se puede apreciar en esta imagen, algunos animales que pueden ser catalogados como presas, tiene los ojos localizados a los lados de la cabeza. | 3 |
| 1.3. | En esta imagen se puede apreciar como animales catalogados como depredadores, tiene los ojos al frente de la cabeza. | 3 |
| 1.4. | Comparación del campo de visión de una Paloma y un Búho. La paloma es considerada un animal tipo presa, mientras que el búho es considerado depredador. Como se puede apreciar en la comparación, la visión monocular le permite a la paloma tener una perspectiva de casi 360 grados, mientras que para el búho, el campo de visión binocular es más amplio, lo que facilita el proceso de estereopsis. Imagen descargada de Wikipedia. | 4 |
| 1.5. | Sistema binocular natural del ser humano. | 5 |
| 1.6. | Sistema estereoscópico artificial. | 5 |
| 1.7. | Imagen del modelo de visión Binocular de Descartes. | 6 |
| 1.8. | En la figura se muestra un ejemplo básico de como funciona la medida parallaxa. Como se puede observar, se tienen dos vistas diferentes de un objeto en el espacio, cuando la tierra se encuentra exactamente en la posición opuesta de su órbita, con respecto del sol, para lo cual, tienen que pasar 6 meses. Imagen descargada de Wikipedia. | 7 |
| 1.9. | Un estereograma, es una imagen generada a partir de dos imágenes con desplazamientos laterales de elementos. | 8 |
| 1.10. | Estereoscópio de Wendell Holmes, funcional a partir de una imagen estereoscópica. | 8 |
| 1.11. | Ejemplo del funcionamiento de los anaglifos. | 9 |
| 1.12. | Cámara Fujifilm FinePix Real 3D W3 | 10 |
| 1.13. | Cámara Panasonic Lumix DMC-GH2 con lente para imágenes 3D. | 10 |
| 1.14. | Cámara estereoscópica Holga con dos lentes de 60mm. | 11 |
| 1.15. | Videocámara Panasonic FULL HD 3D Camcorder para grabar video en 3D con alta definición. | 11 |
| 1.16. | Videocámara JVC Everio GS-TD1B Full HD 3D. | 11 |
| 2.1. | Representación de una imagen digital. | 16 |
| 2.2. | Comportamiento del Modelo PinHole usando una caja. | 17 |

| | | |
|-------|--|----|
| 2.3. | Modelo matemático PinHole. Imagen original descargada de Wikipedia. . . | 18 |
| 2.4. | Modelo estereoscópico para el plano X, Z | 20 |
| 2.5. | Modelo estereoscópico para el plano Z, Y | 23 |
| 2.6. | Planos coordenados de imagen para el sistema estereoscópico. | 24 |
| 2.7. | La diferencia de perspectivas por la distancia entre los ojos, es lo que permite medir la profundidad de los objetos en una escena. | 29 |
| 3.1. | Dimensiones de la cámara Logitech C920. | 35 |
| 3.2. | Modelo de cámaras PinHole estéreo. | 35 |
| 3.3. | Ejemplo básico del proceso llevado a cabo por la calibración, al obtener la información necesaria para alinear las cámaras de un sistema estereoscópico de forma artificial. | 37 |
| 3.4. | Posición de las cámaras en el sistema. | 38 |
| 3.5. | Diseño de la caja para el empotrado de las cámaras. | 38 |
| 3.6. | Presentación del ensamblado del modelo. | 39 |
| 3.7. | Ensamblado final del modelo. | 39 |
| 3.8. | Imágenes del tablero de ajedrez con distintas orientaciones. Gary Bradski <i>et</i> <i>al.</i> [Bradski08] | 41 |
| 3.9. | Patrón de tablero de ajedrez impreso en papel opalina opaco, pegado sobre papel cascaron blanco y acrílico, para mantener la rigidez. | 42 |
| 3.10. | Muestra de 24 pares de imágenes usadas para la calibración. | 43 |
| 3.11. | Detección de esquinas en el tablero de ajedrez con FindChessboardCorners(). | 46 |
| 3.12. | Comparativa entre imágenes tomadas por el sistema estéreo, antes y después de calibrar y rectificar dicho sistema. | 49 |
| 3.13. | Punto P_1 en el espacio tridimensional con coordenadas $[10.0 \text{ cm}, 5.0 \text{ cm}, 20.0 \text{ cm}]^T$. | 51 |
| 3.14. | Mapeo del punto P_1 a coordenadas de la imagen izquierda y derecha respec- tivamente, en comparación con las coordenadas calculadas con el modelo de las cámaras. | 52 |
| 3.15. | Funcionamiento de las gafas especiales para observar imágenes anaglifo. Imagen de Vicente Fuentes [Fuentes14b] | 53 |
| 3.16. | Ejemplo de creación de imágenes anaglifo a partir de un par de imágenes estereoscópicas. | 54 |
| 3.17. | Anaglifo generado a partir de imágenes con el sistema estéreo calibrado. . . | 54 |
| 3.18. | En este anaglifo es posible observar como las componentes de color de los objetos que se encuentran a mayor profundidad, se juntan o tienen una menor disparidad. Por el contrario, se puede observar que para el objeto más cercano (la cinta métrica), las componentes de color están completamente separadas, indicando que hay una mayor disparidad. | 56 |
| 3.19. | Anaglifo generado a partir del sistema de cámaras implementado para este trabajo de tesis. | 57 |
| 3.20. | Esta gráfica muestra de forma clara como la curva se comporta de forma asintótica respecto del eje de la profundidad Z , es decir, conforme la dispari- dad disminuye, la profundidad se aproxima a cero, pero nunca alcanza este valor. | 58 |

| | |
|---|----|
| 3.21. Modelo de las cámaras antes y después de aplicar un desplazamiento a los centros de cámaras con la variable α , en la Figura 3.5, y 3.5 respectivamente. | 60 |
| 3.22. Desplazamiento en las imágenes tomadas con el sistema (imágenes izquierda y derecha respectivamente), después de modificar los centros de la cámaras. | 63 |
| 3.23. Coordenadas tridimensionales del punto observado. | 64 |
| 3.24. Comparación entre los puntos calculados con los parámetros del sistema con el valor de $\alpha = 38$ pixeles, y las coordenadas de imagen el punto observado. | 65 |
| 3.25. Comparación entre el anaglifo de la Figura 3.17, con un anaglifo generado a partir del modelo con un desplazamiento de centros de $\alpha = 38$ pixeles, donde claramente se puede ver el corrimiento que existe entre las dos imágenes. . . | 66 |
| 3.26. Comparación entre un anaglifo sin desplazamiento de centros, y un anaglifo con un desplazamiento de $\alpha = 38$ pixeles. | 67 |
| 3.27. Anaglifo de un cañón antiguo. [Fuentes14a] | 68 |
| 3.28. Esta gráfica muestra como la curva corta el eje de la profundida, en $Z = 120$ centímetros, indicando que la disparidad en ese punto es $x_l - x_r = 0$. Esto sucede debido a que se añadió el valor de $\alpha = 38$ pixeles, al valor de los centros de las cámaras. | 69 |
| 3.29. Imágenes tomadas de un objeto a diferentes distancias del sistema de cámaras. | 70 |
| 3.30. Posición de los objetos del experimento de la Figura 3.29, en relación a la gráfica de la Figura 3.28 | 71 |
| 3.31. En esta Figura se observa el acercamiento al anaglifo de la Figura 3.29(1), donde claramente se puede apreciar que el objeto a una distancia de 120 centímetros, se puede visualizar a la perfección sin ningún tipo de descomposición de color, indicando así, que la disparidad a esa distancia, es igual a cero. | 72 |
| | |
| 4.1. Ejemplo de la correspondencia de un pixel en la imagen derecha, y su empate con un pixel en la imagen izquierda. | 74 |
| 4.2. Funcionamiento básico de las búsquedas basadas en correlación. | 78 |
| 4.3. Modelo de la geometría epipolar. Robert Laganier [Laganier11] | 84 |
| 4.4. En esta figura, se aprecia como las líneas epipolares pasan por los mismos puntos. | 85 |
| 4.5. En esta figura se muestra la escala de color que se tomo como base para diferenciar las diferentes profundidades. | 89 |
| | |
| 5.1. Anaglifo tomado con el sistema estéreo usando el modelo del sistema modificado. | 94 |
| 5.2. Anaglifo al exterior del edificio del Omega 2 en la DEP-FIE, tomado con el sistema estéreo usando el modelo del sistema modificado. | 94 |
| 5.3. Comparación entre un anaglifo sin desplazamiento de centros (imagen inciso a), y un anaglifo con desplazamiento de $\alpha = 38$ pixeles, aplicado a los centros. Escena: laboratorio de computo del edificio Omega 2, en la DEP-FIE. . . . | 95 |
| 5.4. Imágenes ejemplo para llevar a cabo las diferentes pruebas de cálculo de disparidad. | 96 |

| | |
|--|-----|
| 5.5. Resultados del cálculo de la disparidad con las diferentes medidas de distancia, y utilizando las imágenes a color. | 97 |
| 5.6. Resultados del cálculo de la disparidad con los diferentes métodos, y utilizando las imágenes en escala de grises. | 98 |
| 5.7. Imágenes ejemplo en formato de color HSV. | 100 |
| 5.8. Resultados del cálculo de la disparidad con las diferentes Medida de Distancia, utilizando el esquema de color HSV. | 101 |
| 5.9. Resultados del cálculo de la disparidad con las diferentes medidas de distancia, implementando líneas epipolares. | 102 |
| 5.10. En la figura se muestra dos objetos, una copa de cobre, y una figurilla de arcilla. Ambos objetos fueron colocados a una distancia aproximada de entre 115.0 cm y 120.0 cm con respecto de las cámaras, con la finalidad de probar el funcionamiento del modelo. | 104 |
| 5.11. Mapa de disparidad en x , de las imágenes de la Figura 5.10, utilizando SSDN (4.2). | 104 |
| 5.12. Resultados de aplicar el Algoritmo 7, coloreando únicamente los objetos con una profundidad de 120 centímetros con respecto de las cámaras. | 105 |
| 5.13. En la figura se muestra un nuevo par de imágenes de los objetos usados anteriormente, pero esta vez con la figurilla de arcilla a 40. centímetros, y la copa de cobre a 120.0 centímetros. | 105 |
| 5.14. Mapa de disparidad a partir de las imágenes de la Figura 5.13, usando el método SSDN (4.2). | 106 |
| 5.15. Profundidad calculada con el Algoritmo 7, donde se puede observar claramente la diferencia entre los colores de los objetos en la escena, indicando que se encuentran a diferentes profundidades. | 106 |
| 5.16. En este diagrama se muestra la secuencia de pasos necesaria para poder llevar a cabo la estimación de la profundidad, de los objetos en una escena, a partir de un par de imágenes estereoscópicas. | 107 |
| 5.17. En la figura se observan 3 objetos, una figura de arcilla a una distancia de 40 centímetros, una caja a 80 centímetros, y una copa de cobre a 115 centímetros de las cámaras. | 108 |
| 5.18. Resultados de Disparidad Usando el Método SSDN (4.2). | 108 |
| 5.19. Resultados de la estimación de la profundidad, a partir de las disparidad mostrada en la figura 5.18 | 109 |
| 5.20. En la figura se observa una caja a 18 centímetros, la figurilla de arcilla a 50 centímetros aproximadamente, y una copa de cobre a 65 centímetros de las cámaras. | 110 |
| 5.21. Resultados de Disparidad Usando el Método SSDN (4.2). | 110 |
| 5.22. Resultados de la estimación de la Profundidad a partir de la disparidad. . . | 111 |

Lista de Tablas

| | |
|--|-----|
| 4.1. Tabla de relación entre la escala de color, y la distancia. | 89 |
| 5.1. Tabla de tiempos del cálculo de la disparidad, utilizando las imágenes a color. | 96 |
| 5.2. Tabla de tiempos del cálculo de la disparidad, utilizando las imágenes en escala de grises. | 99 |
| 5.3. Tabla de tiempos del cálculo de la disparidad, utilizando las imágenes con el modelo HSV. | 100 |
| 5.4. Tabla de tiempos del cálculo de la disparidad, implementando líneas epipolares. | 103 |
| 5.5. Tabla de tiempos de ejecución, para la estimación de la profundidad, a partir de las imágenes de la Figura 5.11 y de la Figura 5.14 respectivamente. . . . | 106 |

Lista de Símbolos

Mayúsculas

- \hat{C} — Matriz de valores de similitud.
- C_x — Centro de una cámara en x .
- C_y — Centro de una cámara en y .
- Cl_x — Centro de la cámara izquierda en x .
- Cl_y — Centro de la cámara izquierda en y .
- Cr_x — Centro de la cámara derecha en x .
- Cr_y — Centro de la cámara derecha en y .
- CR — Matriz de correspondencias.
- D_x — Mapa de disparidad en x .
- D_y — Mapa de disparidad en y .
- F — Matriz fundamental de un sistema de cámaras.
- I — Matriz de una imagen.
- I_A — Imagen anaglifo.
- I_R — Imagen tomada por la cámara derecha.
- I_L — Imagen tomada por la cámara izquierda.
- I_r — Imagen derecha.
- I_l — Imagen izquierda.
- K — Matriz de parámetros intrínsecos.
- K_l — Matriz de parámetros intrínsecos de la cámara izquierda.
- K_r — Matriz de parámetros intrínsecos de la cámara derecha.
- N — Número de imágenes muestra.
- N_r — Número de renglones en una imagen.
- N_c — Número de columnas en una imagen.
- P — Representación de un punto en coordenadas tridimensionales.
- P_1 — Representación de un punto en coordenadas tridimensionales.
- R — Matriz de rotación de una cámara.
- R_r — Matriz de rotación de la cámara derecha.
- R_l — Matriz de rotación de la cámara izquierda.
- T — Vector de traslación de un sistema de cámaras.
- T_l — Vector de traslación de la cámara izquierda.
- T_r — Vector de traslación de la cámara derecha.
- T_f — Vector de traslación optimizado.

| | | |
|-------|---|--|
| W | — | Tamaño de una ventana de píxeles. |
| X | — | Eje coordenado tridimensional de anchura. |
| Y | — | Eje coordenado tridimensional de altura. |
| Z | — | Eje coordenado tridimensional de profundidad. |
| X_l | — | Coordenada tridimensional en X en la cámara izquierda. |
| Y_l | — | Coordenada tridimensional en Y en la cámara izquierda. |
| X_r | — | Coordenada tridimensional en X en la cámara derecha. |
| Y_r | — | Coordenada tridimensional en Y en la cámara derecha. |
| X_0 | — | Coordenada de anchura en el espacio tridimensional. |
| Y_0 | — | Coordenada de altura en el espacio tridimensional. |
| Z_0 | — | Coordenada de profundidad en el espacio tridimensional. |
| Z' | — | Profundidad definida para el marco de referencia de profundidad. |
| V | — | Ventana de píxeles. |
| MD | — | Medida de distancia. |

Minúsculas

| | | |
|-----------|---|---|
| a | — | Primer coeficiente para la ecuación de una línea recta. |
| b | — | Segundo coeficiente para la ecuación una línea recta. |
| b | — | Tercer coeficiente para la ecuación una línea recta. |
| d | — | Distancia entre el centro de dos cámaras. |
| nr | — | Número de renglones en una imagen. |
| nc | — | Número de columnas en una imagen. |
| x | — | Coordenada en x en el plano de la cámara. |
| y | — | Coordenada en y en el plano de la cámara. |
| x_l | — | Coordenada de imagen en x de la cámara izquierda. |
| y_l | — | Coordenada de imagen en y de la cámara izquierda. |
| x_r | — | Coordenada de imagen en x de la cámara derecha. |
| y_r | — | Coordenada de imagen en y de la cámara derecha. |
| \hat{x} | — | Coordenada de imagen en x de una ventana muestra. |
| \hat{y} | — | Coordenada de imagen en y de una ventana muestra. |
| f | — | Distancia focal de una cámara. |

Letras griegas

| | | |
|----------|---|--|
| α | — | Desplazamiento de centro de cámaras(+, -). |
|----------|---|--|

Glosario de Términos

HSV modelo HSV (del inglés Hue, Saturation, Value – Matiz, Saturación, Valor), también llamado HSB (Hue, Saturation, Brightness – Matiz, Saturación, Brillo), define un modelo de color en términos de sus componentes.

RGB (sigla en inglés de red, green, blue, en español «rojo, verde y azul») o RVA (sigla preferida por la ASALE y la RAE) es la composición del color en términos de la intensidad de los colores primarios de la luz.

Capítulo 1

Introducción

Como seres humanos, la forma de interactuar con nuestro entorno se lleva a cabo a través de diferentes medios conocidos como los cinco sentidos. Entre estos sentidos, uno de los más importantes es el sentido de la vista, ya que a través de este sentido, un individuo es capaz de recopilar información respecto de su entorno. Por otro lado, según Gary Bradski *et al.* [Bradski08], todo comienza con un haz de luz emitido como un rayo desde alguna fuente origen (como un foco, una lámpara, o el mismo sol), que viaja por el entorno chocando con diferentes objetos en el camino. Cuando la luz impacta con algún objeto en particular, parte de la energía es absorbida por dicho objeto, mientras que el resto es reflejada y es lo que el ojo humano percibe como color e iluminación como se muestra en la Figura 1.1. Gracias a esto, no solo el ser humano, si no también la mayoría de los animales, tienen la capacidad de reconocer su entorno. Ahora bien, otra característica importante del sentido de la vista, es la capacidad de calcular las distancias a las que se encuentran diferentes objetos. Esta característica tan particular, es conocida como visión estereoscópica, o visión binocular. Por definición, la palabra binocular proviene de dos raíces latinas, ‘bini’ doble, y ‘oculus’ ojo, lo cuál hace referencia al tipo de visión que hace uso de los dos ojos en conjunto. Por otro lado, el término estéreo de estereoscopia, proviene del griego estéreo, que significa ‘dos’, y scopía, que significa ‘examinar’, haciendo referencia a sistemas que implementan el uso de dos canales de comunicación para la obtención de información. Ahora bien, en conjunto con la visión binocular, una tercera característica del sentido de la vista es el campo de

visión. El campo de visión consiste en superponer las imágenes visualizadas por el sistema estereoscópico, con la finalidad de obtener un panorama tridimensional del entorno. Pero para que esto sea posible, debe existir una distancia entre cada ojo de tal forma que cada receptor ocular perciba una imagen distinta del ambiente.

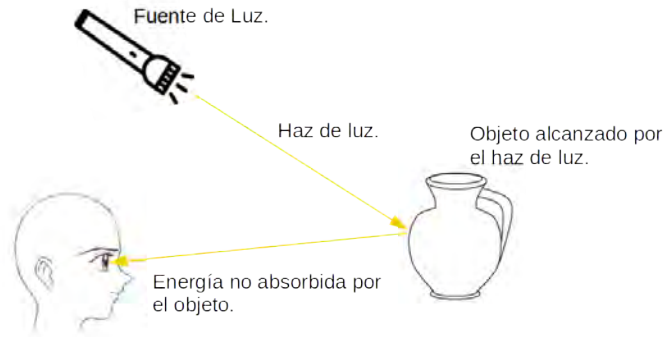


Figura 1.1: Interacción entre un haz de luz y el ojo humano.

En general en el reino animal, los animales catalogados como presas, tienen los dos ojos localizados en las zonas laterales de la cabeza con la finalidad de obtener un campo de visión lo más amplio posible, para determinar si se encuentran en peligro en algún momento dado. Como ejemplo, se pueden observar animales como los venados, las cebras, o los antílopes (por mencionar algunos). Normalmente, este tipo de animales tienen la capacidad de mover los ojos de forma independiente, lo que les permite ampliar su campo de visión de forma considerable, incluso, algunas aves pueden llegar a tener un campo de visión de hasta 360 grados, sin necesidad de mover los ojos. Un ejemplo de la posición de los ojos en este tipo de animales, se muestra en la Figura 1.2.

Por otro lado, los animales normalmente catalogados como depredadores, tienen situados los ojos al frente de la cabeza, por ejemplo: el águila, el león, la serpiente, etc. (Figura 1.3). También, otros animales que no necesariamente deben ser incluidos en la categoría de depredadores, como los murciélagos de frutas, o los primates (entre ellos el hombre), tienen los ojos colocados en una posición frontal en la cabeza. La Figura 1.4 muestra una comparación entre un animal dentro de la clasificación de las presas, y un



Figura 1.2: Como se puede apreciar en esta imagen, algunos animales que pueden ser catalogados como presas, tiene los ojos localizados a los lados de la cabeza.

animal catalogado como depredador. En general, para este tipo de animales, es necesario tener la capacidad de discriminar en base a la profundidad de los objetos, por tanto, la visión binocular mejora la capacidad de agarrar una fruta, capturar una presa o alcanzar una rama, por mencionar algunos ejemplos. Por tanto, a pesar de que la posición de los ojos de este tipo de animales, reduce el campo de visión a solo 180 grados generalmente, incrementa las condiciones a favor de un proceso llamado estereopsis.

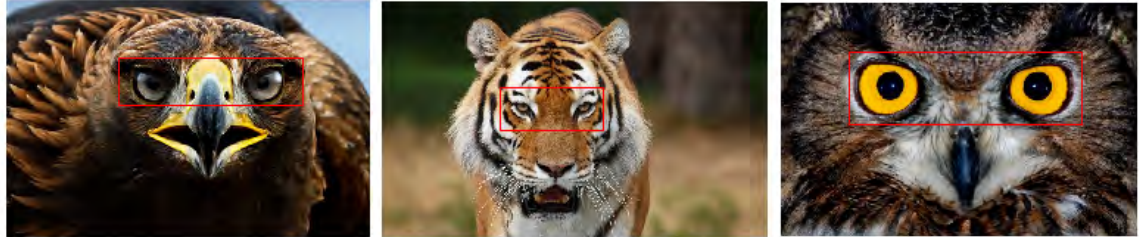


Figura 1.3: En esta imagen se puede apreciar como animales catalogados como depredadores, tiene los ojos al frente de la cabeza.

Según Robert Laganier [Lecumberry14], el proceso de estereopsis consiste en tres etapas. Dichas etapas son: seleccionar un punto característico en una de las imágenes, buscar el punto correspondiente al seleccionado en la otra imagen, y por último calcular la diferencia en las posiciones de los dos puntos seleccionados. Dicho procedimiento, es llevado a cabo por el cerebro del individuo, y permite calcular la posición espacial de los objetos en la escena. Con esta información es entonces posible triangular la posición de un objeto en el espacio.

Como se mencionó anteriormente, el ser humano entra en la categoría de animales

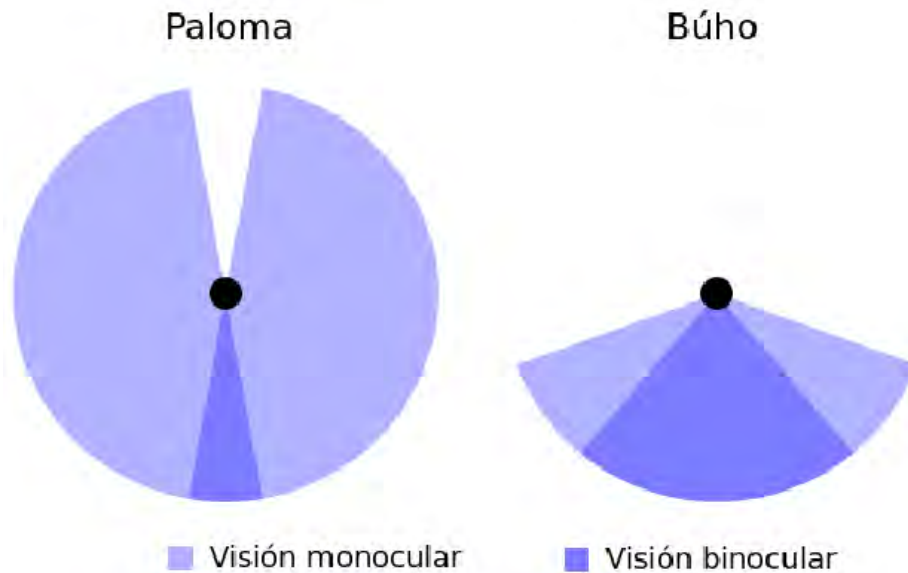


Figura 1.4: Comparación del campo de visión de una Paloma y un Búho. La paloma es considerada un animal tipo presa, mientras que el búho es considerado depredador. Como se puede apreciar en la comparación, la visión monocular le permite a la paloma tener una perspectiva de casi 360 grados, mientras que para el búho, el campo de visión binocular es más amplio, lo que facilita el proceso de estereopsis. Imagen descargada de Wikipedia.

que cuentan con un panorama de visión que favorece la visión binocular, y por medio del cuál según Patricia Compañ [Compañ04], obtienen tres cuartas partes de la información de su entorno. Por tal motivo, existe un amplio interés en el simular el modelo de visión binocular del hombre, con un sistema de cámaras estereoscópicas y una computadora. Partiendo de esto, es posible elaborar una comparación entre el sistema binocular humano, y uno simulado con una par de cámaras y una computadora.

Como se muestra en la Figura 1.5, el ser humano tiene un sistema binocular formado por dos ojos, establecidos a una distancia aproximada de 6.5 cm de distancia tomando como punto de medida la retina ocular. Gracias a la distancia existente entre las retinas oculares (sensores ópticos del sistema), el cerebro humano es capaz de triangular un objeto en el espacio, y determinar la profundidad de este por medio de la estereopsis. Los canales por medio de los cuales se comunican los ojos con el cerebro, son conocidos como nervios ópticos.

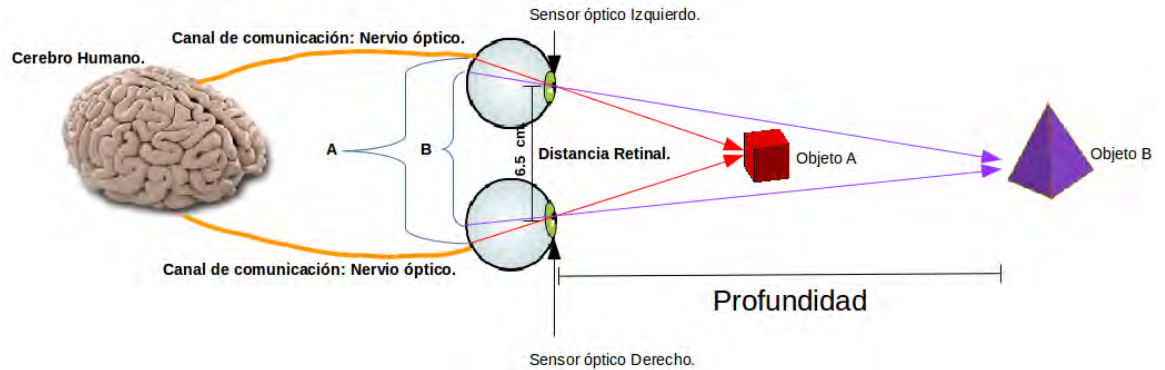


Figura 1.5: Sistema binocular natural del ser humano.

Por otro lado, en la Figura 1.6 se muestra un modelo de visión artificial. En comparación con el modelo natural del ser humano, se puede apreciar que existen muchas similitudes. En primer lugar se tienen las cámaras que simulan los ojos, implementando además la misma distancia entre ellas, existente en el sistema binocular humano. Para este caso, el centro de cada una de las cámaras captura la luz de la escena, mientras que la computadora recibe esta información y la procesa, efectuando la función de cerebro. La comunicación entre dispositivos se da por medio de los cables USB de cada cámara, simulando los nervios ópticos.

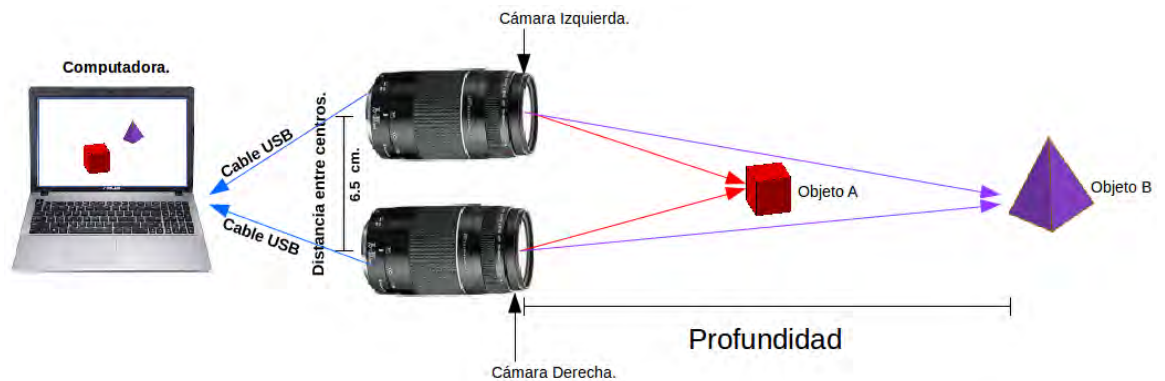


Figura 1.6: Sistema estereoscópico artificial.

1.1. Visión Estereoscópica Artificial

Desde hace años, la visión estereoscópica ha sido un tema de interés para estudiosos como Descartes (Figura 1.7), Euclides, Da Vinci, o Kepler, los cuales son considerados como pioneros en el campo de la visión binocular.

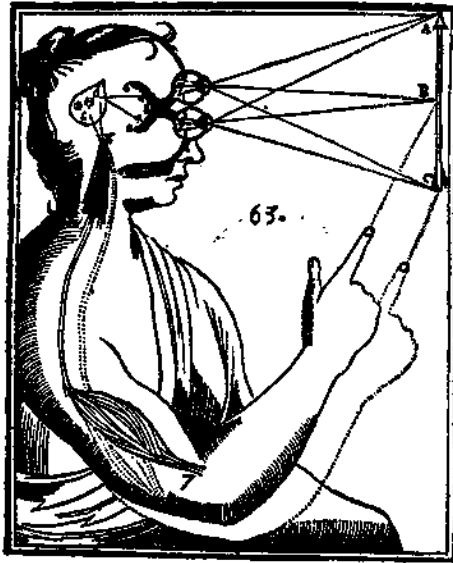


Figura 1.7: Imagen del modelo de visión Binocular de Descartes.

Por su parte, en el año 1838, el astrónomo Friedrich Bessel, fue el primero en lograr medir el Paralaje (Parallax por su traducción al inglés), de la estrella 61 Cygni. El paralaje es un desplazamiento o diferencia en la aparente posición de un objeto a lo largo de dos perspectivas, y es medido por un ángulo o semiángulo de inclinación entre dos líneas. El término se deriva de la palabra griega *parallaxis*, que significa alteración. De esta forma, en astronomía, el paralaje se usa para calcular distancias de estrellas (u otros objetos), en unidades astronómicas (UA), y esto sucede gracias a las diferentes posiciones orbitales de la tierra. Cuando el planeta tierra se encuentra exactamente en lados opuestos con respecto del sol en su órbita, se obtienen dos vistas diferentes de una estrella o un objeto, estableciendo un marco de referencia, para el cálculo de la distancia de dicha estrella u objeto, como se muestra en la Figura 1.8.

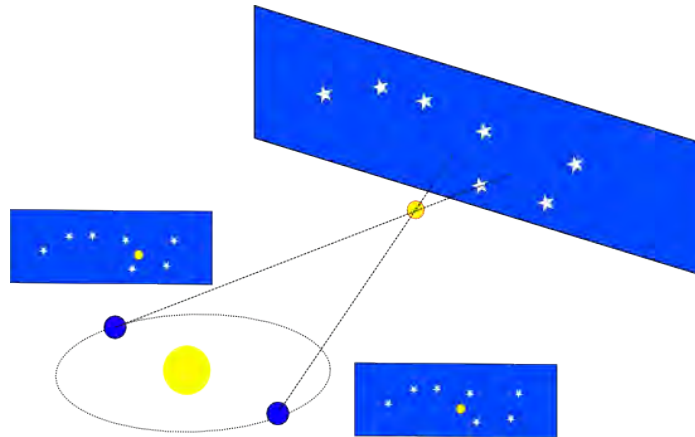


Figura 1.8: En la figura se muestra un ejemplo básico de como funciona la medida paralaxa. Como se puede observar, se tienen dos vistas diferentes de un objeto en el espacio, cuando la tierra se encuentra exactamente en la posición opuesta de su órbita, con respecto del sol, para lo cual, tienen que pasar 6 meses. Imagen descargada de Wikipedia.

Por otro lado, a mediados del siglo 19, científicos como Charles Wheatstone creador de los estereogramas (Figura 1.9), Sir David Brewster, Oliver Wendell Holmes (Figura 1.10) o Don Santiago Ramón y Cajal, comenzaron a trabajar más a fondo en el área de la visión binocular con el objetivo de crear distintos dispositivos capaces de generar el efecto de profundidad o relieve, implementando dos imágenes de la misma escena, pero con desplazamientos laterales en su posición horizontal.

Ya en el siglo 20, alrededor de los años 30's, resurgió la visión estereoscópica pero aplicada a películas en cámaras con rollos de 35 mm. como la *realistic* o la *ViewMaster*. Sin embargo, no fue hasta los años 50's que se intentó la comercialización de las películas 3D sin tener mucho auge. Por otro lado, ya en los años 80's, se obtuvieron resultados más espectaculares como el formato IMAX (alta resolución en imágenes gigantes), lo que incrementó su popularidad, Patricia Compañ [Compañ04]. Cabe mencionar que para lograr el efecto 3D en el formato IMAX, se implementó una técnica conocida como *anaglifo*. Según Brian Coe [Coe78], esta técnica fue descubierta y patentada por Louis Ducos du Hauron en 1868, que a diferencia del estereoscopio de Wendell Holmes, el cuál usaba dos imágenes distintas de la escena vistas cada una por un ojo diferente, los *anaglifos* son imágenes compuestas por dos capas de color sobrepuestas y con un desplazamiento ligero una con

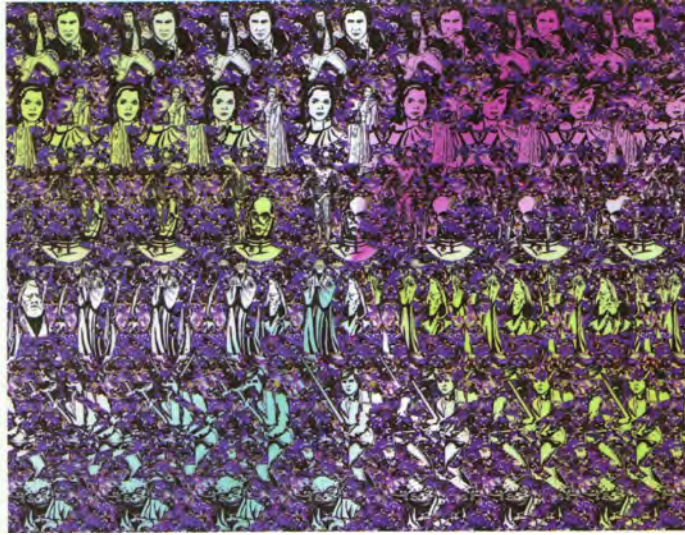


Figura 1.9: Un estereograma, es una imagen generada a partir de dos imágenes con desplazamientos laterales de elementos.



Figura 1.10: Estereoscópio de Wendell Holmes, funcional a partir de una imagen estereoscópica.

respecto de la otra en forma horizontal. Pero dado que son dos capas de color distintas, para poder apreciar el efecto 3D, es necesario utilizar lentes anaglifo, los cuales se encargan de filtrar una capa de color distinta para cada ojo. Con estas gafas, el cerebro fusiona la información recibida por los ojos en una misma percepción, y genera el efecto tridimensional (Figura 1.11).

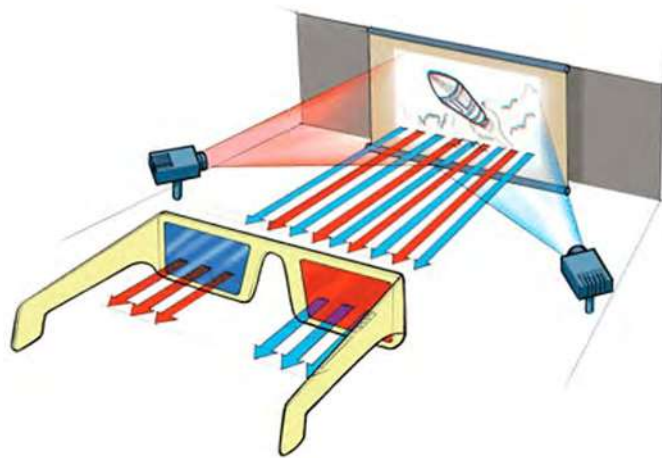


Figura 1.11: Ejemplo del funcionamiento de los anaglifos.

Ahora bien, con el constante avance en el desarrollo tecnológico tanto de cámaras, como de hardware computacional, es hasta años más recientes que se ha visto a las imágenes tridimensionales como una forma de esparcimiento. Sin embargo, a pesar del increíble avance que se ha dado en el campo de la fotografía, en la actualidad las cámaras estereoscópicas se han popularizado muy poco y por tanto sus costos son muy elevados. Por ejemplo, algunas de las cámaras más comunes son, la Fujifilm FinePix Real 3D W3 de la empresa Fujifilm con un costo de aproximadamente 6 mil pesos (Figura 1.12), la Panasonic Lumix DMC-GH2 que incorpora un lente para imágenes 3D con un costo aproximado de 12 mil pesos (Figura 1.13), y también la Holga, que es una cámara no digital más sencilla y hecha completamente de plástico (Figura 1.14), con un costo aproximado de 2 mil quinientos pesos.

Por otro lado, en la industria del cine y la televisión, la creación de secuencias de video tridimensionales se encuentra en auge debido a la experiencia que esta brinda al usuario. Este tipo de trabajos son creados con cámaras estereoscópicas especialmente manufacturadas para la grabación de secuencias de video tridimensionales y con calidad de alta definición. Sin embargo los costos siguen siendo poco accesibles, los dispositivos poco comunes y dado que sus costos son muy elevados, solo estudios profesionales de producción tienen la solvencia económica para adquirir una. Algunos ejemplos de este tipo de cámaras



Figura 1.12: Cámara Fujifilm FinePix Real 3D W3



Figura 1.13: Cámara Panasonic Lumix DMC-GH2 con lente para imágenes 3D.

son por ejemplo la Panasonic FULL HD 3D Camcorder con un costo aproximado de 378 mil pesos (Figura 1.15), o la JVC Everio GS-TD1B Full HD 3D con un costo aproximado de 27 mil pesos aproximadamente (Figura 1.16).

Pero a pesar de que las imágenes y secuencias tridimensionales han sido muy utilizados en la actualidad con fines comerciales, este tipo sistemas estereoscópicos manufacturados no suelen ser usados con objetivos más específicos como determinar la distancia de un objeto con respecto de las cámaras.

Algunos autores como Vibin N Valsan *et al.* [Valsan15], y Anwar Hasni Abu Hasan *et al.* [Anwar Hasni Abu Hasan09], dan muestra de las posibilidades que existen al utilizar



Figura 1.14: Cámara estereoscópica Holga con dos lentes de 60mm.



Figura 1.15: Videocámara Panasonic FULL HD 3D Camcorder para grabar video en 3D con alta definición.



Figura 1.16: Videocámara JVC Everio GS-TD1B Full HD 3D.

la visión estereoscópica como herramienta de estimación de distancias, o para la autolocalización de un robot. Sin embargo, sus trabajos no llegan a determinar de forma clara una forma de medir la profundidad de un objeto con respecto de una sistema estereoscópico, o tampoco una metodología para generar secuencias tridimensionales.

1.2. Propuesta de Solución

A partir de un sistema de cámaras estéreo, implementando cámaras digitales USB Logitech C920 HD, se propone la estimar la profundidad de diferentes objetos en una escena, utilizando un marco de referencia calculado a partir de imágenes anaglifo, modificando los parámetros intrínsecos de las cámaras, con la finalidad de reducir la distancia entre los centros de las cámaras de manera artificial. Se propone además, calcular mapas de disparidad densos, utilizando la búsqueda de correspondencia basada en correlación, implementando además, un esquema de color similar al utilizado en imágenes térmicas, con la finalidad de diferenciar a que distancia se encuentran los objetos, dependiendo del color de sus píxeles.

1.3. Objetivos de la Tesis

1.3.1. Objetivo general

Desarrollar el algoritmo para la estimación de la profundidad, el cual será nombrado como Algoritmo calcularProfundidad, capaz de estimar la profundidad de los objetos contenidos en una determinada escena, a partir de un par de imágenes estereoscópicas tomadas con un sistema de cámaras estéreo.

1.3.2. Objetivos particulares

- Calibrar un sistema estereoscópico, construido a partir de un par de cámaras digitales USB Logitech C920 HD.
- Utilizar imágenes anaglifo como técnica para evaluar visualmente los resultados del proceso de calibración del sistema estéreo, y en base al modelo de las cámaras, establecer un marco de referencia que facilite la estimación de la profundidad.
- Calcular mapas de disparidad densos y de buena calidad, con la intención de obtener información de todos los píxeles en las imágenes, con el objetivo de facilitar el proceso de estimación de la profundidad, en las escenas tomadas con el sistema estéreo.

- Desarrollar el algoritmo calcularProfundidad, capaz de estimar la profundidad de los objetos en una escena, tomando como parámetros iniciales, los mapas de disparidad de dicha escena, utilizando un mapa de color conocido como Hot color map (mapa de color caliente por su traducción al español), para colorear los pixeles en las imágenes con un color en particular dependiendo su valor de disparidad, y así identificar de manera aproximada, a que distancia se encuentran con respecto de las cámaras.

1.4. Descripción de Capítulos

El presente trabajo está formado por 6 capítulos distribuidos de la siguiente manera:

- En el capítulo 1 se presenta una introducción a los sistemas de visión estereoscópicos, explicando la procedencia de esta idea, y una corta remembranza sobre su evolución a lo largo de tiempo. Además de mencionar la propuesta de solución desarrollada a lo largo de esta tesis, para dicho tópico.
- En el capítulo 2 se hace una revisión del funcionamiento del modelo PinHole para las cámaras convencionales, se desarrolla el modelo matemático PinHole para un sistema de cámaras estéreo, y se determinan las ecuaciones de los parámetros del sistema.
- En el capítulo 3 se desarrolla una explicación del proceso de construcción del sistema estéreo usado para este trabajo, así como una breve explicación de las características técnicas de los dispositivos usados para construirlo. Por otro lado, utilizando la librería OpenCV, se describen las funciones utilizadas para llevar a cabo el proceso de calibración, los resultados de dicha calibración, y la generación de imágenes anaglifo y su impacto en la evaluación de los resultados de calibración, y en el proceso de estimación de la profundidad. Se analiza la relación disparidad - profundidad visualizada por medio de las imágenes anaglifo, y se proponen cambios a esta.
- En el capítulo 4 se plantea el problema de la búsqueda de correspondencia, y diferentes formas de abordar el problema. Además, se aborda la obtención de mapas de disparidad densos y su impacto en la estimación de la profundidad.

- En el capítulo 5 Se presentan resultados tanto de las imágenes anaglifo generadas con el sistema estéreo calibrado y aplicando desplazamiento de centros, así como resultados obtenidos a partir de la ejecución del algoritmo de estimación de profundidad.
- En el capítulo 6 se plantean las conclusiones y se mencionan posibles trabajos futuros.

Capítulo 2

Modelos de visión estereoscópica

Dado que el cálculo de las distancias con un sistema estéreo de cámaras es un tema relacionado con el campo de la visión computacional, es necesario dejar en claro algunos conceptos previos para poder entender la complejidad de este proceso. Por tal razón, quizá el concepto inicial más básico para analizar, es el de una “Imagen Digital”. Una imagen digital se puede definir como la representación artificial que ilustra la forma física de un objeto en el mundo real. Sin embargo, dado que lo que se busca es procesar y obtener información de estas representaciones visuales “digitales”, es necesario extender este concepto un poco más. Según Robert Laganier [Laganier11], una imagen se define como la representación de un escenario tridimensional en un plano bidimensional, pero dado que la información del escenario en cuestión, no puede ser capturada en su totalidad, esta puede ser discretizada y representada como una matriz de dos dimensiones compuesta por valores numéricos, por lo que una imagen se definiría entonces como una matriz bidimensional I de tamaño $(N_r \times N_c)$, donde N_r es el número de renglones y N_c el número de columnas. Ahora bien, para poder inspeccionar los elementos que componen a esta matriz, es necesario determinar un sistema coordenado, el cuál utilizará a la letra x para recorrer las filas, y a la letra y para recorrer las columnas. Por otro lado, la imagen I estará compuesta por pixeles en cada uno de sus elementos, por lo que es posible definir a un pixel como el componente más pequeño que forma parte de una imagen digital, y en este caso, de la imagen I . Cada píxel a su vez estará formado por un pequeño vector con tres valores que pueden variar desde 0 hasta 255;

estos valores son para tres colores, rojo, verde y azul. Dependiendo del valor que adquiera cada uno de los elementos del pixel, se formará una combinación de colores conocida como profundidad de color. Sabiendo esto, si se desea hacer referencia a la información de un pixel, es posible hacerlo como $I(x, y)$, donde la posición estará denotada por un par de coordenadas (x, y) , pertenecientes a la imagen I . Un ejemplo de esto se muestra en la Figura 2.1.

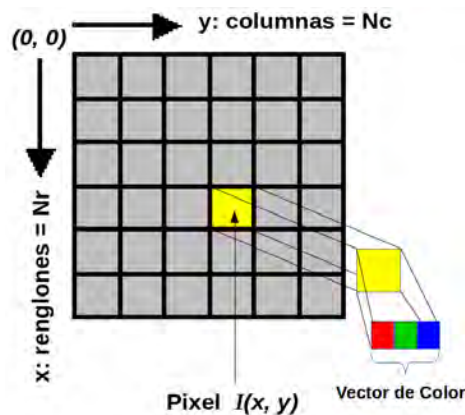


Figura 2.1: Representación de una imagen digital.

El reto entonces, es modelar el mundo tridimensional a partir de imágenes digitales donde la profundidad de los objetos se pierde. Para ello es necesario comenzar con la descripción del modelo más simple de una cámara, conocido como el modelo Agujero de alfiler, y mayormente identificado por su nombre en inglés, como modelo PinHole, como explica Kenneth Dawson [Dawson-Howe14]. Cabe mencionar, que a lo largo de esta tesis, dicho modelo será referenciado por su nombre en inglés.

2.1. Modelo pinhole

Las imágenes digitales son generalmente producidas usando una cámara digital que captura una escena al proyectar la luz a través de un lente y directo a un sensor de imagen. Por lo tanto, el simple hecho de que una escena 3D sea proyectada en un plano 2D, implica el uso de la geometría proyectiva, que es una herramienta para describir y caracterizar, en

términos matemáticos, el proceso de la formación de una imagen, según Gary Bradski *et al.* [Bradski08].

El modelo PinHole, es el modelo más simple utilizado para representar el proceso de captura de una escena. Una cámara tipo PinHole, se describe como una caja con un agujero hecho por una alfiler (Pin-Hole en inglés, Figura 2.2), en uno de sus lados. Dicha cámara muestra como la luz es capturada a través del agujero de alfiler, sufriendo una rotación de 180 grados, quedando plasmada en un plano como se ilustra en la Figura 2.2.

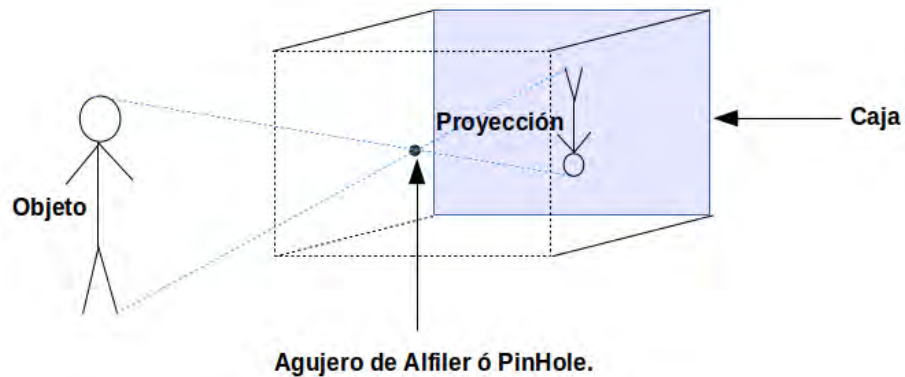


Figura 2.2: Comportamiento del Modelo PinHole usando una caja.

Durante el proceso de captura en la caja, se pierde una dimensión, es decir, se tiene una altura y un ancho del objeto en la escena, pero se desconoce la profundidad. Por otro lado, si el objeto en la escena capturada, se aleja o se acerca a la caja, dicho movimiento se ve reflejado en la imagen proyectada, ya que el objeto se ve más grande si se acerca y más pequeño si se aleja.

Por las razones mencionadas, el modelo PinHole define que un punto P observado en el mundo real con coordenadas $[X_0, Y_0, Z_0]^T$, puede ser representado por un punto q en un plano de la caja con coordenadas homogéneas $[x_0, y_0, 1]^T$, donde X_0 es el ancho, Y_0 la altura y Z_0 la profundidad o distancia a la que se encuentra el punto P de la caja, como se aprecia en la Figura 2.3.

De acuerdo con Richard Hartley *et al.* [Hartley04], el modelo de la cámara PinHole

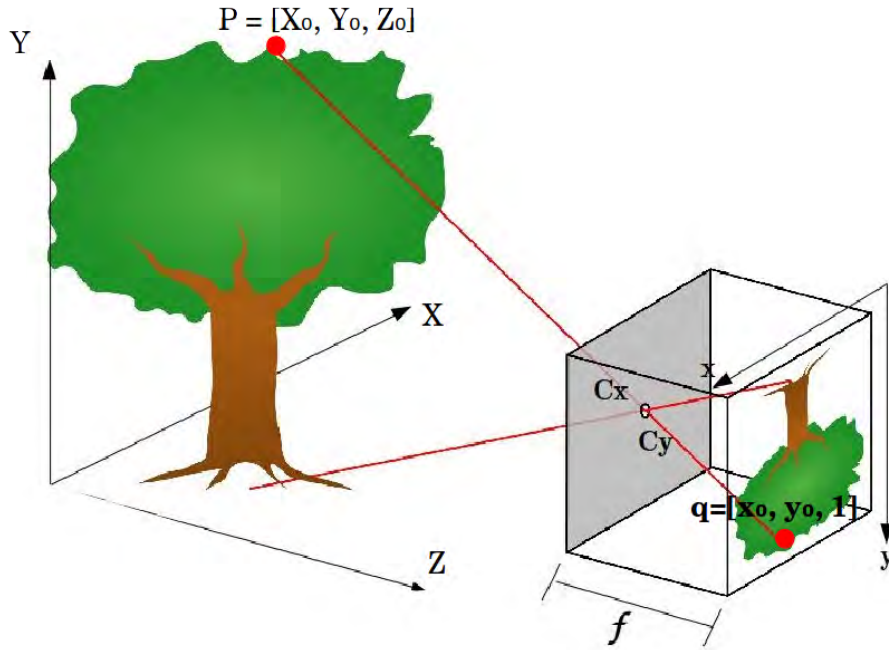


Figura 2.3: Modelo matemático PinHole. Imagen original descargada de Wikipedia.

está dado por (2.1) y (2.2), las cuales representan en coordenadas de la caja, la posición de un punto P tomando como referencia el modelo mostrado en la Figura 2.3, donde se observa que la caja tiene dimensiones $2C_x$, $2C_y$ y f , de tal manera que C_x es la posición en x de la abertura por donde atraviesa la luz, C_y es la posición en y de la misma abertura, y f , la distancia que existe entre la pared de la caja con la abertura, y la pared opuesta donde se proyecta la imagen del objeto, conocida como distancia focal. Los valores C_x , C_y y f , son conocidos como parámetros intrínsecos de la cámara.

$$x_0 = f \frac{X_0}{Z_0} + C_x \quad (2.1)$$

$$y_0 = f \frac{Y_0}{Z_0} + C_y \quad (2.2)$$

Es posible organizar en forma matricial (2.1) y (2.2) del modelo de la Figura 2.3 como en (2.3).

$$\begin{bmatrix} \hat{x}_0 \\ \hat{y}_0 \\ 1 \end{bmatrix} = \begin{bmatrix} f & 0 & Cx \\ 0 & f & Cy \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X_0 \\ Y_0 \\ Z_0 \end{bmatrix} \quad (2.3)$$

Sin embargo, como se mencionaba anteriormente, en el proceso de mapeo del punto P a coordenadas de la caja, se pierde la dimensión de profundidad Z_0 , por lo que para poder llegar a las ecuaciones (2.1) y (2.2), y obtener las coordenadas homogéneas del punto q , es necesario dividir los elementos \hat{x}_0 , \hat{y}_0 y Z_0 , entre Z_0 , obteniendo así a (2.4), lo que permite establecer una congruencia entre (2.4), y (2.1) y (2.2).

$$\begin{bmatrix} Z_0 \cdot x_0 \\ Z_0 \cdot y_0 \\ Z_0 \end{bmatrix} = \begin{bmatrix} f & 0 & Cx \\ 0 & f & Cy \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X_0 \\ Y_0 \\ Z_0 \end{bmatrix} \quad (2.4)$$

Con las operaciones realizadas en (2.4), es entonces posible representar un punto P en el mundo real, en coordenadas de la caja como un punto q . Según Richard Hartley *et al.* [Hartley04], es posible reacomodar la ecuación matricial del modelo PinHole como (2.5).

$$q = K \times P \quad (2.5)$$

Tomando la matriz que contiene la distancia focal f , y los centros de la cámara por donde pasa la luz, C_x y C_y en (2.3), se obtiene la matriz K (2.6), la cual es conocida como la matriz de parámetros intrínsecos de una cámara, y que al ser multiplicada por un punto P en coordenadas tridimensionales, mapea dicho punto a coordenadas de la caja.

$$K = \begin{bmatrix} fx & 0 & Cx \\ 0 & fy & Cy \\ 0 & 0 & 1 \end{bmatrix} \quad (2.6)$$

Ahora bien, si lo que se desea es modelar un sistema estereoscópico de cámaras, es necesario tomar en cuenta algunos otros factores que afectan su comportamiento, por lo cuál, se presenta la deducción del modelo PinHole para un sistema de cámaras estéreo.

2.2. Modelo pinhole para cámaras estéreo

Con el objetivo de lograr un efecto tridimensional similar a la visión humana, es necesario contar con dos cámaras como requisito mínimo, para igualar dicho comportamiento.

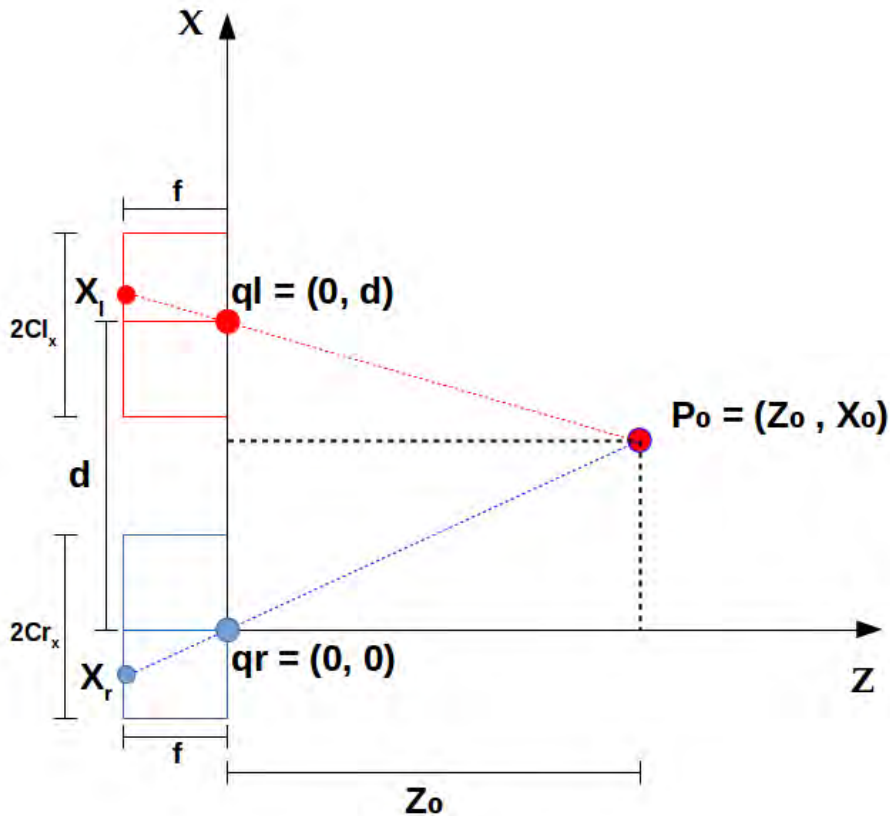


Figura 2.4: Modelo estereoscópico para el plano X, Z .

En la Figura 2.4, se puede apreciar un diagrama de un sistema de cámaras PinHole estéreo, que en condiciones normales, debe estar alineado perfectamente en el eje Y . Bajo esta condición, es posible hacer un análisis únicamente para el plano Z, X , en coordenadas tridimensionales. Como se observa en la Figura 2.4, las dos cámaras del sistema estéreo, se observan alineadas en el eje X . Por otro lado, se observa la distancia d , haciendo referencia a la distancia existente entre los centros de cámaras, que a su vez están representados por

los puntos ql y qr . Como se mencionó anteriormente, la distancia entre el orificio por donde entra la luz, y la superficie de la cámara donde se proyecta el objeto, es conocida como distancia focal, y está representada por la letra f , la cuál para este caso, se encuentra localizada en el cuadrante negativo del eje coordenado Z . Finalmente, el diagrama expone dos líneas rectas: una desde el punto P_0 al punto ql , y otra del mismo punto P_0 pero al punto qr , ejemplificando la trayectoria que sigue la luz al entrar en las cámaras PinHole, mostrando además, que cada cámara tiene una perspectiva diferente del punto P_0 .

Para determinar el modelo PinHole del sistema estéreo, primero se debe determinar el modelo de cada una de las cámaras de forma individual. Para este caso, se toma como base la ecuación de una recta que pasa por dos puntos, mostrada en (2.7).

$$X - X_1 = \frac{X_2 - X_1}{Z_2 - Z_1}(Z - Z_1) \quad (2.7)$$

Para la cámara izquierda, se toma la línea que pasa por el punto P_0 con coordenadas $[Z_0, X_0]$, y el punto ql con coordenadas $[0, d]$. Substituyendo en (2.7), se obtiene la siguiente ecuación:

$$X - X_0 = \frac{d - X_0}{0 - Z_0}(Z - Z_0) \quad (2.8)$$

El objetivo de (2.8), es encontrar la coordenada de proyección X_l del punto P_0 . Por tal razón, el punto proyectado en la cámara izquierda, tiene coordenadas $[X_l, -f]$, debido a que la distancia focal se encuentra en el cuadrante negativo del eje Z . Dicho esto, al substituir y despejar X_l en (2.8), se obtiene la siguiente ecuación:

$$X_l = \frac{d - X_0}{-Z_0}(-f - Z_0) + X_0 \quad (2.9)$$

Desarrollando la multiplicación del primer término en (2.9) y simplificando el resultado, se obtiene (2.10), que permite identificar las coordenada X_l de proyección del punto P_0 , en la superficie de proyección de la cámara izquierda.

$$X_l = (-f) \frac{X_0 - d}{Z_0} + d \quad (2.10)$$

Por otro lado, para la cámara derecha se repite el mismo procedimiento, pero esta vez tomando en cuenta la línea que pasa por los puntos $P_0 = [Z_0, X_0]$ y $qr = [0, 0]$, por lo que al substituir en (2.7), se obtiene la siguiente ecuación:

$$X - X_0 = \frac{0 - X_0}{0 - Z_0}(Z - Z_0) \quad (2.11)$$

Ahora el objetivo con (2.11) es encontrar la coordenada de proyección X_r del punto P_0 , pero en la superficie de proyección de la cámara derecha, este punto de proyección tiene coordenadas $[X_r, -f]$, por lo que al substituir en (2.11), se obtiene que:

$$X_r = \frac{-X_0}{-Z_0}(-f - Z_0) + X_0 \quad (2.12)$$

Finalmente, desarrollando la multiplicación del primer término en (2.12), y simplificando los resultados, se obtiene la ecuación para encontrar la coordenada de proyección X_r para el punto P_0 con (2.13).

$$X_r = (-f) \frac{X_0}{Z_0} \quad (2.13)$$

Por otro lado, siguiendo el diagrama de la Figura 2.5, es posible observar el comportamiento del modelo PinHole estéreo, pero ahora con respecto al plano Z, Y .

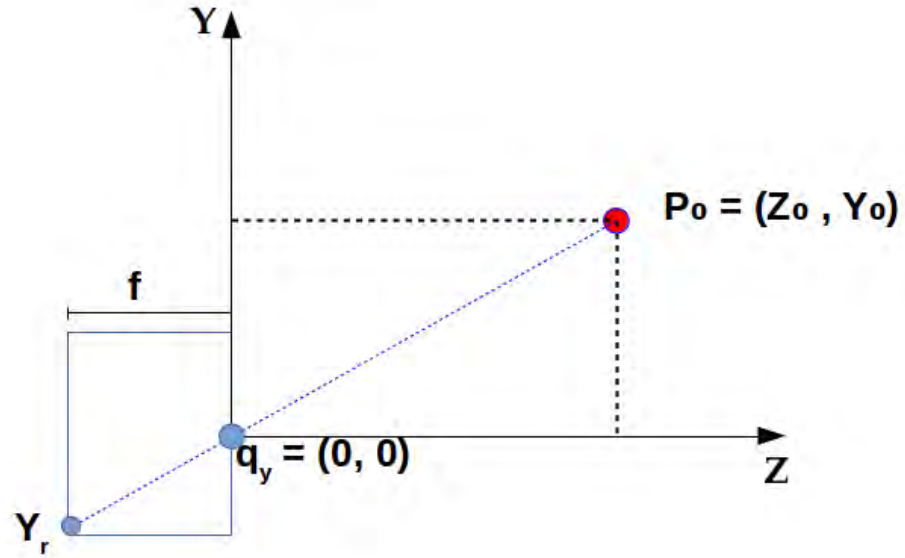
Partiendo de igual forma de la ecuación de la línea recta, se tiene que:

$$Y_r - Y_1 = \frac{Y_2 - Y_1}{Z_2 - Z_1}(Z - Z_1) \quad (2.14)$$

Siguiendo la Figura 2.5, se toma la línea recta que cruza los puntos $P_0 = [Z_0, Y_0]$, y $q_y = [0, 0]$. Substituyendo esta información en (2.15), se obtiene que:

$$Y - Y_0 = \frac{0 - Y_0}{0 - Z_0}(Z - Z_0) \quad (2.15)$$

Para el plano Z, Y , se busca obtener la coordenada de proyección Y_r para el punto P_0 , en la superficie de proyección de la cámara derecha, por lo tanto, se toman las coordenadas $[Y_r, -f]$, y se substituyen en (2.15), obteniendo así (2.16).

Figura 2.5: Modelo estereoscópico para el plano Z, Y .

$$Y_r = \frac{-Y_0}{-Z_0}(-f - Z_0) + Y_0 \quad (2.16)$$

Desarrollando la multiplicación del primer término en (2.16) y simplificando términos, se obtiene (2.17) para el cálculo de la coordenada de proyección Y_r del punto P_0 .

$$Y_r = (-f) \frac{Y_0}{Z_0} \quad (2.17)$$

Dado que para el modelo estéreo, tanto la cámara derecha como la cámara izquierda deben estar completamente alineadas en el eje Y , se puede decir que $Y_l = Y_r$, donde Y_l es la coordenada de proyección en el eje Y para el punto P_0 , en la cámara izquierda.

Ahora bien, si se sintetiza (2.10), (2.13), y (2.17) para la cámara izquierda y derecha respectivamente, se obtienen las siguientes ecuaciones matriciales que permiten el mapeo de un punto P_0 , a coordenadas de la cámara izquierda o derecha según sea el caso:

$$\begin{bmatrix} Z_0 \cdot X_l \\ Z_0 \cdot Y_l \\ Z_0 \end{bmatrix} = \begin{bmatrix} -f & 0 & d \\ 0 & -f & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X_0 - d \\ Y_0 \\ Z_0 \end{bmatrix} \quad (2.18)$$

$$\begin{bmatrix} Z_0 \cdot X_r \\ Z_0 \cdot Y_r \\ Z_0 \end{bmatrix} = \begin{bmatrix} -f & 0 & 0 \\ 0 & -f & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X_0 \\ Y_0 \\ Z_0 \end{bmatrix} \quad (2.19)$$

2.2.1. Modelo en coordenadas de la imágenes

Hasta este punto, con (2.18) y (2.19) es posible mapear un punto P_0 en coordenadas tridimensionales, a coordenadas tridimensionales de la cámara. Sin embargo, para poder llevar un punto en el mundo tridimensional, al plano bidimensional de las imágenes, es necesario involucrar información relacionada con los centros de cámaras, y utilizar el marco de referencia de las imágenes como se muestra en la Figura 2.6.

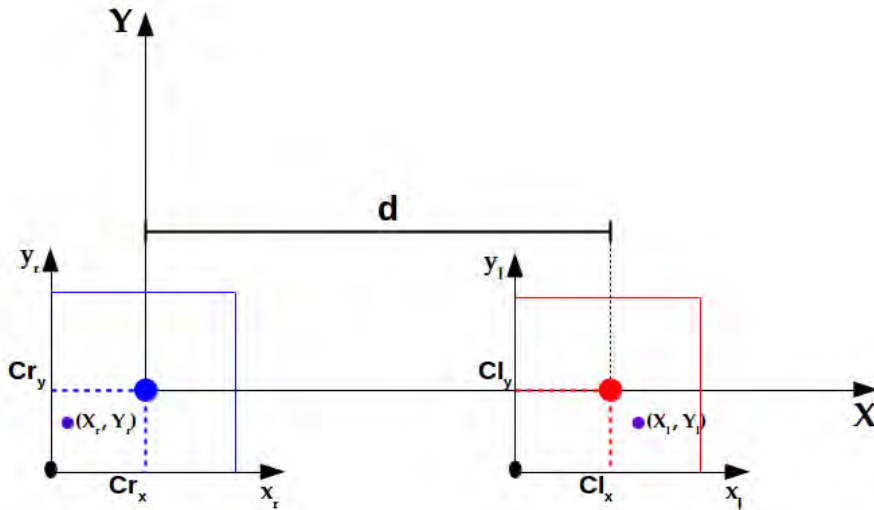


Figura 2.6: Planos coordenados de imagen para el sistema estereoscópico.

En la Figura 2.6, se puede apreciar que el sistema estereoscópico está en función de tres planos coordenados diferentes, el plano coordenado X, Y , el plano x_r, y_r de la cámara derecha, y el plano x_l, y_l de la cámara izquierda. Por tanto, como el objetivo es obtener

las coordenadas bidimensionales, de un punto en coordenadas tridimensionales, es necesario llevar a cabo una traslación para alinear los planos x_r, y_r , y x_l, y_l , con el plano X, Y . Para lograr esta traslación, es necesario sumar los centros de cámaras a (2.10), (2.13), y (2.17), como se muestra a continuación:

$$x_r = X_r + Cr_x \quad (2.20)$$

$$y_r = Y_r + Cr_y \quad (2.21)$$

En (2.20), para el cálculo de la coordenada x_r en términos de la imagen, se suma a X_r el centro de cámara Cr_x , donde X_r representa a (2.13). Por otro lado en (2.21), para el cálculo de la coordenada y_r en términos de la imagen, se suma a Y_r el centro de cámara Cr_y , donde repitiendo el mismo proceso, Y_r representa a (2.17). Ahora bien, tomando a (2.13) y (2.17), y remplazándolas en (2.20) y (2.21) respectivamente, se logra alinear el plano x_r, y_r , al origen coordenado del plano X, Y . Por otro lado, se obtiene a (2.22) y (2.23) para el mapeo de un punto P_0 en coordenadas tridimensionales, a coordenadas de imagen en la cámara derecha.

$$x_r = (-f) \frac{X_0}{Z_0} + Cr_x \quad (2.22)$$

$$y_r = (-f) \frac{Y_0}{Z_0} + Cr_y \quad (2.23)$$

De igual forma, para la cámara izquierda, se lleva a cabo un procedimiento similar. Sin embargo, debido al desplazamiento existente entre el centro de la cámara izquierda con respecto de la derecha, es necesario restar la distancia d a (2.10), con el objetivo de que el origen coordenado X, Y se posicione en el centro de la cámara izquierda. Por ende, realizando esta operación, se tiene que:

$$X_l = (-f) \frac{X_0 - d}{Z_0} + d - d \quad (2.24)$$

Hecho esto, se procede a alinear x_l, y_l , con X, Y , sumando los centros de la cámara izquierda a (2.24) representada por X_l y (2.23) representada por Y_l como se muestra a continuación:

$$x_l = X_l + Cl_x \quad (2.25)$$

$$y_l = Y_l + Cl_y \quad (2.26)$$

Siguiendo la misma mecánica utilizada para la cámara derecha, se substituye en (2.25) y (2.26), la información de (2.24) y (2.23) respectivamente, obteniendo así a (2.27) y (2.28), para llevar un punto en coordenadas tridimensionales, a coordenadas de imagen en la cámara izquierda. No olvidar que las cámaras presentan una alineación perfecta en el plano Z, Y , por lo que es posible utilizar a (2.17) para la cámara izquierda, como se hizo con la cámara derecha.

$$x_l = (-f) \frac{X_0 - d}{Z_0} + Cl_x \quad (2.27)$$

$$y_l = (-f) \frac{Y_0}{Z_0} + Cl_y \quad (2.28)$$

Finalmente, sintetizando (2.22) y (2.23) para la cámara derecha, y (2.27) y (2.28) para la cámara izquierda en forma matricial, se obtiene a (2.29) y (2.30) para la cámara derecha e izquierda respectivamente.

$$\begin{bmatrix} Z_0 \cdot x_r \\ Z_0 \cdot y_r \\ Z_0 \end{bmatrix} = \begin{bmatrix} -f & 0 & Cr_x \\ 0 & -f & Cr_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X_0 \\ Y_0 \\ Z_0 \end{bmatrix} \quad (2.29)$$

$$\begin{bmatrix} Z_0 \cdot x_l \\ Z_0 \cdot y_l \\ Z_0 \end{bmatrix} = \begin{bmatrix} -f & 0 & Cl_x \\ 0 & -f & Cl_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} (X_0 - d) \\ Y_0 \\ Z_0 \end{bmatrix} \quad (2.30)$$

2.2.2. Parámetros extrínsecos del sistema

El modelo de cámaras PinHole estéreo, asume que el sistema de cámaras está perfectamente alineado en todos los sentidos, y que no hay ningún tipo de factor que pueda alterar esta situación. Sin embargo, al momento de trabajar con cámaras digitales, esto no

se acerca a la realidad, ya que físicamente, es prácticamente imposible alinear las cámaras a la perfección (siempre existirá un error de alineación que puede ser causado principalmente por factores físicos). Por esta razón, es necesario añadir una matriz de parámetros extrínsecos a cada cámara del sistema. De acuerdo con Robert Laganier [Laganier11], generalmente cuando el marco coordenado de referencia, no se encuentra en el centro de las cámaras, es necesario añadir una matriz de rotación de tamaño 3×3 y un vector de traslación de tamaño 3×1 . Estas dos nuevas matrices describen una transformación rígida que debe ser aplicada al punto en coordenadas tridimensionales, con el objetivo de ponerlo en coordenadas de la imagen. La matriz de rotación, y el vector de traslación, son conocidas como parámetros extrínsecos del sistema, y en específico, la función principal de la matriz de rotación, es corregir el error de alineación que pueda existir entre las cámaras del sistema, mientras que el vector de traslación, almacena la distancia “ d ”, que existe entre el centro de la cámara izquierda, con respecto del centro la derecha. Por lo que al añadir estos nuevos parámetros en (2.29) y (2.30), se obtiene (2.31) y (2.32). Pero además, para mantener la coherencia en las operaciones matriciales en ambas ecuaciones, es necesario añadir un $\mathbf{1}$ al punto de coordenadas tridimensionales, convirtiéndolas así, en coordenadas homogéneas.

$$\begin{bmatrix} Z_0 \cdot x_r \\ Z_0 \cdot y_r \\ Z_0 \end{bmatrix} = \begin{bmatrix} -f & 0 & Cr_x \\ 0 & -f & Cr_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X_0 \\ Y_0 \\ Z_0 \\ 1 \end{bmatrix} \quad (2.31)$$

$$\begin{bmatrix} Z_0 \cdot x_l \\ Z_0 \cdot y_l \\ Z_0 \end{bmatrix} = \begin{bmatrix} -f & 0 & Cl_x \\ 0 & -f & Cl_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & -d \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X_0 \\ Y_0 \\ Z_0 \\ 1 \end{bmatrix} \quad (2.32)$$

Como se puede observar en (2.31) y (2.32), las matrices de rotación son matrices identidad, esto debido a que dicho modelo, supone que las cámaras están perfectamente alineadas. Por otro lado, se puede observar que en ambos casos, la matriz de rotación es igual en dimensiones que la matriz identidad anexada al vector de traslación, por lo que es posible reemplazar esta matriz identidad y poner en su lugar la matriz de rotación, como se

muestra en (2.33) y (2.34).

$$\begin{bmatrix} Z_0 \cdot x_r \\ Z_0 \cdot y_r \\ Z_0 \end{bmatrix} = \begin{bmatrix} -f & 0 & Cr_x \\ 0 & -f & Cr_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} Rr_{0,0} & Rr_{0,1} & Rr_{0,2} & 0 \\ Rr_{1,0} & Rr_{1,1} & Rr_{1,2} & 0 \\ Rr_{2,0} & Rr_{2,1} & Rr_{2,2} & 0 \end{bmatrix} \begin{bmatrix} X_0 \\ Y_0 \\ Z_0 \\ 1 \end{bmatrix} \quad (2.33)$$

$$\begin{bmatrix} Z_0 \cdot x_l \\ Z_0 \cdot y_l \\ Z_0 \end{bmatrix} = \begin{bmatrix} -f & 0 & Cl_x \\ 0 & -f & Cl_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} Rl_{0,0} & Rl_{0,1} & Rl_{0,2} & -d \\ Rl_{1,0} & Rl_{1,1} & Rl_{1,2} & 0 \\ Rl_{2,0} & Rl_{2,1} & Rl_{2,2} & 0 \end{bmatrix} \begin{bmatrix} X_0 \\ Y_0 \\ Z_0 \\ 1 \end{bmatrix} \quad (2.34)$$

De esta forma, los valores Rr y Rl en (2.33) y (2.34), representan los valores de las matrices de rotación para la cámara derecha e izquierda respectivamente, mientras que la columna que almacena el desplazamiento “ d ”, será identificado como el vector de traslación “ T ”.

Se puede observar que (2.33) y (2.34) son muy similares a las ecuaciones desarrolladas previamente para el modelo PinHole, por lo que si se desea mapear un punto con coordenadas tridimensionales, en el plano bidimensional de las imágenes de cada cámara, es posible hacerlo con estas ecuaciones. Para más detalles de este modelo, refiérase a Richard Hartley *et al.* [Hartley04].

2.3. Cálculo de la disparidad

Según la Real Academia Española, la palabra disparidad, hace referencia a la desigualdad o diferencia que existe en la comparación de alguna cosa, con respecto de otra. Aplicado a la visión binocular del ser humano, se puede hacer referencia a la disparidad como la desigualdad o diferencia existente en la perspectiva con la que cada ojo observa algún objeto. Según Ales Ude [Ude10], esta diferencia se da debido a la distancia existente entre los ojos, lo que permite obtener dos imágenes casi iguales pero con ligeras diferencias de perspectiva con el objetivo de determinar la profundidad de los objetos en una escena observada (Figura 2.7). Por esta razón, a la hora de trabajar con un sistema estéreo artificial,

el cálculo de la disparidad juega un papel importante para poder determinar las diferentes profundidades de los objetos en una escena, añadiendo a esto la posibilidad de triangular un objeto en el espacio, con respecto de las cámaras.

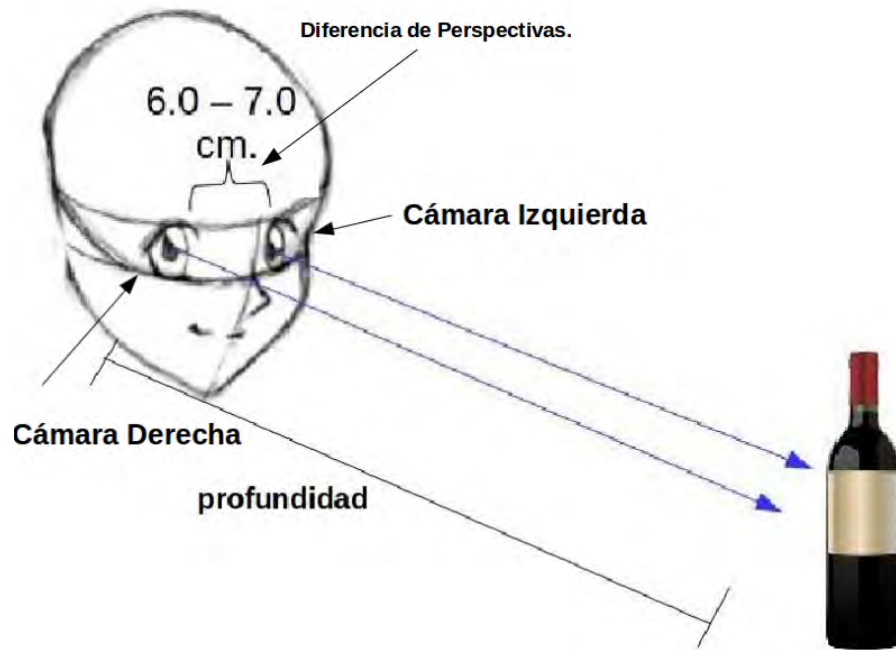


Figura 2.7: La diferencia de perspectivas por la distancia entre los ojos, es lo que permite medir la profundidad de los objetos en una escena.

Sin embargo, aunque el cálculo de la disparidad pudiera aparentar ser una tarea de mucha complejidad, es un procedimiento relativamente sencillo, y por medio del cuál se recopila información espacial de un punto en el espacio, en el plano de las cámaras. Por lo que tomando como punto de partida el ejemplo de la Figura 2.6, el cálculo de la disparidad necesita como parámetros iniciales dos píxeles en coordenadas de las cámaras, un píxel q_l con coordenadas de cámara $[x_l, y_l]$ en la imagen I_l , perteneciente a la cámara izquierda, y un píxel q_r con coordenadas de cámara $[x_r, y_r]$ en la imagen I_r , perteneciente a la cámara derecha. Posteriormente, lo único que se hace es calcular la diferencia que existe en las posiciones en el eje coordenado de las x , como se muestra en (2.35). Cabe mencionar que dado el desplazamiento existente sobre el eje coordenado x , el mapa de disparidad que nos interesa obtener, es el existente en este plano, ya que como las cámaras están alineadas

verticalmente al eje coordenado y , el mapa de disparidad en el plano y , será cero.

$$D_x = x_l - x_r \quad (2.35)$$

En (2.35), x_l es la coordenada del punto ql en el eje x en la imagen I_l , y x_r la coordenada del punto q_r en x , en la imagen I_r . Si se substituye con la información de (2.22) a x_r y (2.27) a x_l , obtenemos (2.36).

$$D_x = (-f) \frac{(X_0 - d)}{Z_0} + Cl_x - (-f) \frac{X_0}{Z_0} + Cr_x \quad (2.36)$$

Ahora bien, suponiendo que los centros de las cámaras Cl_x y Cr_x son iguales, y simplificando a (2.36), se obtiene que el cálculo de la disparidad se reduce a la siguiente ecuación:

$$D_x = x_l - x_r = f \frac{d}{Z_0} \quad (2.37)$$

Donde D_x es la variable que almacenará la disparidad, y Z_0 la coordenada que representa la profundidad del punto observado, siguiendo el modelo de la Figura 2.4.

Con 2.37 es entonces posible calcular una estimación de la profundidad de todos los puntos contenidos en una par de imágenes estéreo, y diferenciar la profundidad de los objetos en la escena con respecto de las cámaras en la imagen.

Ahora bien, si lo que se desea es triangular un punto en el espacio, despejando a Z_0 en 2.37, y substituyendo la disparidad D_X por la resta equivalente de $x_l - x_r$, se obtiene que:

$$Z_0 = f \frac{d}{x_l - x_r} \quad (2.38)$$

Por lo tanto, para poder calcular una aproximación de la profundidad de un punto en el espacio utilizando el sistema estereoscópico, con 2.38 es posible determinarla.

2.4. Conclusión del capítulo

En conclusión, es posible determinar que el modelo PinHole para un sistema estéreo es relativamente simple, si todo el sistema se encuentra bien alineado. Sin embargo, esto no es así, ya que en primer lugar, las cámaras son digitales y no PinHole, y en segundo lugar, debido a que es sumamente difícil armar un sistema de cámaras estereoscópico, con una alineación perfecta en todos los sentidos. Por lo tanto, para que el sistema funcione correctamente, es necesario abordar el tema de la calibración, donde se obtendrá información detallada como: la matriz de parámetros intrínsecos K , la matriz de rotación R y el vector de traslación T , para cada una de las cámaras con el objetivo de mejorar los resultados del sistema. Por otro lado, para demostrar el correcto funcionamiento de la calibración, y el impacto que tiene, se generarán imágenes anaglifo.

Capítulo 3

Construcción y Calibración de un Sistema Estéreo

El modelo PinHole para sistemas estéreo puede ser implementado en cualquier sistema estéreo de cámaras, ya sea que este cuente con dispositivos de muy buena calidad, o dispositivos de bajo costo. Sin embargo, los resultados que se obtengan a partir del sistema construido, siempre van a estar en función de la calidad de la información que las cámaras recopilen. Además, el sistema estéreo siempre estará expuesto a factores físicos, ya sea defectos de fabricación, o que al momento de construir el sistema, no se hayan fijado las cámaras. Sin embargo, el hecho de fijar las cámaras, no garantiza que tengan una correcta alineación, ya que por mucho cuidado que se haya puesto a la hora de construir el sistema en cuestión, es prácticamente imposible alinearlas perfectamente. Por tal razón, es necesario llevar a cabo el proceso de calibración, de tal forma que con la información obtenida a partir de este procedimiento, sea posible corregir la alineación de las cámaras de forma artificial, como se explica a lo largo de este capítulo.

3.1. Características del sistema estereoscópico

A continuación se presenta el desarrollo y construcción del sistema de cámaras estéreo construido para este proyecto. Se muestran las características de las cámaras utili-

zadas, y la construcción en general del sistema estéreo con sus elementos y componentes. También se muestra el proceso de calibración en general, y su efecto en las cámaras, a partir de la generación de imágenes anaglifo.

3.1.1. Características de las cámaras

Para el presenta trabajo, se utilizaron dos cámaras marca Logitech modelo C920. Estas cámaras además de ser económicas, son dispositivos de alta calidad y que cuentan con diferentes características de bastante utilidad para este proyecto.

Algunas de las características más importantes de las cámaras, se enlistan a continuación:

- Alta definición. Las cámaras Logitech C920 tienen la capacidad para capturar tanto video, como fotografías con una resolución de hasta 1920 x 1080 píxeles de calidad (se conoce como Full HD 1080p).
- Compatibilidad. Cuentan con una interfaz de comunicación USB 2.0 (compatible con USB 3.0), y son compatibles con los sistemas operativos más comunes (entiéndase Windows, MacOS y Linux).
- Iluminación. También incluye la propiedad de corrección automática ante iluminación escasa.
- Autoenfocado. Incluyen la propiedad de autoenfocado, lo cuál permite tener imágenes nítidas en todo momento.

Algunas otras características pueden ser encontradas en la página del fabricante (<http://www.logitech.com/es-mx/product/hd-pro-webcam-c920>).

Las dimensiones de las cámaras son 9.4 cm de ancho, 7.1 cm de profundidad y 4.3 cm de altura (Figura 3.1). Estas medidas ayudaron al desarrollo de la estructura del sistema que posteriormente será explicado.



Figura 3.1: Dimensiones de la cámara Logitech C920.

3.1.2. Características de la estructura del sistema

Para entender más claramente la problemática involucrada en la construcción de un sistema estéreo, es necesario tomar como ejemplo el modelo PinHole estéreo donde las cámaras del sistema están perfectamente alineadas, y son exactamente iguales como se muestra en la Figura 3.2.

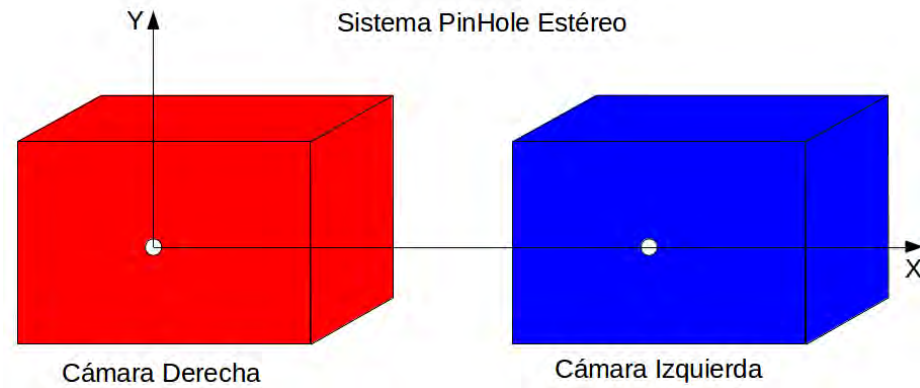


Figura 3.2: Modelo de cámaras PinHole estéreo.

A partir del modelo PinHole estéreo, es entonces posible desarrollar la deducción de (2.37) para el cálculo de la disparidad de un punto q , o la deducción de (2.38) para determinar la profundidad un punto P en coordenadas tridimensionales. Sin embargo, en el trayecto al desarrollo de estas ecuaciones, es necesario tomar en cuenta algunas condiciones particulares relacionadas al modelo PinHole, que no son aplicables a sistemas estereoscópicos reales. Así pues, con el objetivo de mostrar el impacto de la calibración en un sistema

estereoscópico, se toma a (2.33) y (2.34), y se representan como se muestra en (3.1) y (3.2).

$$q_r = K_r[R_r|T_r]P \quad (3.1)$$

$$q_l = K_l[R_l|T_l]P \quad (3.2)$$

Dado que (3.1) y (3.2), son representaciones de (2.33) y (2.34) respectivamente, se debe entender que K_r y K_l son las matrices de parámetros intrínsecos para cada cámara, R_r y R_l las matrices de rotación, que en conjunto con los vectores de traslación T_r y T_l , forman las matrices de parámetros extrínsecos, mientras que P representa el vector de coordenadas homogéneas de un punto en el plano tridimensional. Con este marco de referencias, es posible resaltar que con el modelo PinHole estereo, se puede deducir a (2.35) y (2.36) suponiendo que los valores de las matrices K_r y K_l , son exactamente iguales, y que por consecuencia, las matrices de rotación R_r y R_l , son matrices identidad, ya que no hay necesidad de una rectificación del sistema. Sin embargo, para un sistema estereo real, estas condiciones no se cumplirán de igual forma, ya que las cámaras nunca estarán perfectamente alineadas, y los centros de cámaras tampoco serán iguales, siendo necesario aplicar una transformación geométrica. Es aquí donde entra el proceso de calibración. De acuerdo con Gary Bradski *et al.* [Bradski08], la calibración estereo consiste en encontrar la relación geométrica entre dos cámaras en el mundo real, para posteriormente, con la información obtenida, corregir de forma individual las imágenes de cada cámara, de tal forma que parezca que fueron tomadas por dos cámaras perfectamente alineadas horizontalmente, como se muestra en la Figura 3.3. A partir de esto, con el proceso de calibración será posible obtener las matrices de rotación R_r y R_l , que alinearán de manera artificial las cámaras de un sistema estereo, logrando de esta forma que K_r sea igual a K_l , y cumpliendo así con la deducción de (2.35) y (2.36).

Dadas las razones mencionadas, es necesario remarcar que las cámaras deben mantener una posición fija una vez rectificadas, ya que con el menor movimiento, sería necesario repetir el proceso de calibración, y obtener nuevamente los parámetros necesarios del sistema. Sin embargo, es posible evitar esta situación, fijando las cámaras con alguna estructura

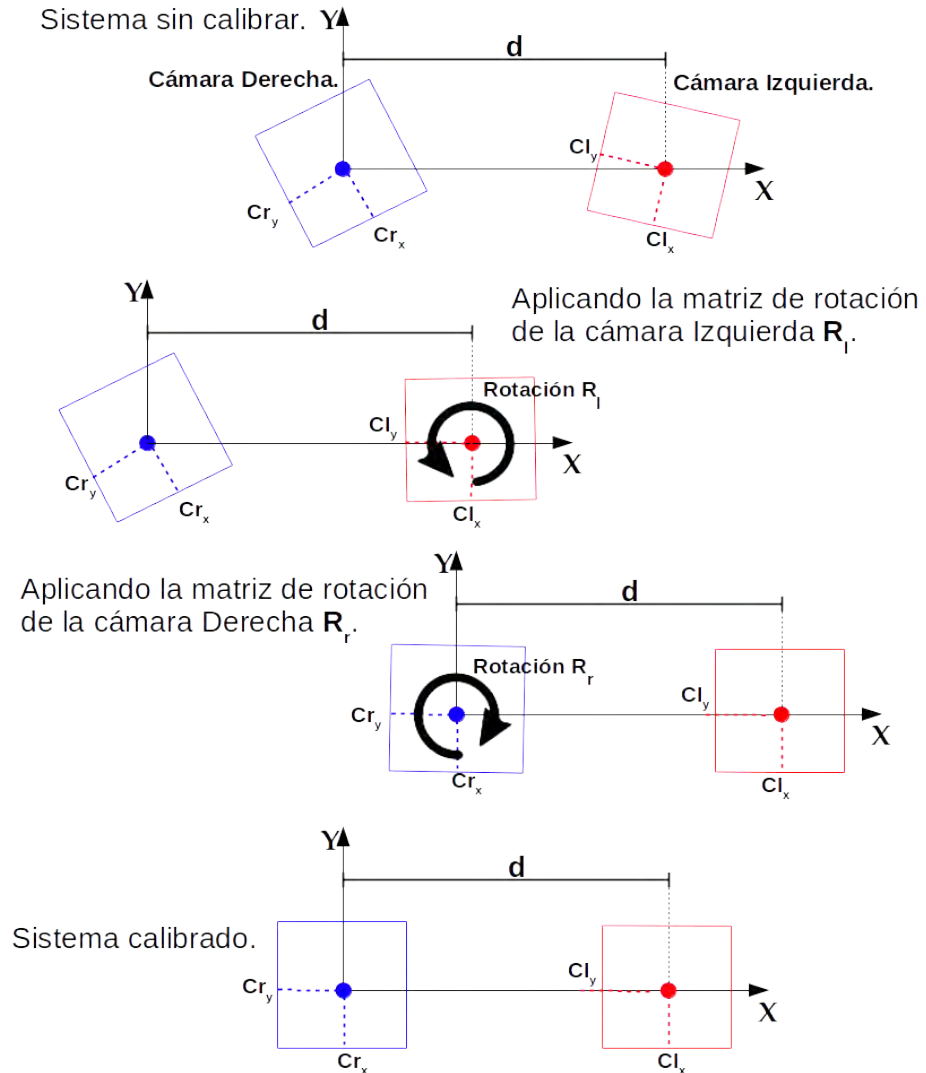


Figura 3.3: Ejemplo básico del proceso llevado a cabo por la calibración, al obtener la información necesaria para alinear las cámaras de un sistema estereoscópico de forma artificial.

que permita mantenerlas alineadas.

Ahora bien, con el objetivo de obtener los mejores resultados posibles, se procuró alinear las cámaras tanto vertical como horizontalmente de la mejor forma posible, procurando además, aproximar la distancia existente entre retinas oculares del ser humano, que es de entre 6.5 cm y 7.5 cm. Sin embargo, dado que las cámaras cuentan con un sistema de micrófono estéreo a los lados, la distancia mínima lograda, fue de 9.5 cm entre centros

de cámara como se muestra en la Figura 3.4.



Figura 3.4: Posición de las cámaras en el sistema.

Por otro lado, ya que se juntaron y alinearon las cámaras, se diseñó una caja de acrílico para empotrarlas y mantenerlas fijas. Las dimensiones de la caja son: 21.5 cm de ancho, 6.0 cm de altura y 9.0 cm de profundidad. Las cámaras cuentan con un orificio para ser atornilladas en la base, por lo cuál, fue posible respetar la posición de los dispositivos según se muestra en la Figura 3.4. Por otro lado, se hicieron dos aberturas en la parte frontal de la caja, para empotrar los lentes de las cámaras y evitar que se muevan. Las dimensiones de las aberturas son de aproximadamente 6.5 a 7.0 cm de ancho y de 2.8 a 3.0 cm de alto. El diseño de la caja se puede observar en la Figura 3.5.

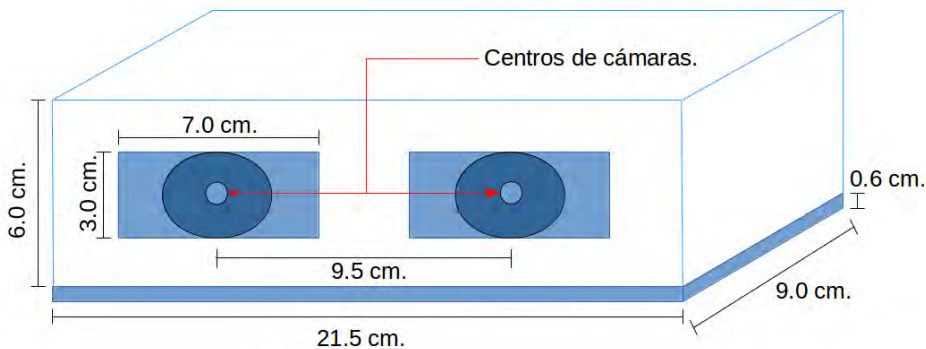


Figura 3.5: Diseño de la caja para el empotrado de las cámaras.

3.1.3. Ensamblado final del sistema

Se añadió un hub usb al sistema de cámaras para energizarlas de forma externa, mientras que para la protección tanto de las cámaras, como del cableado, se colocó una tapadera de acrílico, como se observa en las Figuras 3.6 y 3.7.



Figura 3.6: Presentación del ensamblado del modelo.



Figura 3.7: Ensamblado final del modelo.

Ya con el sistema armado, físicamente alineado y sin riesgo de que las cámaras se muevan, es posible continuar con el proceso de calibración del sistema.

3.2. Calibración del sistema estéreo

La calibración de un sistema estéreo es un proceso común, y no es difícil encontrar información al respecto. Sin embargo, una herramienta que ha sido crucial en los últimos años, es la librería OpenCV. A grandes rasgos, esta librería se encuentra disponibles para lenguajes como C, C++, Python, e inclusive Java, y cuenta con un conjunto de diferentes algoritmos y funciones optimizadas para todo lo relacionado con el tratamiento y procesamiento de imágenes digitales, sin mencionar el amplio soporte y bibliografía que respaldan dicha librería (Gary Bradski *et al.* [Bradski08]). Autores como Kenneth Dawson-Howe [Dawson-Howe14], ó Robert Laganiere [Laganiere11], y trabajos como los de Roberto Rangel [Heras] y Patricia Compañ [Compañ04], por mencionar algunos, fundamentan o involucran OpenCV en su investigación. Por esta razón, el presente proyecto se desarrolló utilizando la librería OpenCV, en la parte de implementación.

3.2.1. Patrón de calibración para el sistema estéreo

En la literatura, existen diferentes metodologías basadas en el uso de objetos tridimensionales para la calibración de un sistema estéreo de cámaras (generalmente cubos o cajas con un conjunto de marcas), sin embargo, el uso de patrones planos como un tablero de ajedrez, son más sencillos de utilizar y más sencillos de construir, ya que de hecho, es más difícil construir, almacenar y transportar un patrón tridimensional de calibración, que un patrón impreso en una hoja de papel (Gary Bradski *et al.* [Bradski08]). Por tanto, para el presente proyecto, se utilizó un tablero de ajedrez con un patrón alternado entre cuadros de color blanco y negro (Figura 3.9), que además de ser fácil de utilizar, asegura la existencia de parcialidad a la hora de obtener referencias del patrón, en todos los ejes coordenados de este. Además, como la calibración se dá a partir de un conjunto de puntos esquina, el uso de un tablero de ajedrez, facilita el orden de la relación existente entre las esquinas de los cuadros de forma natural. Dicho tablero puede ser de diferentes dimensiones, y la cantidad de recuadros blancos y negros puede ser configurada a gusto del usuario, sin embargo, OpenCV ofrece la opción de descargar la imagen del patrón de ajedrez, el cuál está compuesto por 6 puntos esquina en las filas, 9 puntos esquina en las columnas, y di-

mensiones por recuadro de 2.5×2.5 centímetros. Es posible descargar dicho patrón del link: [http : //docs.opencv.org/2.4/downloads/pattern.png](http://docs.opencv.org/2.4/downloads/pattern.png). Además, OpenCV ofrece una función para la detección de los puntos en el patrón, con la función `findChessboardCorners()`, la cuál será explicada más adelante.

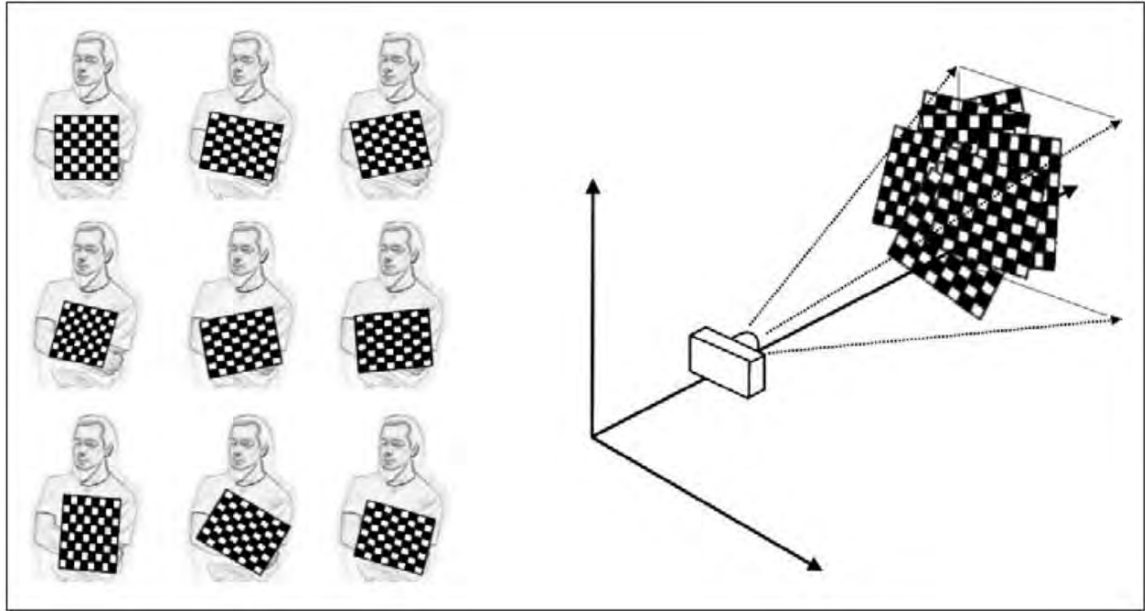


Figura 3.8: Imágenes del tablero de ajedrez con distintas orientaciones. Gary Bradski *et al.* [Bradski08]

Por otro lado, de acuerdo con Gary Bradski *et al.* [Bradski08], para poder calibrar un sistema de cámaras, mientras mayor sea el número de imágenes utilizadas, con el tablero en diferentes posiciones, mejores son los resultados, ya que el algoritmo de calibración tendrá más información para determinar la localización de los puntos esquina, con coordenadas del mundo real (Figura 3.8).

Por lo antes mencionado, para el presente trabajo, se tomó una cantidad de 24 pares de imágenes del patrón de ajedrez, para 30 distintas pruebas de calibración realizadas, por lo que se determinó que para obtener mejores resultados de calibración, se deben tomar en cuenta las siguientes recomendaciones:

- Se debe imprimir el patrón en papel de consistencia rígida, o mantenerse adherido a

una superficie sólida para evitar deformaciones en el papel que dificulten el proceso de calibración (Figura 3.9).

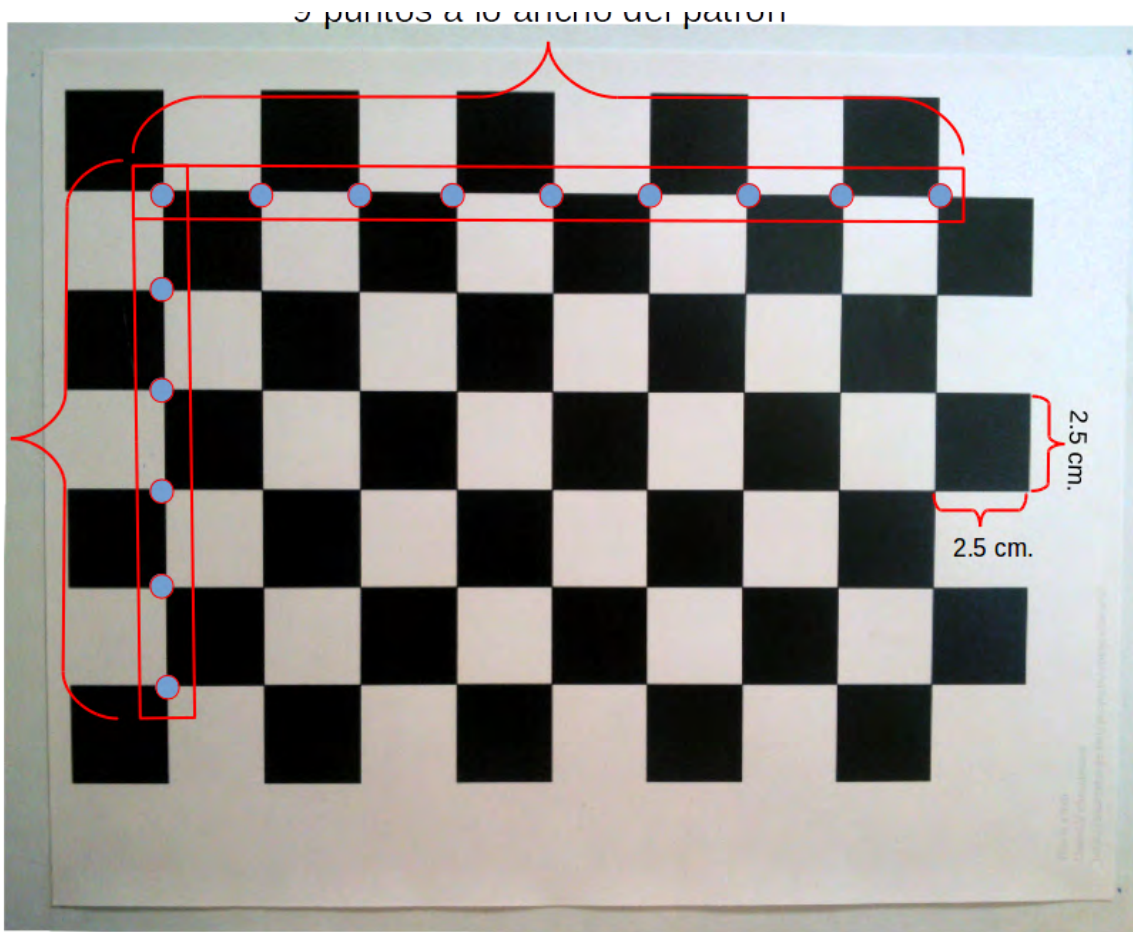


Figura 3.9: Patrón de tablero de ajedrez impreso en papel opalina opaco, pegado sobre papel cascaron blanco y acrílico, para mantener la rigidez.

- Si el patrón utilizado tiene una base color de blanco, se debe procurar mantener la superficie limpia y libre de polvo para evitar puntos falsos de información.
- Es recomendable imprimir el patrón en papeles opacos para evitar el brillo excesivo en las imágenes a la hora de tomar las fotos.
- Para las muestras no existe un número definido de capturas, pero se recomienda tomar tantas fotos como sea necesario, procurando abarcar todo tipo de orientaciones en el

espacio como se muestra en la Figura 3.10.



Figura 3.10: Muestra de 24 pares de imágenes usadas para la calibración.

- La iluminación del lugar debe ser adecuada, evitar luz excesiva, y baja intensidad lumínica, ya que esto puede provocar que los objetos de la escena se vean con diferentes intensidades de color para las cámaras.

Tomando en cuenta estas recomendaciones, determinadas a partir de diferentes pruebas realizadas, es posible obtener buenos resultados a la hora de calibrar.

3.2.2. Calibración con OpenCV

Ya que se hizo un análisis matemático de lo que son los parámetros intrínsecos y extrínsecos, y su importancia para un sistema estéreo, además de tener una noción práctica de como obtener un conjunto de imágenes muestra para asegurar la obtención de buenos resultados en la calibración, conviene revisar las funciones que OpenCV ofrece a la hora de llevar a cabo el proceso de calibración, y los resultados de dichas funciones a la hora de rectificar las cámaras del sistema.

De acuerdo con autores como Kenneth Dawson-Howe [Dawson-Howe14], ó Robert Laganier [Laganier11], la función `cvStereoCalibrate()` es la encargada de llevar a cabo el cálculo de la matriz de rotación y traslación que relacionan geoméricamente a la cámara derecha con la izquierda. Sin embargo son Gary Bradski *et al.* [Bradski08] quienes proponen el algoritmo **StereoCalib**, el cuál conjunta diferentes funciones de OpenCV, no solo para calibrar, si no también para rectificar las cámaras del sistema estéreo, como se muestra en el Algoritmo 1, donde *listaDeImágenes* es una lista de las imágenes muestra previamente capturadas, *nx* es el número de recuadros que hay verticalmente en el patrón, *ny* los recuadros que hay horizontalmente, y *tamañoCuadros* es el tamaño de los recuadros en el patrón, medido en centímetros. Por otro lado, el algoritmo entrega como resultado las matrices de parámetros intrínsecos K_l y K_r para las cámaras izquierda y derecha respectivamente, la matriz de rotación R_r para la cámara derecha y la matriz de rotación R_l para la cámara izquierda, además de un vector de traslación T_f optimizado, y un conjunto de matrices de mapeo $mapeo_x_r$, $mapeo_x_l$, $mapeo_y_r$ y $mapeo_y_l$, que al ser aplicadas en las imágenes tomadas por el sistema, remapean los puntos contenidos en dichas imágenes, a la nueva posición rectificadas.

Función `cvFindChessboardCorners`

OpenCV proporciona una función para la obtención de información de un sistema estéreo, partiendo de un patrón de tablero de ajedrez, con la función llamada `FindChessboardCorners()`. Esta función, localiza los puntos esquina en cada una de las imágenes muestra, para posteriormente almacenarlos en una matriz de puntos, que será utilizada pa-

Algoritmo 1: StereoCalib(*listaDeImágenes*, *nx*, *ny*, *tamañoCuadros*).

- 1: **Para** toda imagen $I \in \text{listaDeImágenes}$ **hacer**
 - 2: **Calcular** *puntosEsquina* usando `cvFindChessboardCorners()` y w .
 - 3: **fin Para**
 - 4: **Calcular** R y T con `cvStereoCalibrate()` usando *puntosEsquina*
 - 5: **Calcular** K_r , K_l , R_r , R_l , T_r y T_l con `cvStereoRectify()` usando R y T
 - 6: **Calcular** $\text{mapeo}x_r$, $\text{mapeo}x_l$, $\text{mapeo}y_r$ y $\text{mapeo}y_l$ con `cvInitUndistortRectifyMap()` usando R_r , R_l , T_r y T_l
 - 7: **Calcular** K_r , K_l , R_r , R_l , T_f , $\text{mapeo}x_r$, $\text{mapeo}x_l$, $\text{mapeo}y_r$ y $\text{mapeo}y_l$
-

ra el cálculo de R y T . En la Figura 3.11, se observa como para cada conjunto de puntos esquina encontrado, se asigna un color diferente, indicando que pertenecen a un mismo renglón, además, dichos puntos se unen con una línea del mismo color con su vecino correspondiente, en el orden que son encontrados. `FindChessboardCorners()` ofrece dos formas de búsqueda de los puntos esquina: de forma horizontal de izquierda a derecha, o de forma vertical, de arriba hacia abajo. Para el caso de esta tesis, se implementó dicha función con una búsqueda horizontal.

Función `cvStereoCalibrate`

Con la matriz de puntos esquina calculada por `cvFindChessboardCorners()`, es posible obtener los parámetros extrínsecos de un sistema estereoscópico utilizando la función `cvStereoCalibrate()`. De acuerdo con Gary Bradski *et al.* [Bradski08], esta función usa información a-priori de las cámaras, y los puntos esquina encontrados, con el objetivo de inicializar un algoritmo de optimización robusto llamado Levenberg-Marquardt (Manolis I. A. Lourakis [Lourakis05]), el cuál se enfoca en encontrar el error mínimo de re-proyección de los puntos esquina del tablero de ajedrez, para las vistas de ambas cámaras. Finalmente `cvStereoClibrate()` retorna una matriz de rotación R y un vector traslación T que relacionan geoméricamente a las cámaras del sistema.

Los parámetros para `cvStereoCalibrate()`, son los siguientes:

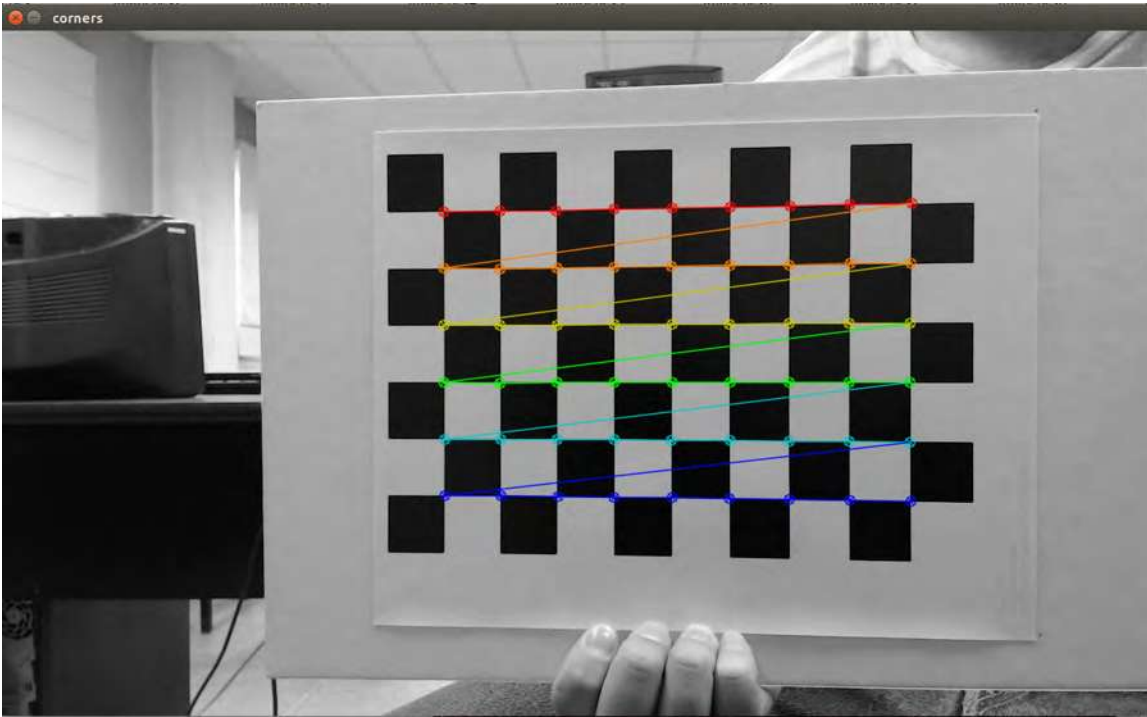


Figura 3.11: Detección de esquinas en el tablero de ajedrez con `FindChessboardCorners()`.

- `objectPoints`. Matriz de tamaño $N \times 3$ que almacena las coordenadas tridimensionales de cada punto “ P ” en función de las N imágenes muestra tomadas por el sistema estéreo, es decir, $objectPoints = P \times N$.
- `ImagePoints1` y `ImagePoints2`. Matrices de tamaño $N \times 2$ que contienen las coordenadas de los pixeles de la imagen derecha e izquierda respectivamente, en referencia de la información de la matriz $objectPoints$.
- `CameraMatrix1` y `CameraMatrix2`. Matrices de la cámara de 3×3 que la misma función calcula.
- `DistCoeff1` y `DistCoeff2`. Matrices de distorsión que de tamaño 5×1 , compuestas en las dos primeras posiciones por coeficientes de distorsión radial, seguidos por dos coeficientes de distorsión tangencial y un tercer coeficiente de distorsión radial que más actualmente fue añadido a la función para el efecto de ojo de pez, como lo

menciona Gary Bradski *et al.* [Bradski08].

- `ImageSize`. Tamaño de las imágenes en píxeles.
- R y T . Matriz de Rotación y vector de traslación entregados por la función al finalizar su ejecución.
- E y F . Parámetros opcionales para el cálculo de la matriz esencial y la matriz fundamental para el par estéreo.
- `CvTermCriteria`. Parámetros para el algoritmo de minimización Levenberg-Marquardt, donde se especifica el número de iteraciones y el umbral para la condición del parada.

Función `cvStereoRectify`

Una vez calculada la matriz de rotación R , y el vector de traslación T , se procede a calcular los parámetros intrínsecos y extrínsecos del sistema que rectificarán las cámaras, utilizando la función llamada `cvStereoRectify()`. Esta función calcula las matrices de rotación y traslación para cada cámara, que artificialmente alinearán ambas cámaras al mismo plano de imagen, es decir, las volverá coplanares. Consecuentemente, esto permitirá que los píxeles en las imágenes de ambas cámaras, queden correctamente alineados. La función retornará las siguientes matrices: K_r , K_l , R_r y R_l .

Los parámetros necesarios para `stereoRectify()` son:

- `cameraMatrix1` y `cameraMatrix2`. Matrices de las cámaras.
- `distCoeffs1` y `distCoeffs2`. Coeficientes de distorsión de la primer y segunda cámara respectivamente.
- `ImageSize`. Tamaño de las imágenes usadas en la función anterior.
- R . Matriz de rotación obtenidas con la función `StereoCalibrate()`.
- T . Vector de traslación obtenido con la función `StereoCalibrate()`.
- R_r . Matriz de rotación rectificadas para la cámara derecha de tamaño 3×3 .

- R_l . Matriz de rotación rectificadora para la cámara izquierda de tamaño 3×3 .
- K_r . Matriz de proyección rectificadora para la cámara derecha, de tamaño 3×4 , añadiendo el vector de traslación T_r a la matriz de parámetros intrínsecos.
- K_l . Matriz de proyección rectificadora para la cámara izquierda, de tamaño 3×4 , añadiendo el vector de traslación T_l a la matriz de parámetros intrínsecos.
- Q . Matriz de mapeo de disparidad-a-profundidad de tamaño 4×4 .

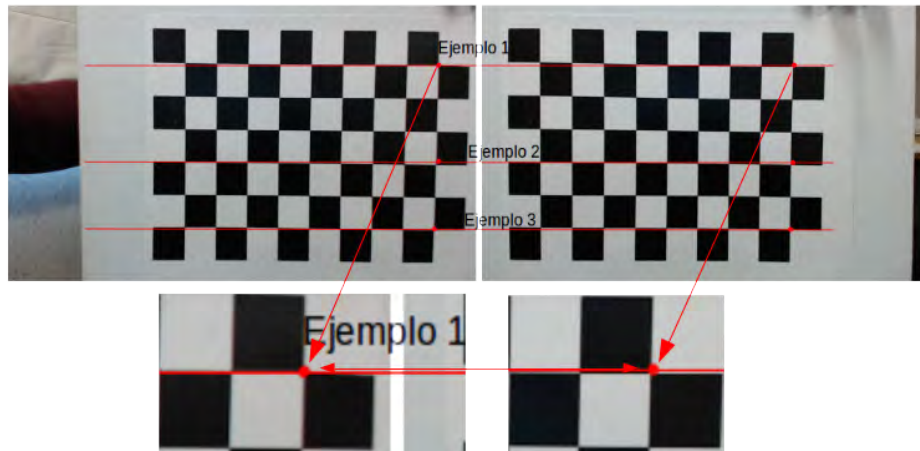
Entre otros parámetros opcionales que pueden variar dependiendo el sistema estereoscópico.

Función `cvInitUndistortRectifyMap`

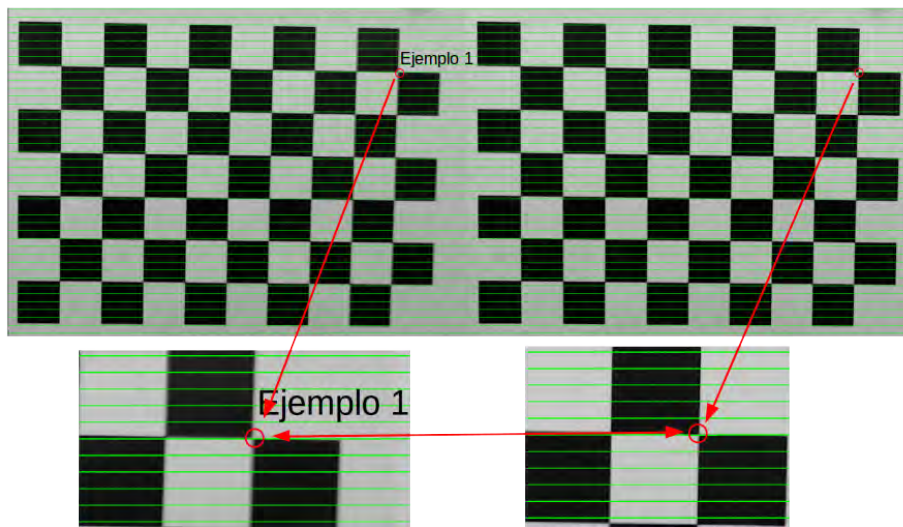
Ya con con las matrices de parámetros intrínsecos y extrínsecos calculadas, OpenCV implementa la función `InitUndistortRectifyMap()`. La finalidad de esta función, es mapear cada punto de las imágenes originales a una nueva posición en coordenadas de imagen. Esta función obtiene una matriz de remapeo para el eje de la x y otra para el eje de la y de forma individual para cada una de las cámaras por separado. Finalmente se aplican los resultados al sistema por medio de la función `Remap()`, y se obtienen las imágenes rectificadas. En la Figura 3.12, se aprecia un ejemplo del resultado de calibración llevado a cabo para el sistema estereoscópico implementado en este trabajo, donde se observa un par de imágenes rectificadas tomando como referencia tres líneas horizontales en color rojo, que como se puede observar, tocan los mismo puntos en ambas imágenes, remarcados con color rojo también.

En la Figura 3.12(a), posible notar que la línea roja no coincide en las esquinas marcadas con un punto rojo en ambas imágenes. La línea roja que atraviesa ambas imágenes a lo ancho, aparece ligeramente más arriba sobre la esquina marcada en la imagen derecha, indicando que las imágenes están desalineadas. En contraste, en la Figura 3.12(b) aparecen un conjunto de líneas verdes conocidas como líneas epipolares, las cuales pueden ser usadas como referencia, para visualizar si las imágenes fueron correctamente rectificadas. Como

ejemplo, es posible notar que en las esquinas circuladas en rojo, las líneas epipolares pasan por el mismo lugar en ambas, indicando que las cámaras fueron correctamente rectificadas.



(a) Imágenes tomadas con las cámaras sin calibrar.



(b) Imágenes tomadas con las cámaras calibradas.

Figura 3.12: Comparativa entre imágenes tomadas por el sistema estéreo, antes y después de calibrar y rectificar dicho sistema.

Ejemplo de mapeo usando el modelo y los parámetros del sistema estereo.

Con las funciones descritas anteriormente implementadas en el algoritmo desarrollado por Gary Bradski *et al.* [Bradski08], y la calibración aplicada al sistema estereoscópico, se obtuvieron los intrínsecos K_r y R_r para la cámara derecha, y K_l y R_l para la cámara izquierda, extrayendo el parámetro de distancia d del vector de traslaciones T para el sistema de cámaras, como se muestra en (3.3), (3.4), (3.5), y (3.6).

$$K_r = \begin{bmatrix} -944.54 & 0.0 & 657.92 \\ 0 & -944.54 & 364.67 \\ 0.0 & 0.0 & 1.0 \end{bmatrix} \quad (3.3)$$

$$[R_r|T_r] = \begin{bmatrix} 0.9999559 & -0.0038788 & -0.0085438 & 0 \\ 0.0038833 & 0.9999923 & 0.0005166 & 0 \\ 0.0085417 & -0.0005417 & 0.9999633 & 0 \end{bmatrix} \quad (3.4)$$

$$K_l = \begin{bmatrix} -944.54 & 0.0 & 660.89 \\ 0.0 & -944.54 & 364.67 \\ 0.0 & 0.0 & 1.0 \end{bmatrix} \quad (3.5)$$

$$[R_l|T_l] = \begin{bmatrix} 0.9999728 & -0.0018403 & -0.0071395 & -9.48 \\ 0.0018365 & 0.9999981 & -0.0005397 & 0.0 \\ 0.0071405 & 0.0005266 & 0.9999674 & 0.0 \end{bmatrix} \quad (3.6)$$

Como ejemplo, en la Figura 3.13 se observan dos imágenes tomadas por la cámara izquierda y derecha respectivamente. En ambas imágenes es posible observar un punto en el espacio, con coordenadas tridimensionales de 10.0 cm en el eje X , 5.0 cm de altura en el eje Y , y 20.0 cm de profundidad en el eje Z , tomando como referencia el centro de la cámara derecha, para las coordenadas del punto P_1 .

Utilizando los parámetros del modelo mostrados en (3.3), (3.4), (3.5), y (3.6), y tomando los valores del punto P_1 , es posible mapear las coordenadas de imagen del punto observado, tanto en la imagen izquierda, como en la imagen derecha. Esto se lleva a cabo, usando (3.7) y (3.8).

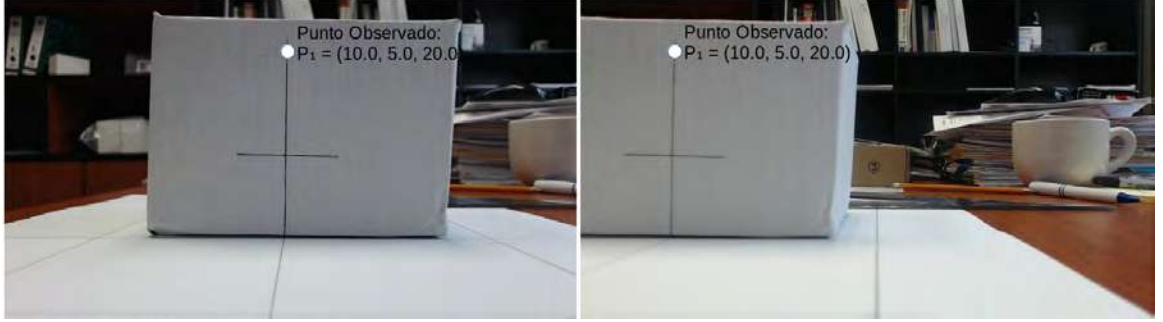


Figura 3.13: Punto P_1 en el espacio tridimensional con coordenadas $[10.0 \text{ cm}, 5.0 \text{ cm}, 20.0 \text{ cm}]^T$.

$$\begin{bmatrix} z_0 \cdot x_r \\ z_0 \cdot y_r \\ z_0 \end{bmatrix} = \begin{bmatrix} -944.54 & 0.0 & 657.92 \\ 0 & -944.54 & 364.67 \\ 0.0 & 0.0 & 1.0 \end{bmatrix} \begin{bmatrix} 0.9999559 & -0.0038788 & -0.0085438 & 0 \\ 0.0038833 & 0.9999923 & 0.0005166 & 0 \\ 0.0085417 & -0.0005417 & 0.9999633 & 0 \end{bmatrix} \begin{bmatrix} 10.0 \\ 5.0 \\ 20.0 \\ 1 \end{bmatrix} \quad (3.7)$$

$$\begin{bmatrix} z_0 \cdot x_l \\ z_0 \cdot y_l \\ z_0 \end{bmatrix} = \begin{bmatrix} -944.54 & 0.0 & 660.89 \\ 0.0 & -944.54 & 364.67 \\ 0.0 & 0.0 & 1.0 \end{bmatrix} \begin{bmatrix} 0.9999728 & -0.0018403 & -0.0071395 & -9.48 \\ 0.0018365 & 0.9999981 & -0.0005397 & 0.0 \\ 0.0071405 & 0.0005266 & 0.9999674 & 0.0 \end{bmatrix} \begin{bmatrix} 10 \\ 5.0 \\ 20.0 \\ 1 \end{bmatrix} \quad (3.8)$$

Mediante (3.7) y (3.8), se determina que las coordenadas de imagen del punto P_1 en la cámara derecha e izquierda respectivamente, son los siguientes:

$$\begin{bmatrix} x_r \\ y_r \\ z_0 \end{bmatrix} = \begin{bmatrix} 196.545 \\ 127.193 \\ 1 \end{bmatrix} \quad (3.9)$$

$$\begin{bmatrix} x_l \\ y_l \\ z_0 \end{bmatrix} = \begin{bmatrix} 643.769 \\ 129.05 \\ 1 \end{bmatrix} \quad (3.10)$$

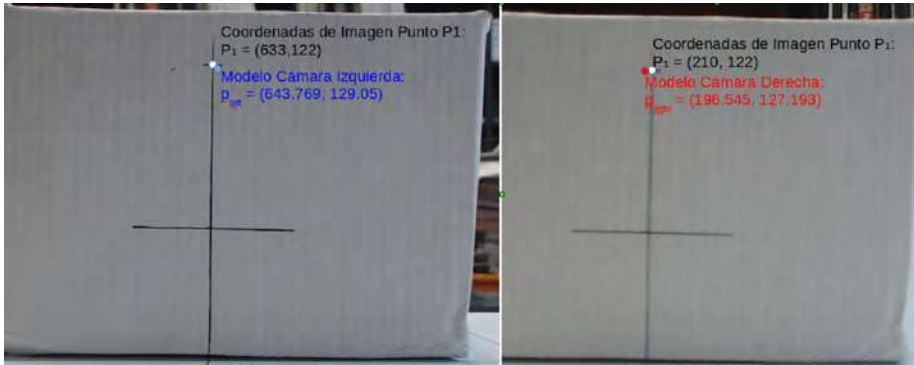


Figura 3.14: Mapeo del punto P_1 a coordenadas de la imagen izquierda y derecha respectivamente, en comparación con las coordenadas calculadas con el modelo de las cámaras.

Como se muestra en la Figura 3.14, los puntos obtenidos en (3.9) y (3.10) son muy próximos a las coordenadas de imagen del punto observado P_1 , demostrando que tanto la calibración del sistema es correcta, como el modelo deducido en (2.33) y (2.34) es adecuado. Por otro lado, con el objetivo de apreciar la profundidad de los objetos en una escena, es posible crear imágenes anaglifo, utilizando el sistema de cámaras estéreo, como se describe en la sección siguiente.

3.3. Imágenes anaglifo como marco de referencia para estimar profundidad

Es posible definir como anaglifo, a las imágenes bidimensionales que pueden generar un efecto tridimensional, o de profundidad, cuando estas son vistas con gafas especiales. Este tipo de imágenes van muy de la mano con la estereoscopia, ya que ofrecen una forma práctica y didáctica de apreciar el efecto de profundidad, a partir de la disparidad ocular si son generados con un sistema estéreo de cámaras. Sin embargo, este tipo de imágenes pueden ser creadas con diferentes técnicas, siendo de lo más común utilizar editores de imágenes. Por otra parte, las gafas anaglifo, están diseñadas para filtrar específicamente el color Rojo con una de las lentes, y el color Verde y Azul con la otra lente como se muestra en la Figura 3.15, y son pieza fundamental para apreciar el efecto de profundidad en este tipo de imágenes.

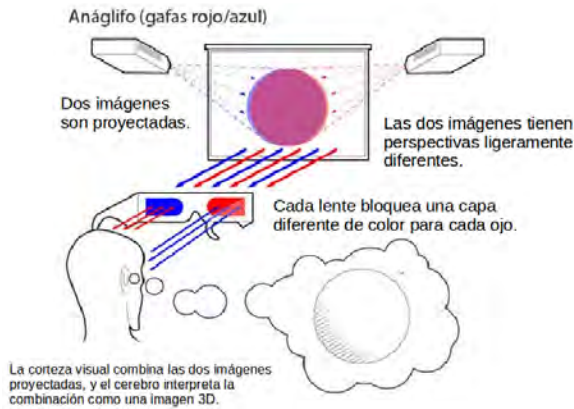


Figura 3.15: Funcionamiento de las gafas especiales para observar imágenes anaglifo. Imagen de Vicente Fuentes [Fuentes14b]

3.3.1. Imágenes anaglifo a partir de un sistema estéreo

La creación de un anaglifo, puede ser modelado siguiendo el ejemplo mostrado en la Figura 3.16, donde se puede observar que I_A se define como la nueva imagen anaglifo, siendo 0 el índice que identifica al canal rojo en un pixel, 1 el índice del canal verde, y 2 el índice para el canal azul en el mismo pixel. Por otro lado, I_L es la imagen tomada por la cámara izquierda, de la cuál, para todos los pixeles con coordenadas x, y en I_A se tomarán los valores de los canales verde y azul para todos los pixeles correspondientes en I_L . Por el contrario, siendo I_R la imagen tomada por la cámara derecha, se extraen únicamente los valores del canal rojo de todos los pixeles en dicha imagen, correspondiendo a todos los pixeles en el anaglifo I_A .

Por otro lado, dando por hecho que ya se tiene un sistema estereoscópico calibrado y rectificado, se propone el Algoritmo 2 para creación de anaglifos, tomando como base a (3.11), (3.12), y (3.13).

$$I_A(0, x, y) = I_R(0, x, y) \quad (3.11)$$

$$I_A(1, x, y) = I_L(1, x, y) \quad (3.12)$$

$$I_A(2, x, y) = I_L(2, x, y) \quad (3.13)$$

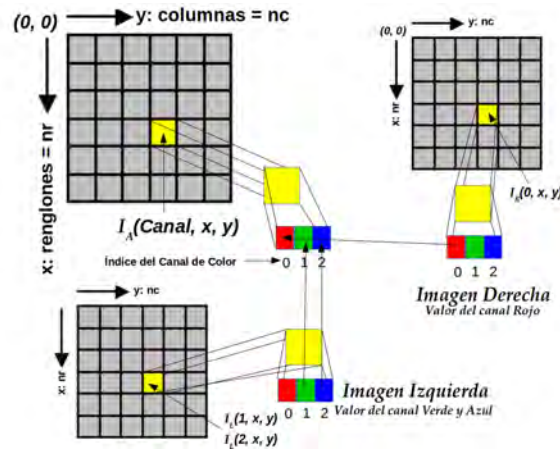


Figura 3.16: Ejemplo de creación de imágenes anaglifo a partir de un par de imágenes estereoscópicas.

3.3.2. Ejemplo de imágenes anaglifo creadas con el sistema estéreo

Siguiendo el proceso mostrado por el Algoritmo 2, y utilizando el sistema estéreo calibrado y rectificado para este trabajo de tesis, se generó el anaglifo mostrado en la Figura 3.17.

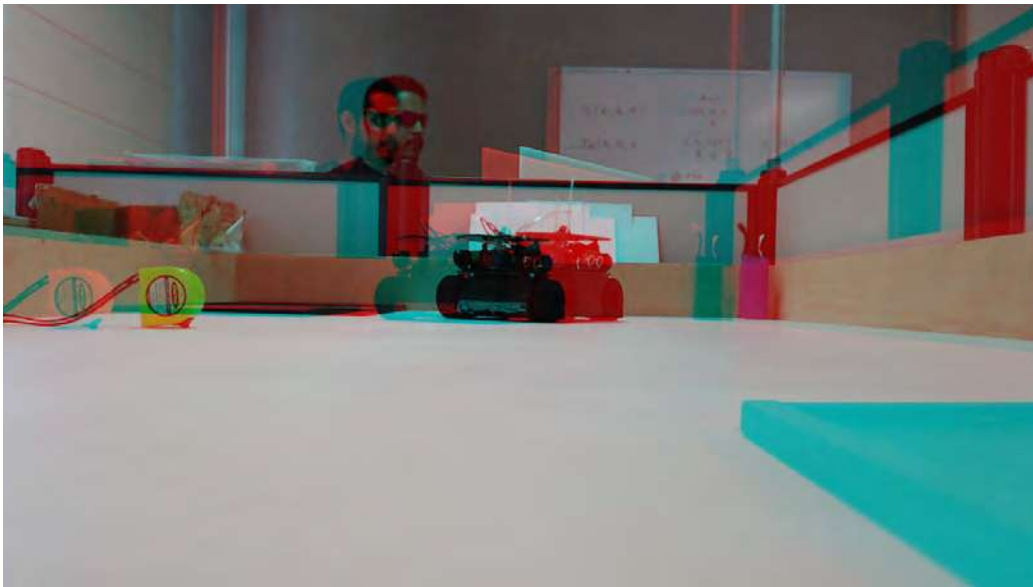


Figura 3.17: Anaglifo generado a partir de imágenes con el sistema estéreo calibrado.

Algoritmo 2: imagenAnaglifo(I_L, I_R).

```

1: Dadas las imágenes  $I_L$  e  $I_R$ :
2: Obtener nr igual al número de renglones en la imagen  $I_R$ 
3: Obtener nc igual al número de columnas en la imagen  $I_R$ 
4: Crear imagen  $I_A$  de tamaño  $nr \times nc$ 
5: para  $x \leftarrow 1, nr$  hacer
6:   para  $y \leftarrow 1, nc$  hacer
7:      $I_A(0, x, y) = I_R(0, x, y)$ 
8:      $I_A(1, x, y) = I_L(1, x, y)$ 
9:      $I_A(2, x, y) = I_L(2, x, y)$ 
10:  fin para
11: fin para
12: Retornar  $I_A$ 

```

Como primer punto, en la imagen se puede apreciar el efecto de la calibración. Los objetos en color rojo pertenecientes a la perspectiva de la cámara derecha, y los objetos en azul o verde pertenecientes a la perspectiva de la cámara izquierda, se ven a la misma altura y correctamente alineados entre sí, por lo que se puede concluir que la calibración y la rectificación alinearon correctamente las cámaras.

Como segundo punto, es claramente visible la relación de profundidad-disparidad. Si se observa con detenimiento la Figura 3.17 sin utilizar las gafas anaglifo, es posible notar que los objetos con mayor distancia respecto de las cámaras, tienen una disparidad menor. Por el contrario, en los objetos más cercanos a las cámaras, se aprecia una separación mayor (o mayor disparidad), en sus componentes de color, como se muestra en la Figura 3.18.

Otro ejemplo de un anaglifo generado con el sistema de cámaras, se muestra en la Figura 3.19, donde como se mencionaba anteriormente, sin usar las gafas especiales, se puede apreciar la diferencia de disparidades conforme más lejos se encuentren los objetos.

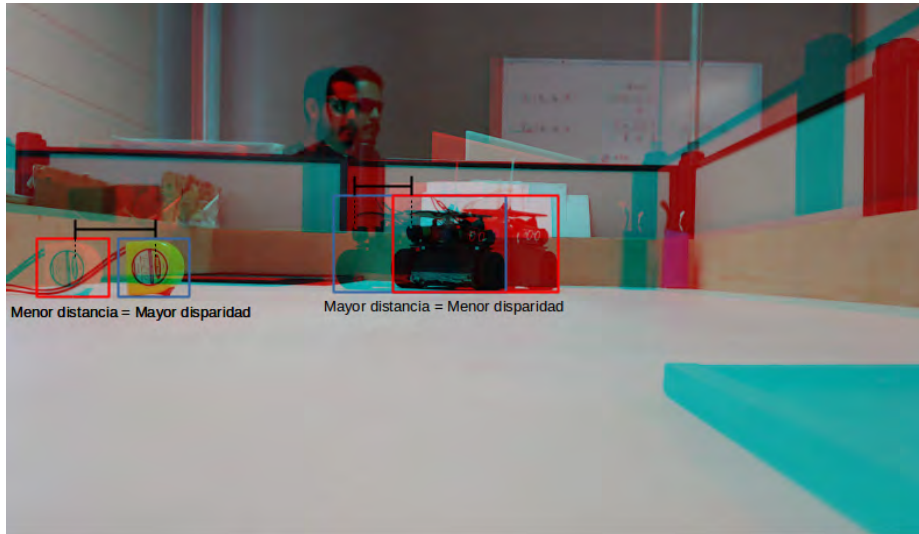


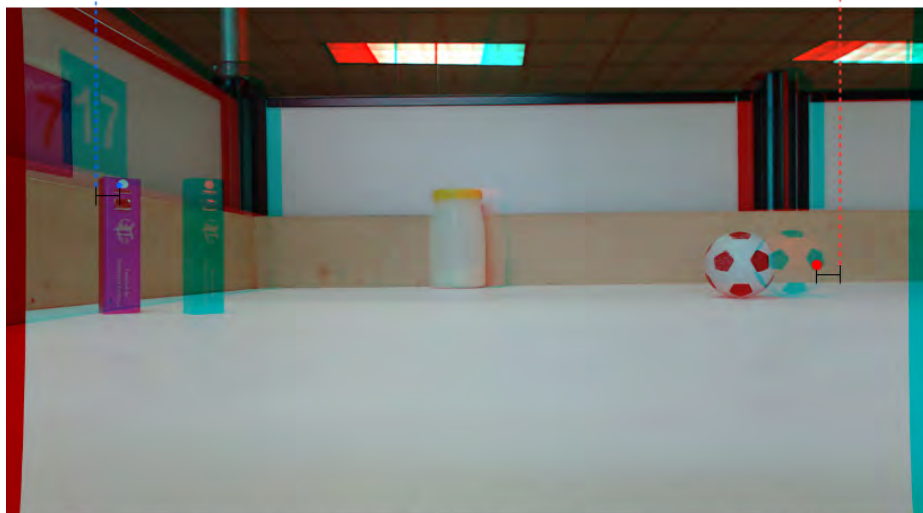
Figura 3.18: En este anaglifo es posible observar como las componentes de color de los objetos que se encuentran a mayor profundidad, se juntan o tienen una menor disparidad. Por el contrario, se puede observar que para el objeto más cercano (la cinta métrica), las componentes de color están completamente separadas, indicando que hay una mayor disparidad.

3.4. Marco de referencia para la estimación de la profundidad con el sistema estéreo

Con los ejemplos anteriores, es posible determinar que las imágenes anaglifo son una forma práctica para obtener una idea preliminar de la profundidad de diferentes objetos en una escena. De hecho, es posible validar la existencia de una relación directa entre la disparidad y la profundidad, como se aprecia en 3.17 y 3.19, respaldada por las ecuaciones de dichos conceptos ((2.37) y (2.38) respectivamente), desarrollados en el capítulo anterior. Por tales razones, y acorde a lo mencionado por Gary Bradski *et al.* [Bradski08], es posible concluir que la relación de proporción “Disparidad-Profundidad”, indica que a mayor distancia, menor disparidad, siendo la disparidad mínima alcanzada, cercana a cero, cuando la profundidad tiende al infinito como se expresa en (3.14), donde Z representa la profundidad de un punto en el espacio tridimensional, y $x_l - x_r$ la disparidad de cada punto.



a) Anaglifo sin desplazamiento de centros.



b) Anaglifo con desplazamiento de centros de $\alpha = 38$ píxeles.

Figura 3.19: Anaglifo generado a partir del sistema de cámaras implementado para este trabajo de tesis.

$$Si \ x_l - x_r \begin{cases} \approx 0 & \therefore Z \rightarrow \infty \\ \neq 0 \text{ y positiva} & \therefore Z \rightarrow 0 \end{cases} \quad (3.14)$$

El comportamiento modelado con (3.14), es posible observarlo en la gráfica de

la Figura 3.20, donde utilizando los parámetros del sistema estereo expresados en (3.3), (3.4), (3.5), y (3.6), se calcularon las coordenadas de imagen $[x_r, y_r, z_0]$ y $[x_l, y_l, z_0]$ de un punto tridimensional $P = [4.7, 0.0, Z]$, con (2.33) y (2.34), variando la profundidad representada por Z , desde 10 cm, hasta 10,000 cm con respecto del sistema de cámaras, para posteriormente calcular la disparidad $x_l - x_r$ de dicho punto, tomando en cuenta los diferentes valores de profundidad.

Curva de la relación existente entre la Disparidad y la Profundidad

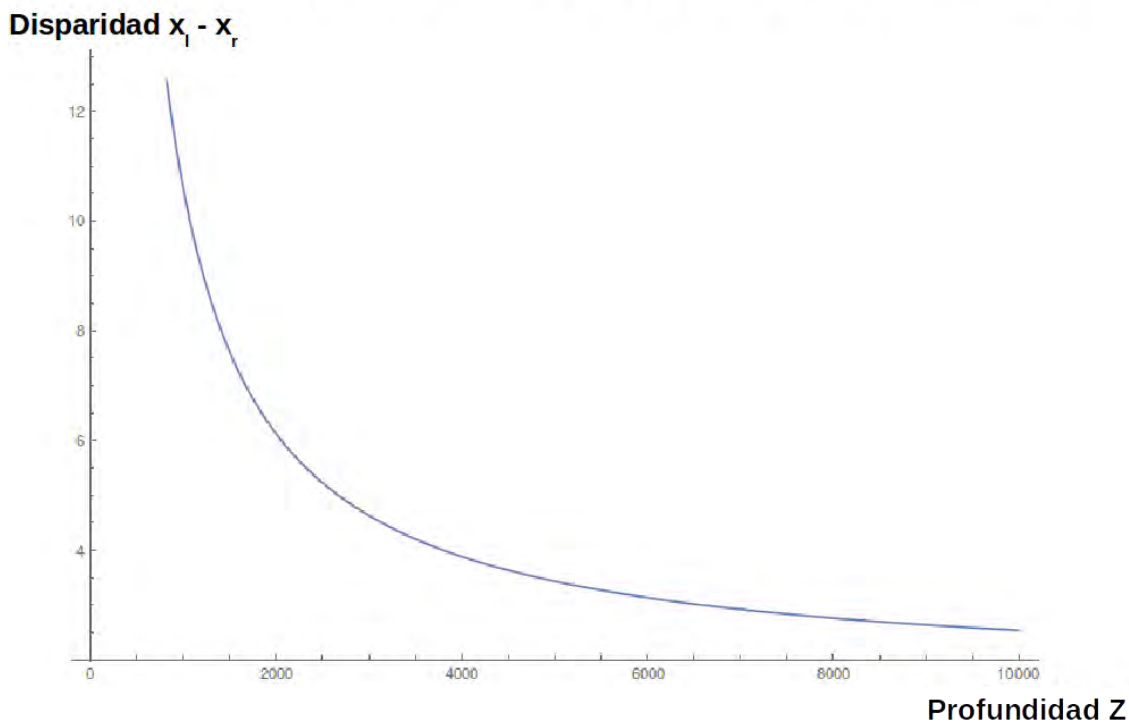


Figura 3.20: Esta gráfica muestra de forma clara como la curva se comporta de forma asintótica respecto del eje de la profundidad Z , es decir, conforme la disparidad disminuye, la profundidad se aproxima a cero, pero nunca alcanza este valor.

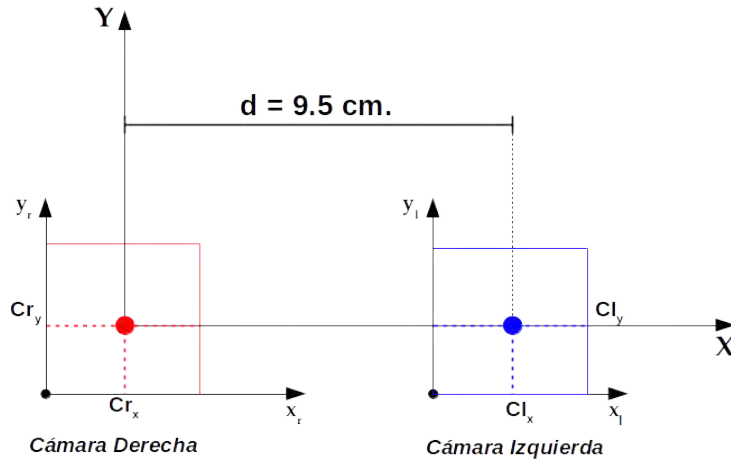
Pero a pesar de la información que los anaglifes generados con el sistema de cámaras puedan ofrecer con respecto de la disparidad y la profundidad, si se utilizan las gafas anaglifo para observar los anaglifes de las Figuras 3.17 ó 3.19, se llegará a la conclusión de que el efecto tridimensional no es fácilmente visible, se dificulta enfocar los objetos con los ojos, o es inclusive agotador para la vista. Esto sucede debido a que la distancia entre

los centros de las cámaras (9.5 cm aproximadamente), no se asemeja a la distancia promedio que existe entre las retinas oculares de la vista humana (6.5 cm aproximadamente), ocasionando que el efecto de profundidad en el anaglifo, no sea fácil de visualizar. Es por esta razón, que es necesario regresar al modelo de las cámaras, y hacer una modificación artificial que permita obtener mejores resultados a la hora de crear anaglifos, y además, permita crear un punto de referencia con el cuál sea posible aproximar la distancia de un punto con respecto de las cámaras.

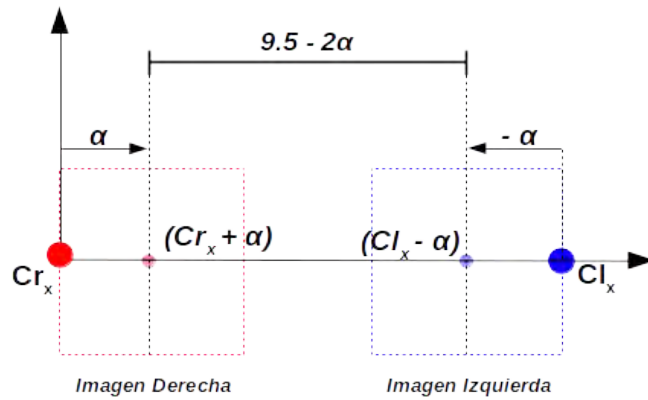
3.5. Reducir la distancia entre los centros de las cámaras de manera artificial

Como se mencionaba anteriormente, lo ideal es poder apreciar claramente los anaglifos sin forzar la vista, y sin dejar de obtener información preliminar de la profundidad de los objetos en la imagen. Sin embargo, por razones que no se pueden controlar, como es el diseño físico de las cámaras en particular, esto no es posible debido a que dicho diseño no permite reducir más la distancia entre las cámaras. Por esta razón, es necesario recurrir a una corrección artificial, tomando en cuenta las ecuaciones que modelan un sistema estéreo, representadas en (2.33) y (2.34). Así pues, es necesario enfocarse en la matriz de parámetros intrínsecos, y en los centros de cámara Cr_x y Cl_x . Se propone la adición de una nueva variable α al modelo de las cámaras, con el objetivo de obtener un desplazamiento artificial en las imágenes, como se muestra en la Figura 3.21.

Siguiendo entonces el diagrama mostrado en la Figura 3.21, es posible observar como α tomará un valor negativo para la cámara izquierda con el objetivo de generar un desplazamiento a la derecha, mientras que para la cámara derecha, tomará un valor positivo para simular un desplazamiento a la izquierda. En consecuencia, α se sumará a Cr_x en los parámetros intrínsecos de la cámara derecha, y se restará a Cl_x en los parámetros intrínsecos de la cámara izquierda, por lo que las ecuaciones desarrolladas en el Capítulo 2 para el modelo de las cámaras (2.33) y (2.34), quedará expresado como se muestra en (3.15) y (3.16).



(a) Modelo convencional de las cámaras con una distancia entre centros de cámaras $d = 9.5$ cm. aproximadamente.



(b) Modelo de las cámaras aplicando α a los centros de cámaras, lo cuál ocasiona que las imágenes derecha e izquierda se aproximen, reduciendo la distancia $d = 9.5 - 2\alpha$

Figura 3.21: Modelo de las cámaras antes y después de aplicar un desplazamiento a los centros de cámaras con la variable α , en la Figura 3.5, y 3.5 respectivamente.

$$\begin{bmatrix} z_0 \cdot x_r \\ z_0 \cdot y_r \\ z_0 \end{bmatrix} = \begin{bmatrix} -f & 0 & Cr_x + \alpha \\ 0 & -f & Cr_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} Rr_{0,0} & Rr_{0,1} & Rr_{0,2} & 0 \\ Rr_{1,0} & Rr_{1,1} & Rr_{1,2} & 0 \\ Rr_{2,0} & Rr_{2,1} & Rr_{2,2} & 0 \end{bmatrix} \begin{bmatrix} X_0 \\ Y_0 \\ Z_0 \\ 1 \end{bmatrix} \quad (3.15)$$

$$\begin{bmatrix} z_0 \cdot x_l \\ z_0 \cdot y_l \\ z_0 \end{bmatrix} = \begin{bmatrix} -f & 0 & Cl_x - \alpha \\ 0 & -f & Cl_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} Rl_{0,0} & Rl_{0,1} & Rl_{0,2} & -d \\ Rl_{1,0} & Rl_{1,1} & Rl_{1,2} & 0 \\ Rl_{2,0} & Rl_{2,1} & Rl_{2,2} & 0 \end{bmatrix} \begin{bmatrix} X_0 \\ Y_0 \\ Z_0 \\ 1 \end{bmatrix} \quad (3.16)$$

Cabe mencionar que los centros de cámara en el eje y , Cr_y y Cl_y , no se ven afectados, ya que la alineación de las cámaras en ese eje se mantiene. Por otro lado, también es importante señalar que debido a que el desplazamiento artificial se da a nivel de imagen, α debe estar dado en pixeles.

Esta modificación a los parámetros del sistema, no solo ayudará a visualizar más cómodamente un anaglifo. Con este nuevo cambio, es posible obtener un marco de referencia, relacionando la profundidad de un punto con coordenadas tridimensionales en particular, con la disparidad igual a cero. Como ejemplo de este nuevo cambio, se toman nuevamente los parámetros intrínsecos del sistema de cámaras implementado, y siguiendo el modelo de (3.15) y (3.16), se suma y se resta α a los parámetros intrínsecos de la cámara derecha e izquierda respectivamente mostrados en (3.3) y (3.5), como se aprecia en (3.17) y (3.18). Para este trabajo de tesis, se decidió asignar un valor de 38 pixeles a la variable α , lo cuál será explicado más a detalle a lo largo de este tema.

$$K_r = \begin{bmatrix} -944.54 & 0.0 & 657.92 + 38 \\ 0 & -944.54 & 364.67 \\ 0.0 & 0.0 & 1.0 \end{bmatrix} \quad (3.17)$$

$$K_l = \begin{bmatrix} -944.54 & 0.0 & 660.89 - 38 \\ 0.0 & -944.54 & 364.67 \\ 0.0 & 0.0 & 1.0 \end{bmatrix} \quad (3.18)$$

Llevando a cabo las operaciones en (3.17) y (3.18), se obtienen las nuevas matrices de parámetros intrínsecos como se muestra a continuación:

$$K_r = \begin{bmatrix} -944.54 & 0.0 & 692.92 \\ 0 & -944.54 & 364.67 \\ 0.0 & 0.0 & 1.0 \end{bmatrix} \quad (3.19)$$

$$K_l = \begin{bmatrix} -944.54 & 0.0 & 625.89 \\ 0.0 & -944.54 & 364.67 \\ 0.0 & 0.0 & 1.0 \end{bmatrix} \quad (3.20)$$

Con los nuevos valores de los parámetros intrínsecos ((3.19) y (3.20)) para las cámaras, es necesario pasar nuevamente por la función de OpenCv `cvInitUndistortRectifyMap()` con el objetivo de obtener las nuevas matrices de mapeo para generar las nuevas imágenes desplazadas, mostradas en la Figura 3.22. En consecuencia, debido al cambio aplicado, aparece una franja negra en la orilla de las imágenes, como indicativo de que estas sufrieron un desplazamiento hacia la derecha o izquierda, con el objetivo de simular un acercamiento entre los centros de cámaras mayor al desplazamiento que es físicamente posible en las cámaras. Ahora bien, es posible llevar a cabo otra prueba de mapeo, tomando un punto en el espacio con coordenadas tridimensionales de aproximadamente de 4.7 cm de distancia en el eje X con respecto del centro de la cámara derecha, 5.0 cm de altura en el eje Y con respecto del origen de referencia de las cámaras, y 20.0 cm de profundidad en el eje Z con respecto de las cámaras, siendo $X_0 = 4.7$, $Y_0 = 5.0$ y $Z_0 = 20.0$. Dicho punto, identificado como punto P_2 , o punto observado 2, como se muestra en la Figura 3.23.

De igual forma como se hizo en (3.7) y (3.8), pero esta vez con los parámetros intrínsecos mostrados en (3.19) y (3.20), se llevó a cabo la obtención de los puntos en coordenadas de las imágenes del punto observado, obteniendo los siguientes valores:

$$\begin{bmatrix} x_r \\ y_r \\ z_0 \end{bmatrix} = \begin{bmatrix} 480.339 \\ 127.627 \\ 1 \end{bmatrix} \quad (3.21)$$

$$\begin{bmatrix} x_l \\ y_l \\ z_0 \end{bmatrix} = \begin{bmatrix} 858.589 \\ 129.064 \\ 1 \end{bmatrix} \quad (3.22)$$

Mapeando los puntos obtenidos (3.21) y (3.22) en las imágenes derecha e izquierda, respectivamente, se puede observar en la Figura 3.24, que las coordenadas de los puntos calculados se aproximan mucho a las coordenadas de imagen del punto observado, por lo



Figura 3.22: Desplazamiento en las imágenes tomadas con el sistema (imágenes izquierda y derecha respectivamente), después de modificar los centros de la cámaras.

que se puede concluir que los cambios hechos al modelo, funcionan correctamente y este no sufre alteraciones que puedan derivar en errores.

Por otro lado, como se mencionaba anteriormente, uno de los objetivos principales de esta modificación al modelo de las cámaras, es apreciar con mayor claridad el efecto tridimensional de los objetos en una escena, por lo que en la Figura 3.25, se muestra una comparativa entre el anaglifo de la Figura 3.17 generado con la calibración normal, y un nuevo anaglifo generado con los parámetros en (3.17) y (3.18), y $\alpha = 38$ pixeles.

En la Figura 3.25, es fácilmente apreciable como el efecto tridimensional de los objetos es mucho más claro en el anaglifo que incluye el parámetro de α , en comparación con el anaglifo creado a partir de los parámetros convencionales. Otra ejemplo de esta mejora,



Imagen Derecha



Imagen Izquierda

Figura 3.23: Coordenadas tridimensionales del punto observado.

se muestra en la Figura 3.26, donde se lleva a cabo otra comparación entre el anaglifo de la Figura 3.19 generado con los parámetros del modelo sin modificar, y un anaglifo generado a partir del modelo modificado.

Si se compara el anaglifo de la Figura 3.26 con un anaglifo generado de forma sintética, como el que se muestra en la Figura 3.27, se pueden notar dos características en común: la primera, que para ambos anaglifos, existe una sección donde la disparidad claramente es cero. Para el anaglifo de la Figura 3.27, la parte de color amarillo, en la base del cañón, es claramente visible sin las gafas, por su parte, para el anaglifo de la Figura 3.26, el bote blanco es el punto donde la disparidad es cero, y donde dicho objeto es claramente visible sin las gafas. Como segunda característica, se puede observar que

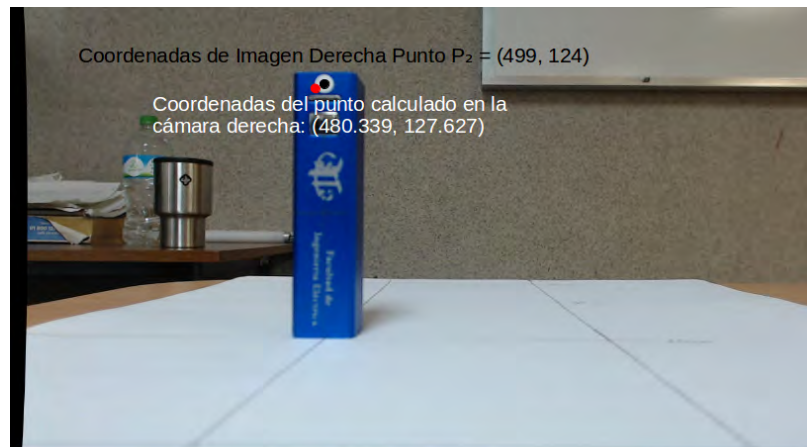


Imagen Derecha



Imagen Izquierda

Figura 3.24: Comparación entre los puntos calculados con los parámetros del sistema con el valor de $\alpha = 38$ pixeles, y las coordenadas de imagen el punto observado.

para ambos anaglifos, las componentes de color de los pixeles de los objetos al frente, se descomponen de una forma en particular, mientras que para los objetos en el fondo de la escena, los colores se cruzan, y se descomponen en un sentido contrario al de los objetos del frente.



a) Anaglifo sin desplazamiento de centros.

b) Anaglifo con desplazamiento de centros de $\alpha = 38$ píxeles.

Figura 3.25: Comparación entre el anaglifo de la Figura 3.17, con un anaglifo generado a partir del modelo con un desplazamiento de centros de $\alpha = 38$ píxeles, donde claramente se puede ver el corrimiento que existe entre las dos imágenes.

3.5.1. Relación disparidad-profundidad con α como nuevo parámetro del modelo de las cámaras

Además de mejorar la creación de anaglifos, modificar los centros de las cámaras va más allá de solo apreciar mejor el efecto de profundidad de los objetos en las imágenes. Con el nuevo parámetro α , la relación disparidad-profundidad también se ve afectada, ya que este nuevo cambio implica que la disparidad tomará el valor de cero, a una distancia en

específico, la cuál puede ser definido según la necesidad del problema. Este comportamiento se puede observar en la gráfica de la Figura 3.28, donde utilizando el modelo modificado expresado en (3.15) y (3.16), con los parámetros intrínsecos y extrínsecos en (3.17), y (3.18), y (3.4) y (3.6), se calcularon las coordenadas de imagen de un punto $P = [4.7, 0.0, Z]$, variando la profundidad Z desde 0 cm, hasta 180 cm con respecto del sistema de cámaras. Por consecuencia, en la gráfica se observa como la curva corta en cero el eje Z , cuando la disparidad $x_l - x_r$ es igual a cero.



a) Anaglifo sin desplazamiento de centros.

b) Anaglifo con desplazamiento de centros de $\alpha = 38$ pixeles.

Figura 3.26: Comparación entre un anaglifo sin desplazamiento de centros, y un anaglifo con un desplazamiento de $\alpha = 38$ pixeles.



Figura 3.27: Anaglifo de un cañón antiguo. [Fuentes14a]

Observando el cambio mostrado en la Figura 3.28, es posible concluir lo siguiente: los puntos con una disparidad igual a 0, son aquellos puntos que se encuentren a una distancia definida nombrada como Z' , y se tomará como referencia para el cálculo de otras profundidades. Ahora bien, dado que α vuelve cero la disparidad de un punto en específico, sumar α a los centros de las cámaras, es equivalente a calcular la disparidad como se muestra en (3.23).

$$D_x = (x_l - \alpha) - (x_r + \alpha) \quad (3.23)$$

Por lo tanto, con (3.23) es necesario revisar la relación “disparidad-profundidad” mostrada en (3.14), y reestructurar dicha ecuación como se muestra en (3.24).

$$(x_l - \alpha) - (x_r + \alpha) \begin{cases} = 0 & \text{Si } Z = Z' \\ > 0 & \text{Si } Z < Z' \\ < 0 & \text{Si } Z > Z' \end{cases} \quad (3.24)$$

Así pues, de acuerdo a (3.24), se puede determinar que si la disparidad de un punto

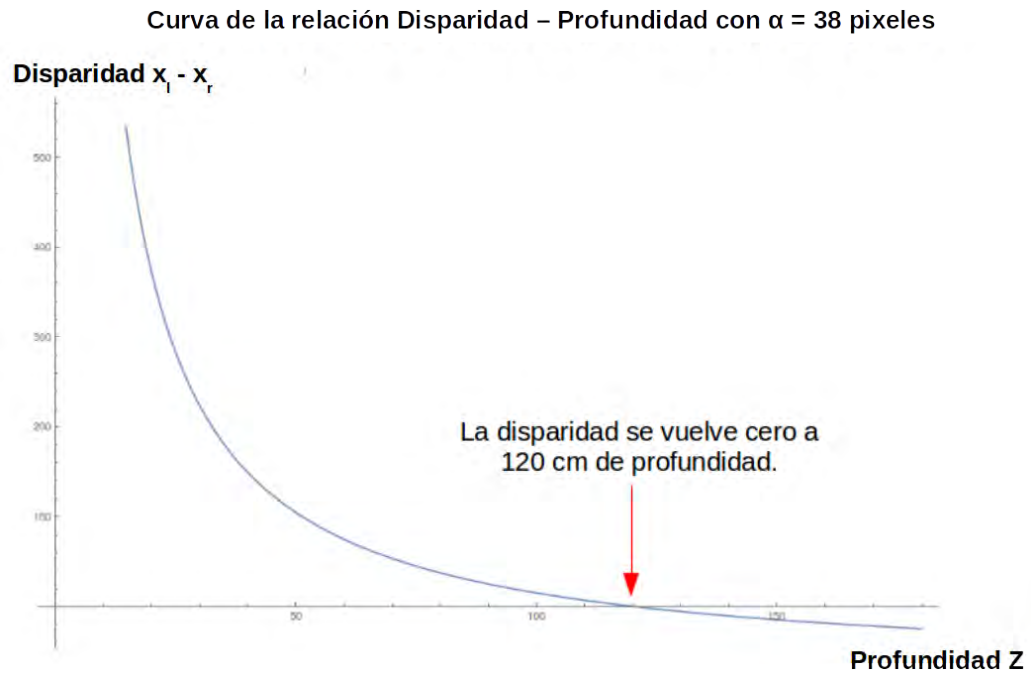


Figura 3.28: Esta gráfica muestra como la curva corta el eje de la profundidad, en $Z = 120$ centímetros, indicando que la disparidad en ese punto es $x_l - x_r = 0$. Esto sucede debido a que se añadió el valor de $\alpha = 38$ pixeles, al valor de los centros de las cámaras.

es cero, la profundidad de ese punto será Z' . Por otro lado, si la disparidad de un punto es mayor a cero, implica que la profundidad de dicho punto será menor a Z' ; de lo contrario, se infiere que su profundidad será mayor a Z' con una disparidad negativa.

Para demostrar el correcto funcionamiento del modelo, se llevó a cabo un experimento de medición de distancia, en la cual se tomaron diferentes capturas de un objeto a diferentes distancias. Se comenzó colocando el objeto a una distancia de 10 cm del sistema estereoscópico, para posteriormente ir incrementando la distancia del objeto en 10 cm con respecto de las cámaras hasta llegar a una distancia medida de 2.10 metros. En la Figura 3.29 se presenta el conjunto de imágenes tomadas para este experimento.

Del experimento se pudo determinar, que cualquier objeto que se encuentre a una distancia de 1.20 metros, tendrá una disparidad igual a cero, demostrando lo antes mencionado con (3.24). Como se puede apreciar en la Figura 3.31, la forma del objeto en la escena (una lata de aluminio) se ve bien definida, inclusive es visible sin los lentes anaglifo

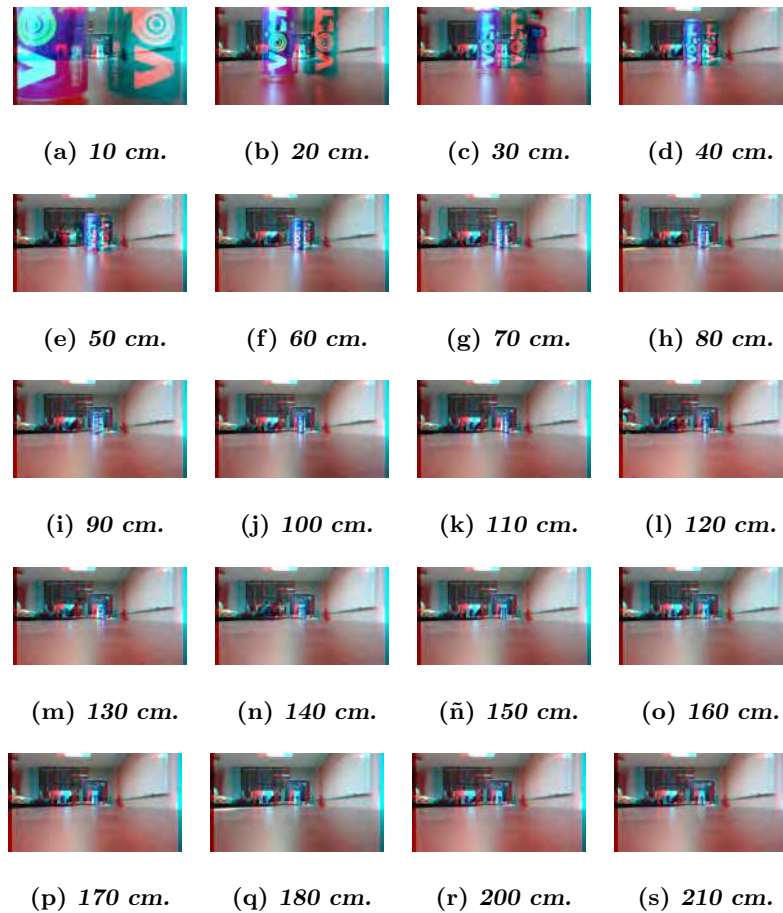


Figura 3.29: Imágenes tomadas de un objeto a diferentes distancias del sistema de cámaras.

a una distancia de 1.20 metros. Por otro lado, el mismo objeto a una distancia de 50 cm, se ve dividido en sus componentes de color. Si se empatan las imágenes de la Figura ??, con la gráfica de la Figura 3.28, se ven de forma más clara los resultados del experimento, como se muestra en la gráfica de la Figura 3.30.

Otro ejemplo de este comportamiento se observa cuando el objeto se encuentra a una distancia de 1.50 metros. Cabe mencionar que esta medida es dependiente de las características físicas del sistema, y de los resultados de la calibración, por lo que tratar de implementarlo con otro sistema, utilizando el mismo valor de $\alpha = 38$ píxeles, podría arrojar resultados diferentes. Sin embargo, el modelo desarrollado puede ser aplicado a otro sistema bajo los mismos criterios y obtener resultados similares. Para este trabajo de tesis

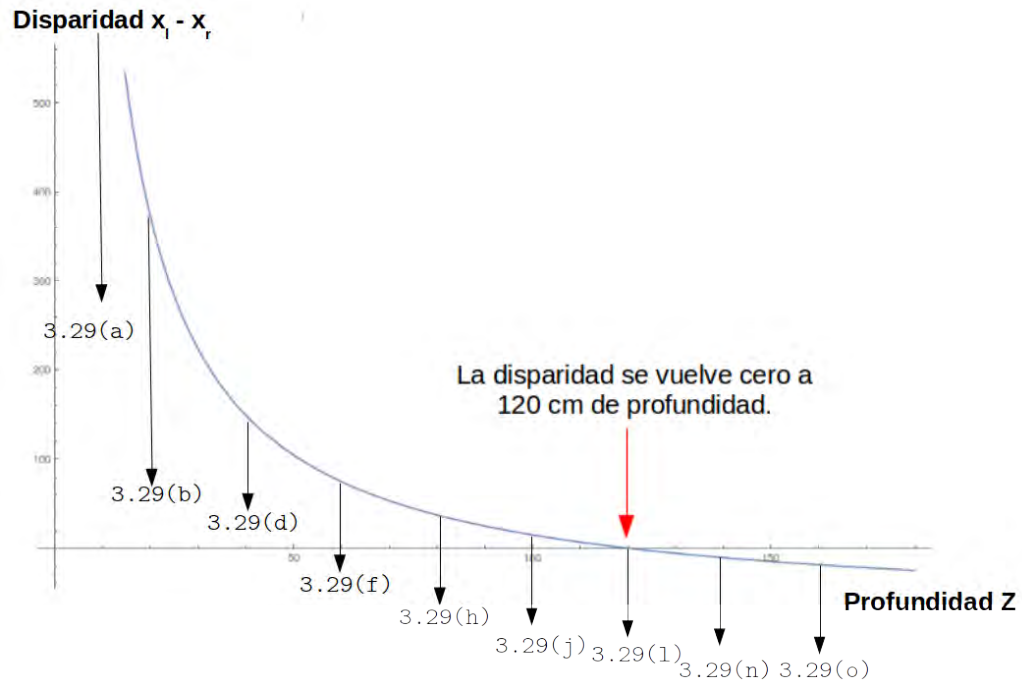


Figura 3.30: Posición de los objetos del experimento de la Figura 3.29, en relación a la gráfica de la Figura 3.28

se decidió utilizar el valor de 38 píxeles, ya que se concluyó que funcionaba mejor para visualizar las imágenes anaglifo generadas con el sistema de cámaras, de forma más clara.

3.6. Conclusiones del capítulo

En este capítulo se abordó el proceso de construcción de un sistema estéreo a partir de un par de cámaras web convencionales, demostrando que por medio de un patrón de tablero de ajedrez, y un algoritmo de calibración formado a partir diferentes funciones de la librería OpenCV, es posible corregir los problemas de alineación que dicho sistema presenta debido a sus características físicas. Además, como parte del proceso de calibración, se demostró que es posible obtener información específica del sistema de cámaras (parámetros intrínsecos y extrínsecos), con el objetivo de probar que el sistema funciona correctamente, y que es posible mapear un punto en coordenadas tridimensionales, a coordenadas de



Figura 3.31: En esta Figura se observa el acercamiento al anaglifo de la Figura 3.29(1), donde claramente se puede apreciar que el objeto a una distancia de 120 centímetros, se puede visualizar a la perfección sin ningún tipo de descomposición de color, indicando así, que la disparidad a esa distancia, es igual a cero.

imagen, con bastante precisión.

Por otro lado, se demostró que a partir de imágenes tomadas por el sistema estereo calibrado, es posible generar imágenes anaglifo y observar el impacto positivo de la calibración en el sistema, al ser visible el efecto de la rectificación, en las imágenes tomadas por el sistema de cámaras. Además, utilizando la técnica de imágenes anaglifo, fue posible demostrar de forma didáctica, cuál es la relación que existe entre la disparidad, y la profundidad de los objetos en una escena. De igual forma, se mejoró la calidad de los anaglifos al añadir una nueva variable α a los parámetros intrínsecos del sistema, lo que en términos simples representó la reducción de la distancia d , existente entre las cámaras, de forma artificial.

Finalmente, con la modificación al modelo del sistema estereo, fue posible establecer un marco de referencia claro y preciso, que tiene como finalidad, facilitar la estimación de la profundidad, generando un cambio en la relación convencional Disparidad-Profundidad, de la cuál se tomará ventaja, para la medición de la profundidad de diferentes objetos, en una escena capturada por el sistema estereo.

Capítulo 4

Estimación de la Profundidad Basado en Imágenes-Estéreo

A lo largo de esta tesis, se han desarrollado diferentes temas que son pieza fundamental para llevar a cabo la estimación de la profundidad a partir de imágenes estereoscópicas, tomadas con un sistema de cámaras estéreo. Haciendo remembranza, se desarrolló el análisis del modelo PinHole para una cámara convencional, y de igual forma para un sistema estéreo de cámaras. Por otro lado, se analizó el proceso de calibración de un par estéreo, y a partir de la información resultante de este proceso, fue posible generar imágenes anaglifo, y mejorar la calidad de estos. A su vez, gracias a las imágenes anaglifo fue posible establecer un marco de referencia con la nueva relación de disparidad – profundidad modificada, con la adición de la variable α a los parámetros intrínsecos de las cámaras. Se determinó que el valor de α adecuado para el sistema en cuestión, es de 38 pixeles, colocando un valor de disparidad igual a cero, para aquellos objetos a una distancia de 120 centímetros con respecto de las cámaras. Sin embargo, para poder estimar la profundidad de los objetos de una escena, los temas abordados son un preámbulo necesario, pero no concluyen con el proceso. Para poder continuar con el proceso de estimación de la profundidad, es necesario analizar dos temas que van muy de la mano: el problema de la búsqueda de correspondencia, y posteriormente, el cálculo de los mapas de disparidad. Por estas razones, el presente capítulo se enfoca en el análisis de diferentes formas de calcular la correspondencia, y finalmente, en

la obtención de mapas de disparidad, tomando un enfoque más orientado a la calidad de los resultados, y dejando un poco más de lado, el problema del tiempo.

4.1. Correspondencia y mapa de disparidad

Con los parámetros del modelo del sistema bien definidos, y las cámaras calibradas, es posible continuar con el proceso para la estimación de la profundidad. Para esto, es necesario abordar el tema de la obtención del mapa de disparidad, que aportará información preliminar de los objetos en la escena capturada, y previo a esto, analizar el concepto del cálculo de la correspondencia, como factor primordial para la obtención de mapas de disparidad.

De acuerdo con Patricia Compañ [Compañ04], la obtención del mapa de disparidad en un par de imágenes estereo, puede verse como un problema de búsqueda, ya que dado un pixel de la imagen derecha, se debe buscar un pixel correspondiente en la imagen izquierda como se muestra en la Figura 4.1, donde se observa un punto marcado de blanco en la imagen izquierda, señalado por una flecha, indicando que es el pixel correspondiente al pixel blanco marcado en la imagen derecha.

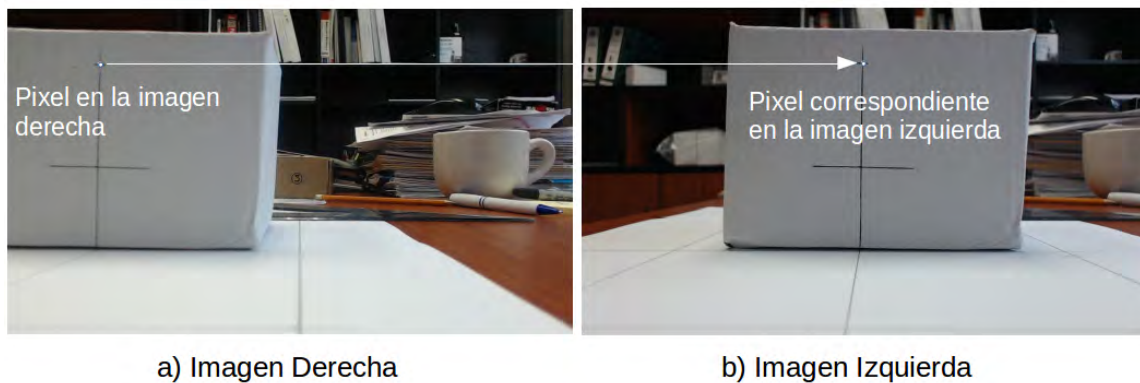


Figura 4.1: Ejemplo de la correspondencia de un pixel en la imagen derecha, y su empate con un pixel en la imagen izquierda.

Si embargo, a pesar de que esta tarea pudiera verse trivial o fácil de llevar a cabo, el problema de la búsqueda de la correspondencia se dificulta debido a factores que no se

pueden controlar, como el ruido en las imágenes (exceso de luminosidad), imágenes borrosas o poco definidas, discontinuidad en las imágenes capturadas, u oclusiones parciales. Por otro lado cabe mencionar, que la calibración pudiera resolver factores tales como discontinuidades en las imágenes, pero sin tener el mismo efecto con los demás problemas.

Por las razones mencionadas, para llevar a cabo la búsqueda de correspondencia entre un par de imágenes, es necesario tomar en cuenta dos decisiones importantes: la cantidad de píxeles con los que se llevará a cabo la búsqueda, y la medida de similitud que definirá el parecido entre los píxeles comparados. Con estas dos decisiones básicas, es posible englobar a los algoritmos de búsqueda de correspondencia en dos tipos: algoritmos basados en correlación, y algoritmos basados en puntos característicos (Lorenzo Tardón [García99]). Es necesario recalcar que los dos tipos de algoritmos se implementan de manera diferente, e inclusive pueden arrojar resultados diferentes entre mismos pares de imágenes.

4.1.1. Métodos Basados en Puntos Característicos

Este tipo de algoritmos se basan en la obtención de la correspondencia por medio de la extracción de características solo de ciertos píxeles en particular de la imagen, tomando un conjunto de medidas de similitud definidas. En general, este tipo de metodologías son iterativas, y procuran la minimización de algún criterio de error. A diferencia de los algoritmos basados en correlación, los algoritmos de puntos característicos no implementan el uso de ventanas de tamaño determinado, por el contrario, utilizan propiedades simbólicas y características numéricas en los píxeles. Algunos ejemplos de búsqueda de características para este tipo de metodologías son: búsqueda por aristas, búsqueda por esquinas, búsqueda de bordes, similitud entre vecinos, etc.

Según Lorenzo Tardón [García99], este tipo de metodologías son en general más robustas, menos sensibles a distorsiones geométricas y cromáticas, pero más difíciles de implementar, ya que precisan de la decisión sobre qué tipo de primitivas emplear en la correspondencia, qué medidas de similitud utilizar, y cómo relacionar cada correspondencia con la de los vecinos. Por otro lado, metodologías como estas, arrojan resultados más dispersos en el cálculo de la disparidad, es decir, la información en el mapa de disparidad, es menor, ya que como se mencionaba anteriormente, utilizan una cantidad definida de píxeles.

Por mencionar algunos ejemplos, se enlistan los más comunes.

- **Detector de esquinas de Harris.** El detector de esquinas de Harris, propuesto por Chris Harris *et al.* [Harris88], es un algoritmo basado en la teoría de Moravec. El principio básico del algoritmo de Moravec, es tomar una zona pequeña y poner como objetivo el pixel central. Después, mover dicha zona seleccionada hacia arriba y hacia abajo a lo largo de 4 direcciones, revisando los niveles de gris en cada dirección. El valor más pequeño en las 4 direcciones revisadas se evalúa como una posible esquina y si el valor es mayor que un umbral previamente determinado, entonces se considera como pixel esquina (Baigen He Zhu Ming *et al.* [Baigen He11]).
- **Detector de características FAST(Features from Accelerated Segment Test).** El detector de características FAST, basa la búsqueda de correspondencia, en un segmento de máscara circular, donde el radio de la máscara es uno de los parámetros que afectan la velocidad y la calidad de la detección de puntos correspondientes. Cada pixel dentro del radio de la máscara, es clasificado de a cuerdo a diferentes estados, por lo que si existe un segmento que contiene doce pixeles contiguos similares en brillo u opacidad, determina que el pixel evaluado es un pixel esquina, o no-esquina en caso contrario (Yenewondim Biadgie *et al.* [Biadgie14]).
- **Detector de características SURF(Speeded Up Robust Feature).** Este algoritmo define una locación y una escala para cada punto característico detectado. Este factor de escala define una ventana de tamaño determinado alrededor del punto característico, de tal forma que el vecindario definido por esta ventana podría, tener la misma información visual sin importar la escala del objeto a la que dicha característica corresponde. Además, la información visual incluida en este vecindario podría ser usada para caracterizar aún más al punto característico y hacerlo distinguible de los demás (Yenewondim Biadgie *et al.* [Biadgie14], y Robert Laganier [Laganier11]).
- **Detector de características SIFT(Scale Invariant Feature Transform), aplicado a la búsqueda de correspondencia.** Suponiendo que se tiene una imagen derecha, y una imagen izquierda tomadas con el mismo sistema estéreo, para el cálculo

lo de la correspondencia en este par de imágenes, con el algoritmo SIFT se extraen primeramente los puntos característicos en la imagen izquierda, y se almacena en una base de datos. Posteriormente se toma la imagen derecha, y se obtienen los puntos característicos como se hizo con la imagen anterior. Individualmente, los puntos de la segunda imagen son comparados con los candidatos almacenados en la base de datos obtenidos a partir de la imagen izquierda, basándose en el cálculo de la distancia Euclidiana, como medida para determinar la similitud entre los puntos elegidos de ambas imágenes (David G. Lowe [Lowe04]).

Cabe mencionar que la librería OpenCV también incluye estos algoritmos en sus acervo de funciones desarrolladas. Para más información, consultar <http://docs.opencv.org/>.

4.1.2. Métodos de Búsqueda Basados en Correlación

Los métodos de búsqueda de correspondencia basados en correlación, son algoritmos más fáciles de implementar y obtienen información de todos los píxeles en la imagen derecha con respecto de la izquierda. Por consecuencia, calculan mapas de disparidad considerados densos, los cuales se caracterizan por calcular una disparidad para cada píxel en la imagen. Por otro lado, este tipo de propiedades son muy útiles a la hora de reconstruir superficies, ya que la búsqueda de correspondencia se vuelve un verdadero reto, cuando existen cambios de luminosidad entre los pares de imágenes (Patricia Compañ [Compañ04]).

El funcionamiento básico de los algoritmos basados en la correlación, consiste en tomar una ventana de tamaño fijo de píxeles vecinos a un píxel en particular perteneciente a la imagen I_r (imagen derecha), para posteriormente compararlo con ventanas de píxeles del mismo tamaño fijo en la imagen I_l (imagen izquierda), como se muestra en la Figura 4.2. Con esta comparación entre ventanas, la correlación producirá un valor de similitud, el cual puede ser calculado utilizando diferentes medidas de distancia, y tiene como objetivo obtener un valor máximo o mínimo para indicar donde existe un mayor parecido entre píxeles.

De entre las medidas para el cálculo de la distancia más comunes para la obtención de la medida de similitud en la búsqueda de correspondencia, se encuentran: la Suma de

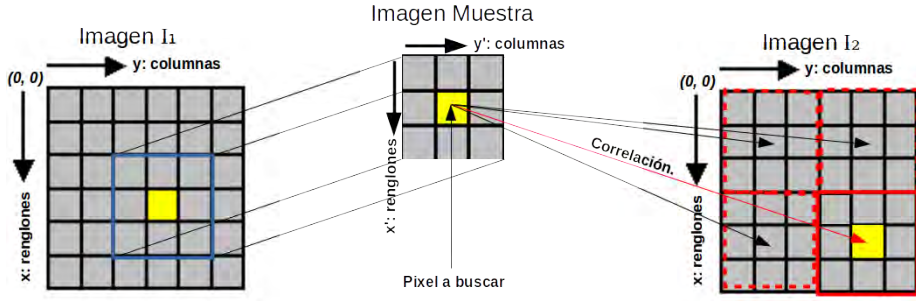


Figura 4.2: Funcionamiento básico de las búsquedas basadas en correlación.

Cuadrados de las Diferencias **SSD** (4.1), la Suma de Cuadrados de las Diferencias Normalizada **NSSD** (4.2), la Correlación Cruzada **CORR** (4.3), y la Correlación Cruzada Normalizada **CORRN** (4.4), donde V es una ventana cuadrada de píxeles a comparar, I es la imagen donde se hará la búsqueda, (\hat{x}, \hat{y}) son coordenadas de los píxeles en la ventana V , y (x, y) las coordenadas de los píxeles en la imagen I . Cabe mencionar que la ventana V debe ser de un tamaño impar $[(2W + 1) \times (2W + 1)]$, con el objetivo de que el píxel del cuál se pretende obtener una similitud, quede situado en el centro de la ventana, donde W es el valor que define el tamaño de los lados de V , y puede ser libremente escogido según las necesidades del proyecto.

$$RSSD(x, y) = \sum_{\hat{x}, \hat{y}} [V(\hat{x}, \hat{y}) - I(x + \hat{x}, y + \hat{y})]^2 \quad (4.1)$$

$$RNSSD(x, y) = \frac{\sum_{\hat{x}, \hat{y}} [V(\hat{x}, \hat{y}) - I(x + \hat{x}, y + \hat{y})]^2}{\sqrt{\sum_{\hat{x}, \hat{y}} V(\hat{x}, \hat{y})^2 \cdot \sum_{\hat{x}, \hat{y}} I(x + \hat{x}, y + \hat{y})^2}} \quad (4.2)$$

$$RCORR(x, y) = \sum_{\hat{x}, \hat{y}} (\hat{T}(\hat{x}, \hat{y}) \cdot I(x + \hat{x}, y + \hat{y})) \quad (4.3)$$

$$RNCORR(x, y) = \frac{\sum_{\hat{x}, \hat{y}} (\hat{T}(\hat{x}, \hat{y}) \cdot I(x + \hat{x}, y + \hat{y}))}{\sqrt{\sum_{\hat{x}, \hat{y}} \hat{T}(\hat{x}, \hat{y})^2 \cdot \sum_{\hat{x}, \hat{y}} I(x + \hat{x}, y + \hat{y})^2}} \quad (4.4)$$

Autores como Vibin N. Valsan *et al.* [Valsan15], Abiel Aguilar *et al.* [Abiel Aguilar González15],

y Beau J. Tippetts *et al.* [Tippetts11], fundamentan sus trabajos en algunas de las metodologías enlistadas, haciendo mayor uso de la descrita con (4.1).

Para la búsqueda de correspondencia basada en la correlación, Patricia Compañ [Compañ04] propone el algoritmo 3 llamado CorrespondenciaEnBaseACorrelacion, donde a partir de dos imágenes I_r e I_l , se obtiene la correspondencia entre pixeles, variando únicamente la medida de distancia MD , evaluando los valores de similitud obtenidos como máximo o un mínimo para definir el grado de parecido entre los pixeles comparados en la matriz \hat{C} , y retornando una matriz CR , con las correspondencias entre los pixeles de la imagen derecha e izquierda.

Algoritmo 3: CorrespondenciaEnBaseACorrelacion(I_r, I_l, W, MD).

- 1: **para todo** pixel $I_r(x, y) \in I_r$. **hacer**
 - 2: Crear V con un tamaño de $(2W + 1) \times (2W + 1)$.
 - 3: **para todo** pixel $V(\hat{x}, \hat{y}) \in V$. **hacer**
 - 4: Calcular $\hat{C}(x, y) =$ a partir de la medida de distancia seleccionada MD
 - 5: **fin para**
 - 6: Buscar el valor de similitud máximo o mínimo en \hat{C} , según MD .
 - 7: Guardar la correspondencia $CR(x, y) = (x_l, y_l)$ del pixel seleccionado.
 - 8: **fin para**
 - 9: Retornar CR .
-

Es importante mencionar que las medidas de distancia expresadas en las ecuaciones (4.1), (4.2), (4.3), y (4.4) son implementados por la función `matchTemplate`, la cuál será descrita con mayor detalle a lo largo de este capítulo.

4.1.3. Obtención de la correspondencia utilizando métodos de búsqueda basados en correlación

Como se mencionaba anteriormente, los algoritmos de búsqueda de correspondencia basados en puntos característicos, utilizan una cantidad determinada de pixeles en las imágenes para obtener la correspondencia y posteriormente calcular el mapa de disparidad. De acuerdo con Patricia Compañ [Compañ04], este tipo de métodos son adecuados cuando

existe información previa sobre las escenas capturadas, y los escenarios más idóneos para estos casos suelen ser aquellos con muchas líneas rectas y pocas texturas. Bajo estas premisas, este tipo de metodologías suelen ser más rápidas en la obtención de mapas de disparidad en comparación con otros métodos, ya que no manejan grandes cantidades de datos, lo cual resulta en mapas de disparidad poco densos, que suelen contener la información suficiente para aplicaciones como la navegación visual. Además, este tipo de algoritmos son menos propensos a pares falsos por cambios de luminosidad, y oclusiones, ya que reducen estos problemas con un conjunto de múltiples restricciones. Por otro lado, la implementación individual de estos métodos suele ser muy diferente, utilizan una cantidad considerable de parámetros que además cambian drásticamente de método a método, y los resultados pueden variar bastante de una metodología a otra, con diferentes pares de imágenes (Federico Lucumberry [Lucumberry14]).

En contraste, los algoritmos de búsqueda de correspondencia basados en correlación, suelen ser algoritmos más simples de implementar, y la cantidad de parámetros utilizados es más limitada, los resultados son mapas de disparidad más densos, y a pesar de que el tiempo implementado en el cálculo de la correspondencia puede ser mayor en comparación con otros métodos, este tipo de metodologías pueden resultar en mapas de disparidad con buenos resultados por la cantidad de información que se tiene sobre las imágenes. No es necesario trabajar con escenarios tan controlados, tomando en consideración que se debe evitar la iluminación excesiva, y además, si a esto se añaden algunos factores como el uso de la geometría epipolar, el uso de los tres canales de color en lugar de una escala de grises, el uso del modelo de color HSV, y el hecho de que las cámaras no estarán tan propensas a movimiento, se pueden obtener resultados considerablemente buenos, además de reducir el tiempo de búsqueda de correspondencias.

Cabe mencionar que para llevar a cabo este proyecto, se optó por utilizar algoritmos basados en correlación, ya que son simples de implementar, y los mapas de disparidad resultantes son densos, aprovechando de esta forma la información de todos los píxeles en las imágenes. Por otro lado, a pesar de que el tiempo de procesamiento puede ser elevado, el objetivo principal es estimar la profundidad de la forma más acertada posible, dejando en segundo plano el tiempo de procesamiento.

Función `matchTemplate`

OpenCV ofrece una herramienta para realizar la búsqueda de correspondencia en un par de imágenes, con la función `matchTemplate()`. Esta función lleva a cabo la búsqueda de correspondencia, implementando cualquiera de las metodologías expresadas en (4.1), (4.2), (4.3), y (4.4), añadiendo además, dos nuevas medidas de distancia, la Correlación de Coeficientes **CCOEFF**, y la Correlación de Coeficientes Normalizada **NCCOEFF**. La función `matchTemplate` tiene como objetivo encontrar el área con mayor similitud entre imágenes utilizando los parámetros que se enlistan a continuación:

- Imagen I_l . Imagen destino donde se llevará a cabo la búsqueda de la correspondencia.
- Ventana V . Identificada por la función como la ventana muestra, o ventana parche. Esta ventana pertenece a la imagen fuente (en este caso I_r), y es la sección que se pretenden buscar en la imagen destino (en este caso la imagen I_l).
- Matriz \hat{C} . Identificada como la matriz de comparación de resultados, almacena los valores de similitud entre las comparaciones de V e I_l .
- *EqMetodo*. Método de similitud a implementar. Con este parámetro se indica el método con el cual se calcularán los valores de similitud.

Posteriormente, con la función `minMaxLoc`, se buscan los valores máximo y mínimo en la matriz resultante \hat{C} , retornando finalmente las coordenadas de (x_l, y_l) de los pixeles con dichos valores pertenecientes a la imagen I_l . Cabe mencionar que para las medidas de distancia expresadas en (4.1) y (4.2), se toma como medida de similitud al pixel con mínimo valor en \hat{c} , mientras que para el resto, se toma el pixel con el valor máximo.

A partir de la función `matchTemplate` y `minMaxLoc`, se propone el algoritmo 4, tomando como base el algoritmo 3 ya explicado.

Con las coordenadas (x_l, y_l) de los pixeles en la imagen I_l , correspondientes para cada pixel (x_r, y_r) en la imagen derecha, es entonces posible continuar con el cálculo de los mapas de disparidad para el par de imágenes en cuestión. Para más información sobre la función `matchTemplate` y `minMaxLoc`, refiérase a la documentación oficial de OpenCV.

Algoritmo 4: CorrespondenciaConMatchTemplate(I_r, I_l, W, MD)

- 1: **para todo** pixel $I_r(x_r, y_r) \in I_r$. **hacer**
 - 2: **Crear** V con un tamaño de $(2W + 1) \times (2W + 1)$
 - 3: **Calcular** $\hat{C}(x, y)$ con `matchTemplate($I_l, V, \hat{C}(x, y), MD$)`
 - 4: **Guardar** la correspondencia $CR(x, y) = \text{minMaxLoc}(\hat{C}(x, y))$
 - 5: **fin para**
 - 6: **Retornar** CR
-

4.1.4. Cálculo de los mapas de disparidad

Utilizando las funciones que OpenCV proporciona, y conociendo diferentes alternativas para el cálculo de la correspondencia, es posible obtener los mapas de disparidad de un par de imágenes.

El cálculo de la disparidad es un procedimiento relativamente sencillo. Una vez que se tienen las coordenadas (x_l, y_l) de un pixel en la imagen I_l , correspondiente a un pixel (x_r, y_r) en la imagen I_r , lo único que se hace es calcular la diferencia entre las posiciones en x , y las posiciones en y , como se expresa en (2.35). Las diferencias calculadas para todos los pixeles en las imágenes, serán almacenadas en una matriz Dx para la disparidad en x , y en una matriz Dy para la disparidad en y . Dichas matrices, al ser presentadas como imágenes en escalas de grises, ofrecen una estimación visual sobre la posición de los objetos con respecto de las cámaras, al tomar el valor de disparidad calculado para cada pixel, y mostrarlo en una tonalidad de color en escala de grises.

Cabe mencionar que a pesar de la calibración y rectificación artificial de las cámaras del sistema, es posible que siga existiendo un pequeño desfase en las imágenes, por lo que será posible obtener un mapa de disparidad tanto para el eje x , como para el eje y , sin embargo, el mapa de disparidad que brindará mayor información, será el calculado para el eje x , ya que es ahí donde radica el mayor desplazamiento entre los pixeles de ambas imágenes.

A partir del algoritmo 4, se implementó el algoritmo 5 como un algoritmo nuevo para el cálculo de la disparidad, donde Dx y Dy son las matrices que almacenan la información de los mapas de disparidad en el eje x y y respectivamente.

Algoritmo 5: CalcularDisparidad(I_r, I_l, W, MD).

- 1: **para todo** pixel $I_r(x_r, y_r) \in I_r$. **hacer**
 - 2: **Crear** V con un tamaño de $(2W + 1) \times (2W + 1)$.
 - 3: **Calcular** $\hat{C}(x, y)$ con `matchTemplate($I_l, V, \hat{C}(x, y), MD$)`.
 - 4: **Obtener** (x_l, y_l) con `minMaxLoc($\hat{C}(x, y)$)`.
 - 5: **Calcular** la disparidad $Dx(x, y) = x_l - x_r$
 - 6: **Calcular** la disparidad $Dy(x, y) = y_l - y_r$
 - 7: **fin para**
 - 8: **Retornar** Dx y Dy .
-

4.1.5. Geometría Epipolar

La geometría epipolar es un modelo geométrico motivado por la búsqueda de la correspondencia entre dos imágenes, el cual consiste en alinear el plano de una imagen, con el plano de la otra, teniendo como base el eje que mantiene alineados a los centros de las cámaras como menciona Richard Hartley *et al.* [Hartley04]. Partiendo de este hecho, de acuerdo con Robert Laganierie [Laganierie11], la geometría epipolar explica que un pixel x en una de las imágenes representando a un punto X con coordenadas tridimensionales, puede ser localizado en la otra imagen al trazar una línea recta imaginaria que une a x con X . La línea recta recién trazada, es proyectada en la segunda imagen, reduciendo el espacio de búsqueda de correspondencia, a una sola línea recta de posibles candidatos. La proyección de esa línea imaginaria sobre la segunda imagen, se conoce como línea epipolar.

En la figura 4.3 se observa la representación de un punto X en el espacio tridimensional, por un punto x en la imagen izquierda. Si se traza una línea imaginaria entre x y X , es posible reflejar dicha línea en la imagen derecha, convirtiéndose en la línea epipolar del punto x . A partir de ese momento, el espacio de búsqueda de correspondencia para el punto x , será reducido a la línea epipolar en la imagen derecha.

Con este modelo se genera una restricción fundamental que debe satisfacer la correspondencia entre dos pixeles, es decir, el empate de un pixel se debe encontrar sobre su línea epipolar recién generada. Por otro lado, la línea epipolar es también representada

por un vector de tamaño 3. Este vector corresponde a un conjunto de puntos de la imagen derecha, que cruza a todos los pixeles que funcionan como posibles candidatos, por lo que x puede tomar un valor entre 0 y n_c , siendo n_c el número de columnas sobre la línea epipolar, representada por (4.5), donde a , b , y c , son los coeficientes de la ecuación de línea epipolar, como se muestra en (4.6), siendo estos calculados a partir de la información que la matriz F o matriz Fundamental, resultante del proceso de calibración del sistema estéreo.

$$ax + by + c = 0 \quad (4.5)$$

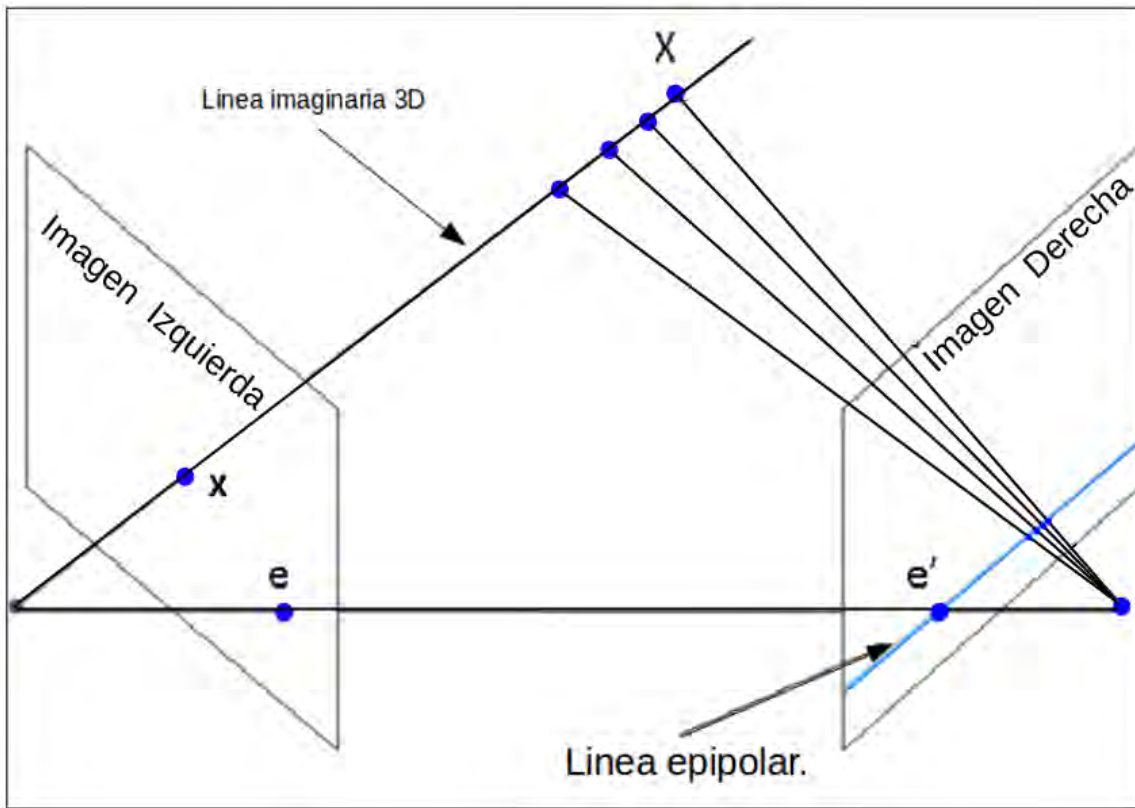


Figura 4.3: Modelo de la geometría epipolar. Robert Laganiere [Laganiere11]

$$\begin{bmatrix} a \\ b \\ c \end{bmatrix} = F \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (4.6)$$

Como se mencionó anteriormente, la matriz F se obtiene como resultado de la calibración del sistema de cámaras, y presenta las siguientes ventajas:

- **Rectificación del sistema estéreo.** Con el cálculo de las líneas epipolares, se puede determinar si el sistema estéreo ha sido bien calibrado, ya que si se grafican las líneas epipolares en ambas imágenes, claramente se puede observar si la correspondencia entre píxeles es adecuada como se aprecia en la Figura 4.4, donde las líneas epipolares se presentan en color verde, como se muestra en la Figura 4.4a), y se hace un acercamiento a un par de puntos esquina de cada imagen como se observa en la Figura 4.4b), mostrando que las líneas epipolares coinciden correctamente.



a) Imágenes Rectificadas con Lineas Epipolares.



b) Acercamiento a los puntos esquina.

Figura 4.4: En esta figura, se aprecia como las líneas epipolares pasan por los mismos puntos.

- **Reducción en el tiempo de búsqueda.** A pesar de que el factor tiempo no es uno de los objetivos que se busca mejorar a la hora de estimar la profundidad, es importante resaltar que este factor puede llegar a ser un problema debido al número tan elevado de operaciones que se pueden llegar a requerir, en relación al tamaño de las imágenes que se están manejando. Por ejemplo, si se tuviera un par de imágenes estereo I_r e I_l tomadas con un sistema de cámaras estereo, y se buscara la correspondencia de los pixeles de la imagen I_r en la imagen I_l , se tendría que comparar a todos y cada uno de los pixeles de la imagen I_r , con todos y cada uno de los pixeles de la imagen I_l para poder calcular la disparidad. El problema de este enfoque, es que aunque las imágenes fueran muy pequeñas, el tiempo de procesamiento seguiría siendo elevado. Por ejemplo, si las imágenes I_r e I_l fueran de tamaño $N \times M$ donde N representa las columnas de una imagen y es de un tamaño de 256, y M representando a las filas es de igual forma de 256 pixeles, la búsqueda de la correspondencia consistiría en llevar a cabo una cantidad de 4, 294, 967, 296 operaciones, o en otras palabras, N^4 operaciones. Con la implementación de la geometría epipolar, es posible reducir la cantidad de operaciones N^3 , ya que de hacer una búsqueda de todos contra todos, se reduce a todos los pixeles de una imagen contra los pixeles en la línea epipolar correspondiente a cada uno de estos puntos.

Como parte del resultado de la calibración, el Algoritmo 1 StereoCalib, implementado en OpenCV también retorna un archivo con los valores de la matriz fundamental.

4.1.6. Obtención del Mapa de Disparidad, Implementando Geometría Epipolar

Si se toman los valores de la matriz fundamental mediante (4.6), y se utilizan para el cálculo de los coeficientes de la ecuación de una línea recta (2.7) despejando a y de (4.7), se propone el Algoritmo 6 llamado DisparidadEpipolar, para el cálculo de la disparidad implementando líneas epipolares.

$$y = \frac{-c - (a * x)}{b} \quad (4.7)$$

Algoritmo 6: DisparidadEpipolar($I_r, I_l, W, EqMetodo, F$).

- 1: **para todo** pixel $I_r(x_r, y_r) \in I_r$ **hacer**
 - 2: **Crear** V con un tamaño de $(2W + 1) \times (2W + 1)$
 - 3: **Calcular** $a = x_r \times F(0, 0) + y_r \times F(0, 1) + F(0, 2)$
 - 4: **Calcular** $b = x_r \times F(1, 0) + y_r \times F(1, 1) + F(1, 2)$
 - 5: **Calcular** $c = x_r \times F(2, 0) + y_r \times F(2, 1) + F(2, 2)$
 - 6: **Calcular** la línea epipolar = $\frac{-c - a * x}{b}$
 - 7: **Obtener** $\hat{C}(x, y)$ con `matchTemplate(epipolar, V, $\hat{C}(x, y)$, EqMetodo)`
 - 8: **Obtener** (x_l, y_l) con `minMaxLoc($\hat{C}(x, y)$)`
 - 9: **Calcular** la disparidad $D_x(x, y) = x_l - x_r$
 - 10: **Calcular** la disparidad $D_y(x, y) = y_l - y_r$
 - 11: **fin para**
 - 12: **Retornar** $D_x y D_y$
-

4.2. Estimación de la profundidad.

Como se ha hecho mención en diferentes ocasiones anteriormente, el objetivo principal de este trabajo de tesis, es poder estimar la profundidad de los objetos en una escena con respecto de las cámaras, para lo cual, se ha desarrollado desde el tema de construcción y calibración del sistema estéreo, hasta el cálculo de los mapas de disparidad. Sin embargo, para poder continuar con la estimación de la profundidad, es necesario hacer referencia al Capítulo 3, donde se explica como a partir de la información obtenida para el modelo estéreo con la calibración, y de la implementación de las imágenes anaglifo, es posible establecer un marco de referencia modificando el modelo convencional de la relación Disparidad-Profundidad, para establecer una relación de proporcionalidad nueva, expresada en (3.24).

Tomando a (3.24), desarrollada en el Capítulo 3, y siendo el valor de α igual a 38 pixeles el desplazamiento aplicado a los centros de cámara como se muestra en (3.17) y (3.18), es posible determinar que para todos los objetos con disparidad igual a cero, la distancia con respecto de las cámaras será de 120 centímetros como se expresa en (4.8).

$$(x_l - 38) - (x_r + 38) \begin{cases} = 0 & \text{Si } Z = 120.0 \text{ cm.} \\ > 0 & \text{Si } Z < 120.0 \text{ cm.} \\ < 0 & \text{Si } Z > 120.0 \text{ cm.} \end{cases} \quad (4.8)$$

Siguiendo entonces a (4.8), es posible comenzar con la estimación de la profundidad, tomando como base los mapas de disparidad, previamente calculados.

Para poder diferenciar la profundidad de cada objeto en una escena capturada, se decidió implementar un escala de color similar a la utilizada en imágenes térmicas. Con dicha escala térmica de colores, se pinta de un color diferente cada pixel en una imagen, dependiendo de su valor de disparidad en la matriz Dx , tomando como marco de referencia lo expresado en (4.8), donde para todo pixel con un valor de disparidad igual a cero, o en un rango de disparidad mayor a -20 pixeles y menor a 20 pixeles, se asignó una tonalidad de color rojo en formato RGB de $(255, 0, 0)$, indicando que todos los pixeles en color rojo, pertenecen a objetos que se encuentran aproximadamente a una distancia de $Z = 120.0$ centímetros, con respecto de las cámaras.

Posteriormente, para los pixeles con valores de disparidad positivos, y mayores a 20 pixeles, se fueron añadiendo al formato de color $(255, 0, 0)$, valores al canal verde de 10 en 10 unidades, procurando aplicar un color diferente, para rangos de disparidad de 40 pixeles aproximadamente, por cada 5 centímetros de profundidad, llegando a una tonalidad de amarillo, RGB igual $(255, 170, 0)$ para los pixeles pertenecientes a objetos con una distancia prácticamente de cero, con respecto de las cámaras. Nuevamente, respetando una proporción de aproximadamente 40 pixeles por cada 5 centímetros de profundidad, se decrementó el valor del canal rojo de 10 en 10 unidades hasta llegar a cero, indicando así, que para los pixeles con un valor de disparidad negativa, la tonalidad de color se irá tornando oscura, si pertenecen a objetos con una distancia muy lejana, con respecto de las cámaras. En la Figura 4.5 se muestra un ejemplo de la escala mencionada, y en la Tabla 4.1, se muestra una relación de color-profundidad.

Para llevar a cabo la estimación de la profundidad, se propone el Algoritmo 7, calcularProfundidad, el cual recibe como único parámetro el mapa de disparidad Dx previamente calculado, con el objetivo de aplicar un color a cada pixel dependiendo su valor



Figura 4.5: En esta figura se muestra la escala de color que se tomo como base para diferenciar las diferentes profundidades.

Tabla 4.1: Tabla de relación entre la escala de color, y la distancia.

| Rango de disparidad (pxs.) | Tonalidad (RGB) | Distancia (cms.) |
|----------------------------|-----------------|------------------|
| De -1020 a -981 pixeles | (5, 0, 0) | 245 cm. |
| De -980 a -941 pixeles | (15, 0, 0) | 240 cm. |
| . | . | . |
| . | . | . |
| . | . | . |
| De -100 a -61 pixeles | (235, 0, 0) | 130 cm. |
| De -60 a -21 pixeles | (245, 0, 0) | 125 cm. |
| De -20 a 20 pixeles | (255, 0, 0) | 120 cm. |
| De 21 a 60 pixeles | (255, 10, 0) | 115 cm. |
| De 61 a 100 pixeles | (255, 20, 0) | 110 cm. |
| . | . | . |
| . | . | . |
| . | . | . |
| De 941 a 980 pixeles | (255, 240, 0) | 5 cm. |
| De 981 a 1020 pixeles | (255, 250, 0) | 0 cm. |

de disparidad, de acuerdo a la Tabla 4.1. Dado que el mayor desplazamiento se dá en el eje de las x , el mapa de disparidad Dy , es descartado en este punto.

Siguiendo el Algoritmo 7, se lleva a cabo la estimación de la profundidad, donde Dx representa el mapa de disparidad en x , y $profD$ representa la matriz que almacenará las tonalidades de color para cada coordenada (x, y) en la matriz de disparidad. Por lo tanto, si se cumple que para $Dx(x, y)$ el valor de disparidad es igual a cero, entonces $profD$ en la posición (x, y) y en el canal de color 0, tomará un valor de 255, mientras que para el canal 1 y 2, tomará los valores de 0. Por otro lado, si el valor de disparidad en $Dx(x, y)$ es menor a 0, $profD$ en la posición (x, y) tomará un valor para el canal rojo, equivalente a su valor de disparidad, siguiendo la Tabla 4.1. En caso contrario, si el valor de disparidad en $Dx(x, y)$ es positivo, el canal 1 equivalente relacionado con al color verde, tomará una tonalidad de

Algoritmo 7: calcularProfundidad(Dx), Recibe el mapa de Disparidad Dx .

```

1: Crear matriz resultante  $profD$ 
2: para todo pixel  $(x, y) \in Dx$ . hacer
3:   si  $Dx(x, y) == 0.0$  entonces
4:      $profD(0, x, y) = 255$ 
5:      $profD(1, x, y) = 0$ 
6:      $profD(2, x, y) = 0$ 
7:   si no, si  $Dx(x, y) < 0.0$  entonces
8:     si  $Dx(x, y) == -60$  entonces
9:        $profD(0, x, y) = 245$ 
10:      ...
11:     si no, si  $Dx(x, y) == -1020$  entonces
12:        $profD(0, x, y) = 5$ 
13:     fin si
14:   si no, si  $Dx(x, y) > 0.0$  entonces
15:     si  $Dx(x, y) == 60$  entonces
16:        $profD(0, x, y) = 255$ 
17:        $profD(1, x, y) = 10$ 
18:       ...
19:     si no, si  $Dx(x, y) == 1020$  entonces
20:        $profD(0, x, y) = 255$ 
21:        $profD(1, x, y) = 250$ 
22:     fin si
23:   fin si
24: fin para
25: Retornar  $profD$ 

```

acuerdo a su valor de disparidad, siguiendo la Tabla 4.1 en $profD(x, y)$. Finalmente, se retorna $profD$ con los pixeles coloreados.

4.3. Conclusiones del Capítulo

Para la estimación de la profundidad, el cálculo y obtención de los mapas de disparidad es una tarea de suma importancia, por lo que elegir un buen método para la obtención de la disparidad, siempre será primordial. A lo largo de este capítulo, se mostró el desarrollo de diferentes algoritmos de búsqueda de correspondencia basados en correlación. Sin embargo, utilizando este tipo de búsquedas, puede representar un tiempo de procesamiento bastante prolongado, debido a la cantidad de operaciones necesarias para llevar a cabo la obtención de un mapa de disparidad a partir de un par de imágenes de tamaño 1280×720 . Por lo tanto, a pesar de que el objetivo de este trabajo de tesis, no está orientado a la reducción del tiempo a la hora de calcular los mapas de disparidad, se propuso un algoritmo basado en líneas epipolares, para reducir el tiempo de ejecución que este tipo de procedimientos conllevan. Por otro lado, la estimación de la profundidad utilizando los mapas de disparidad calculados con metodologías basadas en correlación, podrían resultar eficientes debido a la densidad de información que este tipo de mapas de disparidad ofrecen.

Capítulo 5

Resultados

Finalmente, es posible decir que el proceso de la estimación de la profundidad a partir de imágenes estereoscópicas, está completo. Después de haber analizado y modificado el modelo de relación Disparidad-Profundidad en el Capítulo 3, y después de haber analizado diferentes formas de llevar a cabo la búsqueda de la correspondencia para la obtención de los mapas de disparidad en el Capítulo 4, fue posible desarrollar el Algoritmo 7, con el objetivo de calcular la profundidad de los objetos en una escena, diferenciando las distancias de los objetos por medio de una escala de color, similar a la escala de colores utilizada en imágenes térmicas. Por estas razones, a lo largo de este capítulo, se presentan los diferentes resultados obtenidos de diferentes formas de calcular los mapas de disparidad. De igual forma, se presentan los resultados en la mejora de la calidad de las imágenes anaglifo, y los resultados de la estimación de la profundidad utilizando el Algoritmo 7, a partir de imágenes estereoscópicas tomadas con el sistema estéreo, creado para esta tesis.

5.1. Resultado para la generación de Imágenes Anaglifo.

Con el modelo modificado del sistema de cámaras estéreo, fue posible generar imágenes anaglifo con un efecto tridimensional bastante bueno. En las Figuras 5.1, 5.2, y 5.3, se observan una serie de anaglifos tomados con el sistema de cámaras utilizando el modelo modificado con el valor de $\alpha = 38$ pixeles, y donde claramente es visible la calidad de los anaglifos.



Figura 5.1: Anaglifo tomado con el sistema estéreo usando el modelo del sistema modificado.



Figura 5.2: Anaglifo al exterior del edificio del Omega 2 en la DEP-FIE, tomado con el sistema estéreo usando el modelo del sistema modificado.

5.2. Resultados y comparativas entre los diferentes algoritmos para el cálculo de los mapas de disparidad.

Para llevar a cabo las pruebas para el cálculo de los mapas de disparidad, se utilizaron un par de imágenes tomadas con el sistema estéreo de tamaño 1280×720 píxeles,



a) Anaglifo sin movimiento de centros.



b) Anaglifo con desplazamiento de centros de $\alpha = 38$ píxeles.

Figura 5.3: Comparación entre un anaglifo sin desplazamiento de centros (imagen inciso a), y un anaglifo con desplazamiento de $\alpha = 38$ píxeles, aplicado a los centros. Escena: laboratorio de cómputo del edificio Omega 2, en la DEP-FIE.

reducidas a un tamaño de 360×203 píxeles (Figura 5.4) para agilizar las pruebas. En las imágenes se muestran 3 objetos a diferentes profundidades: un objeto de metal color rojo, una caja blanca, y una pistola de silicon. El objeto de metal es el más próximo a las cámaras,

la pistola de silicon es el objeto más lejano, y la caja blanca se encuentra entre la pistola y la pieza de metal.



Figura 5.4: Imágenes ejemplo para llevar a cabo las diferentes pruebas de cálculo de disparidad.

Como se mencionó anteriormente, los algoritmos de búsqueda de correspondencia basados en correlación, son principalmente sensibles a las condiciones lumínicas de una escena, por lo que se llevaron a cabo tres diferentes pruebas para determinar bajo que condiciones se obtienen mejores resultados. La primer etapa de pruebas se llevó a cabo con las imágenes a color, tal como se muestran en la Figura 5.4, obteniendo los resultados mostrados en la Figura 5.5, y los tiempos mostrados en la Tabla 5.1.

Tabla 5.1: Tabla de tiempos del cálculo de la disparidad, utilizando las imágenes a color.

| Medida utilizada | Tiempo de ejecución (minutos) |
|--------------------|-------------------------------|
| <i>CCOEFF</i> | 6.70721 min. |
| <i>SSD</i> (4.1) | 7.87605 min. |
| <i>CORR</i> (4.3) | 5.42242 min. |
| <i>CORRN</i> (4.4) | 8.52829 min. |
| <i>SSDN</i> (4.2) | 8.60423 min. |
| <i>CCOEFFN</i> | 9.32172 min. |

Como se puede observar en la Figura 5.5, los mapas de disparidad con mejores resultados son los que se muestran en la Figura 5.5(e), 5.5(b), ya que la definición de los objetos se ve con mayor claridad y es posible tener una noción de qué objeto está más al fondo, y cuál más al frente. En contraste, en la Figura 5.5(c), es posible notar que la imagen solo se ve difuminada, pero la forma de los objetos se pierde totalmente. Por otro lado,

para las Figuras 5.5(a), 5.5(d), y 5.5(f), se observar que la forma de los objetos persiste, sin embargo, las imágenes se ven borrosas, lo cual se considera como un mal resultado si lo que se desea es estimar la profundidad.

En la Tabla 5.1, se observan los tiempos de ejecución para las diferentes medidas de distancia.

Para la segunda etapa de pruebas se utilizo el mismo par de imágenes, pero cambiando su esquema de color a escala de grises, obteniendo los resultados mostrados en la Figura 5.6, y los tiempos de procesamiento en la Tabla 5.2.



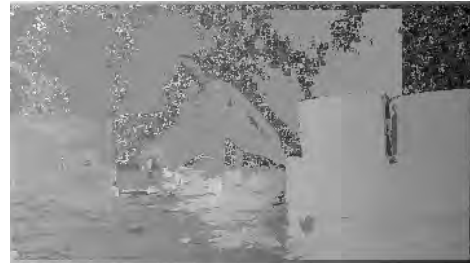
(a) *Medida de Distancia CCOEFF*



(b) *Medida de Distancia SSD (4.1)*



(c) *Medida de Distancia CORR (4.3)*



(d) *Medida de Distancia CORRN (4.4)*



(e) *Medida de Distancia SSDN (4.2)*



(f) *Medida de Distancia CCOEFFN*

Figura 5.5: Resultados del cálculo de la disparidad con las diferentes medidas de distancia, y utilizando las imágenes a color.

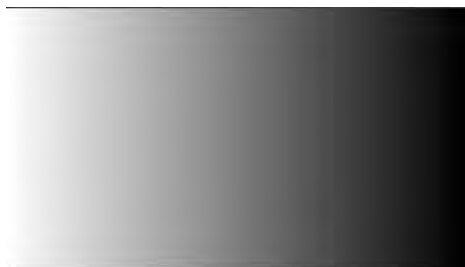
Para esta segunda etapa de pruebas se observa como nuevamente en la Figura 5.6(d) pierde la definición de los objetos por completo, mientras que las medidas de distancia que produjeron mejores resultados de disparidad nuevamente fueron los métodos SSDN en la Figura 5.6(e), y SSD en la Figura 5.6(b), ya que los objetos se ven definidos en su mayoría, sin embargo, los resultados son más pobres que en la prueba anterior, es decir, perdieron calidad. Para las demás medidas de distancia, se obtienen resultados similares a las pruebas anteriores, objetos poco definidos, e imágenes no muy claras. Sin embargo, en cuestión del tiempo de ejecución implementado por cualquier medida, es mucho menor a los observados



(a) *Medida de Distancia CCOEFF*



(b) *Medida de Distancia SSD (4.1)*



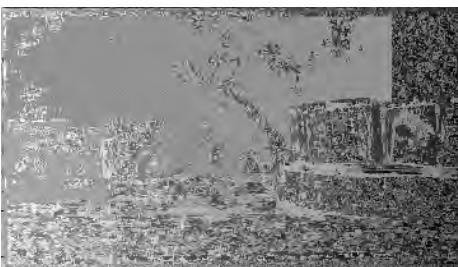
(c) *Medida de Distancia CORR (4.3)*



(d) *Medida de Distancia CORRN (4.4)*



(e) *Medida de Distancia SSDN (4.2)*



(f) *Medida de Distancia CCOEFFN*

Figura 5.6: Resultados del cálculo de la disparidad con los diferentes métodos, y utilizando las imágenes en escala de grises.

Tabla 5.2: Tabla de tiempos del cálculo de la disparidad, utilizando las imágenes en escala de grises.

| Medida utilizada | Tiempo de ejecución (minutos) |
|--------------------|-------------------------------|
| <i>CCOEFF</i> | 1.55892 min. |
| <i>SSD</i> (4.1) | 1.75308 min. |
| <i>CORR</i> (4.3) | 1.13322 min. |
| <i>CORRN</i> (4.4) | 2.40119 min. |
| <i>SSDN</i> (4.2) | 2.5437 min. |
| <i>CCOEFFN</i> | 2.54614 min. |

en la Tabla 5.1, por lo cual se podría concluir que con las imágenes en escala de grises la búsqueda de correspondencia es mucho mas rápida pero con resultados más pobres.

Para la tercera etapa de pruebas, se modificó el modelo de color RGB de las imágenes de la Figura 5.4, al modelo de color HSV (Hue, Saturation, Value, por sus siglas en ingles), donde Hue es la proporción de los dos colores más fuertes, Saturation es la intensidad de brillo de un color, y Value la intensidad de un color. De acuerdo con Aristide Chikando *et al.* [Aristide Chikando15], el esquema de proporciones de color en el modelo HSV, a diferencia del modelo RGB, utiliza una medida angular de valores para determinar el color de un pixel, siendo el rojo puro igual a 0 grados, el verde puro igual a 120 grados, y el azul puro igual a 240 grados. Como maneja un parámetro de saturación de color, lo hace menos sensible al brillo en las imágenes. Por lo tanto, en palabras más simples, HSV agrupa los colores por su tonalidad, variándolos por su saturación y brillo. OpenCV ofrece una herramienta llamada `cvtColor` para llevar acabo la conversión del modelo de color RGB a HSV, recibiendo como parámetros la imagen original, y la matriz destino donde se almacenara la misma imagen pero en formato del nuevo esquema de color. Por lo tanto, para esta etapa de las pruebas, se tomaron las dos imágenes mostradas en la Figura 5.4, y se cambiaron al modelo de color HSV, como se muestra en la Figura 5.7.

Utilizando las imágenes mostradas en la Figura 5.7, se obtuvieron los resultados mostrados en la Figura 5.8, y la Tabla 5.3 donde se muestran los resultados del tiempo de ejecución de todos los métodos.

Bajo el esquema de color HSV se pueden observar múltiples mejoras en los resul-

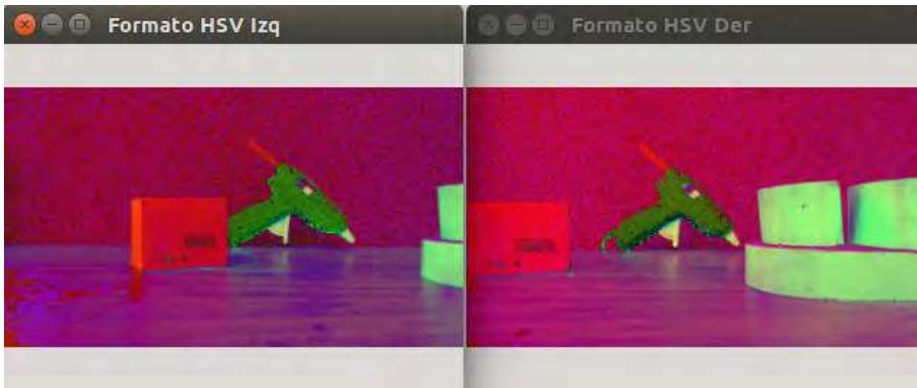


Figura 5.7: Imágenes ejemplo en formato de color HSV.

Tabla 5.3: Tabla de tiempos del cálculo de la disparidad, utilizando las imágenes con el modelo HSV.

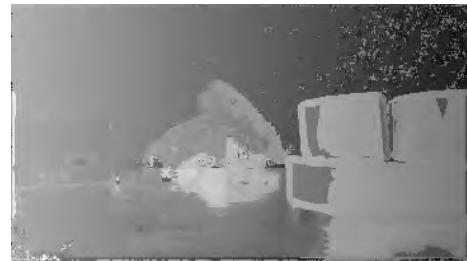
| Medida utilizada | Tiempo de ejecución (minutos) |
|--------------------|-------------------------------|
| <i>CCOEFF</i> | 6.2611 min. |
| <i>SSD</i> (4.1) | 7.95768 min. |
| <i>CORR</i> (4.3) | 5.44603 min. |
| <i>CORRN</i> (4.4) | 8.18237 min. |
| <i>SSDN</i> (4.2) | 8.22002 min. |
| <i>CCOEFFN</i> | 8.69183 min. |

tados para casi todas las medida de distancia, debido a que este esquema es menos sensible a las variaciones lumínicas. Sin embargo, con la aplicación de medidas de distancia como SSDN en la Figura 5.8(e) ó CORR en la Figura 5.8(d), a pesar de la claridad de dos de los objetos, se pierde información de un tercer objeto en la escena, quedando la medida de distancia de la Figura 5.8(a) (*CCOEFF*), como la que mejor resultados ofrece. Por otro lado, en la Tabla 5.3, es posible notar que los tiempos de de procesamiento, se disparan nuevamente en comparación con las pruebas anteriores.

En conclusión, se puede determinar a partir de las pruebas realizadas, que el cálculo de los mapas de disparidad utilizando un esquema de color HSV, ofrece mejores resultados en comparación con el esquema de color RGB, y el esquema de color en escala de grises. Sin embargo, los tiempos de ejecución son muchos mayores inclusive en comparación con el



(a) *Medida de Distancia CCOEFF*



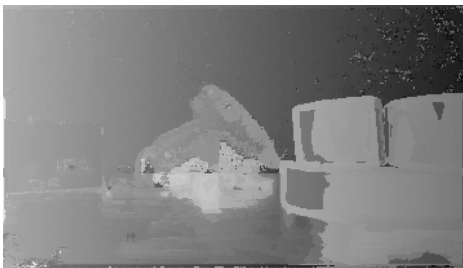
(b) *Medida de Distancia SSD (4.1)*



(c) *Medida de Distancia CORR (4.3)*



(d) *Medida de Distancia CORRN (4.4)*



(e) *Medida de Distancia SSDN (4.2)*

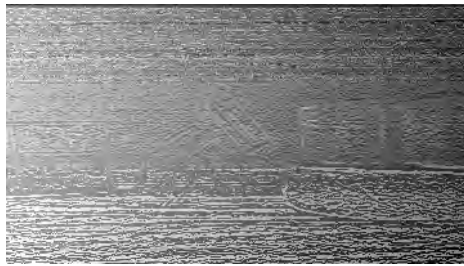


(f) *Medida de Distancia CCOEFFN*

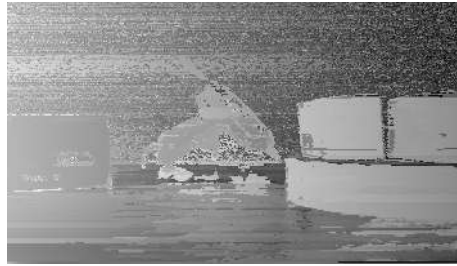
Figura 5.8: Resultados del cálculo de la disparidad con las diferentes Medida de Distancia, utilizando el esquema de color HSV.

esquema de color RGB. Se podría determinar que la mejor opción es utilizar las imágenes en la escala de grises debido al poco tiempo de ejecución que implementan para la búsqueda de correspondencia, pero dado que aún falta tomar en cuenta algunos factores relacionados con la búsqueda de correspondencia, y es necesario tener un enfoque orientado a la calidad de los resultados, y no al tiempo, se determinó que bajo el modelo de color HSV, se podrían obtener mejores resultados en pruebas futuras, sin embargo, los resultados también pueden estar sujetos a la calidad de las imágenes, y a las condiciones de la escena, por lo que para otros pares de imágenes, los resultados podrían variar.

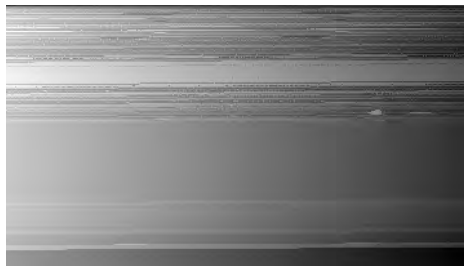
Por otro lado, tomando las imágenes de la Figura 5.7, y llevando a cabo el proceso de obtención del mapa de disparidad pero esta vez con el Algoritmo 6, se obtienen los resultados de la Figura 5.9.



(a) *Medida de Distancia CCOEFF*



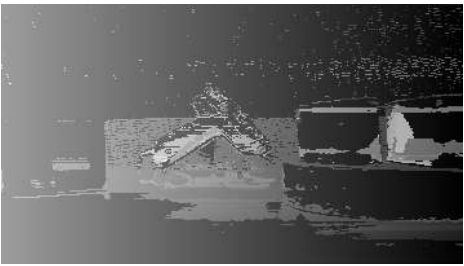
(b) *Medida de Distancia SSD (4.1)*



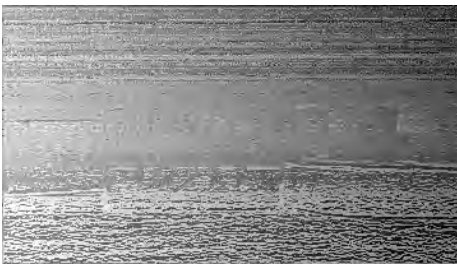
(c) *Medida de Distancia CORR (4.3)*



(d) *Medida de Distancia CORRN (4.4)*



(e) *Medida de Distancia SSDN (4.2)*



(f) *Medida de Distancia CCOEFFN*

Figura 5.9: Resultados del cálculo de la disparidad con las diferentes medidas de distancia, implementando líneas epipolares.

Como se puede apreciar en la Figura 5.9, añadiendo las líneas epipolares al cálculo de la disparidad, se mejora el resultado para las medidas SSD (4.1), y SSDN (4.2) observadas en la Figura 5.9(b), y en la Figura 5.9(e) las cuales delimitan la forma de los objetos, logrando un mapa de disparidad claro. En contraste, otras medidas de distancia como el CCOEFFN, empeoran sus resultados como se aprecia en la Figura 5.9(f), calculando mapas

Tabla 5.4: Tabla de tiempos del cálculo de la disparidad, implementando líneas epipolares.

| Medida utilizada | Tiempo de ejecución (minutos) |
|--------------------|-------------------------------|
| <i>CCOEFF</i> | 0.120706 min. |
| <i>SSD</i> (4.1) | 0.128103 min. |
| <i>CORR</i> (4.3) | 0.109939 min. |
| <i>CORRN</i> (4.4) | 0.126058 min. |
| <i>SSDN</i> (4.2) | 0.126592 min. |
| <i>CCOEFFN</i> | 0.127061 min. |

de disparidad con resultados peores en comparación con pruebas realizadas posteriormente como las mostradas en la Figura 5.8. Lo que si es claramente visible, es que el tiempo de ejecución se ve reducido súbitamente, pasando de tiempos de ejecución en minutos, a tiempos de ejecución en segundo, lo cuál resulta muy conveniente, ya que se logra obtener buenos resultados, y buenos tiempos de procesamiento. Y a pesar de que le objetivo inicial, no era enfocarse en los tiempos de ejecución, la implementación de líneas epipolares añade ese plus al cálculo y obtención de los mapas de disparidad.

Por un lado, se podría concluir que en general el cálculo de mapas de disparidad utilizando búsqueda basada en correlación, son bastante buenos si se deja de lado el factor del tiempo de ejecución. Por otro lado, añadiendo algunos factores como el cambio de modelo de color a HSV, y la inclusión de líneas epipolares al cálculo de mapas de disparidad, este tipo de metodologías pueden obtener resultados bastante buenos, y en tiempos de ejecución medidos en segundos, lo cuál resulta conveniente para llevar a cabo las tareas descritas de forma eficiente. Y a pesar de que el objetivo no está orientado a la mejora del tiempo, implementar los métodos anteriores para imágenes de 1280×720 , incrementará el tiempo de procesamiento de forma exponencial, por lo que tomando en cuenta un equilibrio entre calidad y tiempo, se optó trabajar con el cálculo de disparidad usando líneas epipolares, y el Algoritmo 6.

5.3. Resultados para la Estimación de la Profundidad

Siguiendo la escala de color mostrada en la Tabla 4.1, se llevó a cabo una prueba para demostrar que la disparidad, y el modelo de la relación Disparidad-Profundidad, expresado en (4.8), coinciden y funcionan correctamente.



Figura 5.10: En la figura se muestra dos objetos, una copa de cobre, y una figurilla de arcilla. Ambos objetos fueron colocados a una distancia aproximada de entre 115.0 cm y 120.0 cm con respecto de las cámaras, con la finalidad de probar el funcionamiento del modelo.

A partir del par de imágenes en la Figura 5.10, se calculó el mapa de disparidad Dx mostrado en la Figura 5.11.



Figura 5.11: Mapa de disparidad en x , de las imágenes de la Figura 5.10, utilizando SSDN (4.2).

Ejecutando una vez más el Algoritmo 7, se colorearon únicamente los píxeles con disparidad igual a cero para demostrar que se cumple lo expresado en (4.8). Los píxeles con

una disparidad diferente, fueron coloreados en negro, para resaltar los resultados, como se muestra en la Figura 5.12.



Figura 5.12: Resultados de aplicar el Algoritmo 7, coloreando únicamente los objetos con una profundidad de 120 centímetros con respecto de las cámaras.

Como segunda prueba, se tomaron imágenes de los mismos objetos, pero con diferentes profundidades. La figurilla de arcilla se posicionó a una distancia de 40 centímetros, mientras que la copa de cobre se mantuvo a una distancia de 120.0 centímetros como se muestra en la Figura 5.13. Posteriormente, utilizando de nueva cuenta el Algoritmo 7, se colorearon únicamente los píxeles con disparidades que corresponden a las distancias de los objetos, obteniendo los resultados de la Figura 5.15.

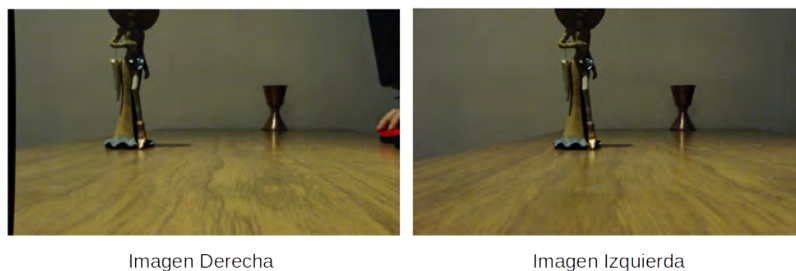


Figura 5.13: En la figura se muestra un nuevo par de imágenes de los objetos usados anteriormente, pero esta vez con la figurilla de arcilla a 40 centímetros, y la copa de cobre a 120.0 centímetros.

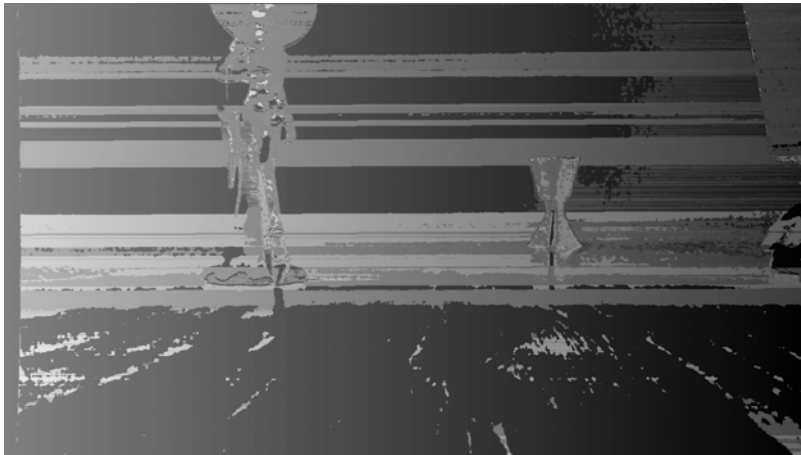


Figura 5.14: Mapa de disparidad a partir de las imágenes de la Figura 5.13, usando el método SSDN (4.2).



Figura 5.15: Profundidad calculada con el Algoritmo 7, donde se puede observar claramente la diferencia entre los colores de los objetos en la escena, indicando que se encuentran a diferentes profundidades.

Tabla 5.5: Tabla de tiempos de ejecución, para la estimación de la profundidad, a partir de las imágenes de la Figura 5.11 y de la Figura 5.14 respectivamente.

| Figura | Tiempo de ejecución (segundos) |
|------------------------------|--------------------------------|
| Resultados de la Figura 5.12 | 0.07129 seg. |
| Resultados de la Figura C | 0.07852 seg. |

En la Figura 5.15, los objetos están coloreados con diferentes tonalidades, la tonalidad RGB(255, 0, 0) para indicar que el objeto se encuentra a 120.0 centímetros, y la tonalidad RGB(255,130,0) para indicar que el objeto se encuentra a 40.0 centímetros. Pero a pesar de los buenos resultados que ofrecen tanto la disparidad como el modelo de la relación Disparidad-Profundidad, expresado en (4.8), existen secciones en las imágenes coloreadas que probablemente no se encuentran a la distancia definida por el color, sin embargo, aparecen marcadas. Este tipo situaciones son ocasionadas por ruido en las imágenes, y por falsas correspondencias en el cálculo de la disparidad.

En la Tabla 5.5 se muestran los resultados del tiempo de ejecución del Algoritmo 7, para los resultados mostrados en las Figuras 5.12 y 5.12, donde se puede observar que este proceso es realmente rápido para estimar la profundidad a partir de los mapas de disparidad dados, siendo el tiempo de ejecución menor a un segundo para ambos ejemplos.

Por otro lado, el diagrama de la Figura 5.16, describe a grandes rasgos el proceso que se lleva a cabo para la estimación de la profundidad.

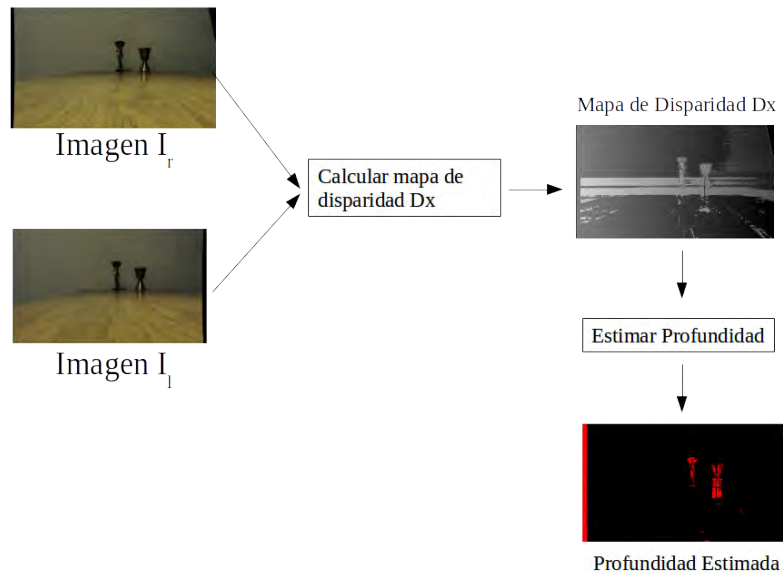


Figura 5.16: En este diagrama se muestra la secuencia de pasos necesaria para poder llevar a cabo la estimación de la profundidad, de los objetos en una escena, a partir de un par de imágenes estereoscópicas.

Finalmente, se presentan los resultados obtenidos a partir de las pruebas llevadas a cabo con las imágenes mostradas en la Figura 5.17, donde se tienen tres objetos diferentes: una figurilla de arcilla, una caja cuadrada y una copa de cobre. La figurilla de arcilla se encuentra a una distancia de 40 centímetros, la caja a 80 centímetros, y la copa de cobre a una distancia de alrededor de 115 centímetros.



Figura 5.17: En la figura se observan 3 objetos, una figura de arcilla a una distancia de 40 centímetros, una caja a 80 centímetros, y una copa de cobre a 115 centímetros de las cámaras.

Con las imágenes de la Figura 5.17, se calculó la disparidad utilizando el método SSDN (4.2). Los resultados obtenidos se presentan en la Figura 5.18.

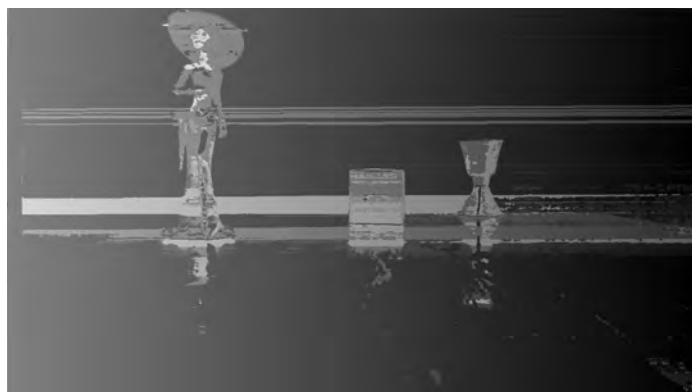


Figura 5.18: Resultados de Disparidad Usando el Método SSDN (4.2).

El tiempo de procesamiento para llevar a cabo la implementación del cálculo de la disparidad, fue de 5.20448 min usando el Método SSDN (4.2).

A partir de la disparidad calculada con las imágenes de la Figura 5.18, se estimó la profundidad obteniendo los resultados que se muestran en la Figura 5.19.



Figura 5.19: Resultados de la estimación de la profundidad, a partir de las disparidad mostrada en la figura 5.18

Como se observa en la Figura 5.19, la estimación de la profundidad tiene buenos resultados a pesar de que el tiempo de cálculo del mapa de disparidad es prolongado, por lo que se podría concluir que la estimación de la profundidad es acertada utilizando la disparidad con el método SSDN (4.2).

Se llevó a cabo otra prueba, tomando un par de imágenes pero con los objetos a menor distancia de las cámaras. Como se observa en la Figura 5.20, la caja es el objeto más cercano con una distancia de entre 23.0 y 28.0 centímetros, mientras que la figura de arcilla se encuentra a una distancia de 50 centímetros, y la copa de cobre a una distancia de entre 60 y 65 centímetros.

Nuevamente se calcula la disparidad utilizando SSDN (4.2), obteniendo los resultados mostrados en la Figura 5.21.

El tiempo de ejecución para el cálculo de la disparidad en esta prueba, fue de 5.24207 min. Sin embargo al estimar la profundidad, es claro observar como las formas de las figuras se colorean lo suficiente como para mantener la forma de los objetos y demostrar que el Algoritmo 7 estima correctamente la profundidad, como se muestra en la Figura 5.22.



Imagen Derecha

Imagen Izquierda

Figura 5.20: En la figura se observa una caja a 18 centímetros, la figurilla de arcilla a 50 centímetros aproximadamente, y una copa de cobre a 65 centímetros de las cámaras.



Figura 5.21: Resultados de Disparidad Usando el Método SSDN (4.2).

5.4. Conclusiones del Capítulo

En este capítulo quedó demostrado que con el sistema de cámaras calibrado y rectificado, es posible generar imágenes anaglifo. Sin embargo, con la modificación a los parámetros intrínsecos del modelo de las cámaras, añadiendo la variable α , fue posible mejorar la calidad de estos anaglifos de tal forma que al visualizarlos, no resulta cansado para los ojos, y es posible visualizar el efecto de tridimensionalidad de forma clara.

Por otro lado, con las diferentes pruebas llevadas a cabo a lo largo de este capítulo, es posible concluir que la relación de Disparidad-Profundidad obtenida a partir de aplicar la técnica de imágenes anaglifo, expresada en (3.24) para el sistema de cámaras, funcionó como



Figura 5.22: Resultados de la estimación de la Profundidad a partir de la disparidad.

marco de referencia para la estimación de la profundidad, facilitando dicha tarea con los mapas de disparidad calculados con la medida de distancia $SSDN$ (4.2).

Se demostró también, que con el Algoritmo 7, es posible determinar la distancia de los objetos en una escena, con respecto de las cámaras. En las diferentes pruebas llevadas a cabo, se logró definir la distancia de los objetos de forma muy acertada. Sin embargo, a pesar de los buenos resultados, quedó claro que es necesario buscar otro tipo de metodologías para el cálculo de mapas de disparidad, ya que los resultados obtenidos en este trabajo de tesis, no estuvieron exentos de valores de falsas correspondencias, o de ruido en las imágenes. Por otro lado, a pesar de que los tiempos de ejecución, no fueron un factor de importancia en esta tesis, queda claro que también es necesario buscar metodologías búsqueda de correspondencia, capaces de resolver dicho problema, en tiempos de ejecución más pequeños, sin importar los tamaños de las imágenes.

Capítulo 6

Conclusiones y Trabajos Futuros

6.1. Conclusiones

Se concluye que con el modelo PinHole es posible establecer un modelo funcional para un sistema de cámaras estereoscópico.

Se abordó el proceso de construcción física del sistema estéreo de cámaras, se identificaron sus ventajas y desventajas mencionadas en el Capítulo 3, y se explicó el proceso de la calibración, utilizando la librería OpenCV.

Por medio de la calibración, se corrigieron artificialmente los problemas de alineación en el sistema, y se logró rectificar las cámaras para efectos de calidad en las imágenes tomadas con dicho sistema. Por otro lado, nuevamente gracias a la librería OpenCV, se logró implementar diferentes algoritmos desarrollados a lo largo de este trabajo de tesis. En consecuencia, fue posible elaborar las propuestas de diferentes algoritmos, con el objetivo de estimar la profundidad de los objetos en una escena capturada con el sistema de cámaras.

Se logró demostrar que a partir de imágenes tomadas por el sistema estéreo calibrado, fue posible generar imágenes anaglifo, las cuales fueron utilizadas para evaluar visualmente los resultados de la calibración, y además utilizarlas como una forma didáctica para observar la relación de proporcionalidad entre la disparidad, y la profundidad de los objetos. De igual forma se logró mejorar la calidad de los anaglifos, modificando los parámetros intrínsecos del modelo del sistema, además de establecer un marco de referencia para

la estimación de la profundidad, a partir de estos cambios.

Se concluye también que la búsqueda de correspondencia basada en la correlación, resultó mejor implementando la medida de distancia SSDN (4.2) con líneas epipolares, logrando desarrollar un algoritmo para la estimación de la profundidad con buena precisión. Sin embargo, utilizando los mapas de disparidad calculados con la medida de distancia SSDN, se notó que cuenta con algunos problemas de falsas correspondencias, lo cuál deja en claro que existe la necesidad de buscar otras alternativas para obtener mejores resultados para calcular los mapas de disparidad.

A pesar de las falsas correspondencias en los mapas de disparidad, utilizando un mapa de color conocido como Hot colormap, se demostró que es posible identificar la profundidad de los objetos con respecto de las cámaras, identificando con un color diferente la posición de los pixeles, en función de los valores de disparidad.

Se concluye que los objetivos de la tesis fueron cumplidos en gran medida, y se logro obtener resultados satisfactorios y claros, como resultado de este trabajo de investigación.

6.2. Trabajos Futuros

En la estimación de la profundidad a partir de sistemas estereoscópicos, existe un camino largo por recorrer con el objetivo de medir de forma precisa la distancia de un objeto. Por lo tanto, se presentan los siguientes puntos, como trabajos futuros, los cuales podrían hacer una diferencia en los resultados a la hora de calcular profundidad con un sistema de cámaras.

1. Una tarea fundamental para la estimación de la profundidad, es el cálculo de los mapas de disparidad. Es necesario invertir tiempo en la búsqueda de metodologías más robustas, que permitan obtener mapas de disparidad de mejor calidad, y que no por esto se incremente el tiempo de procesamiento para obtenerlos. Por el contrario, uno de los objetivos sería encontrar una metodología capaz de llevar a cabo esta tarea, tal vez en segundos de tiempo.
2. Se propone experimentar con metodologías de búsqueda de correspondencia basadas en puntos característicos, ya que debido a la cantidad de información con la que

llevan a cabo el procesamiento de búsqueda, estos pueden ser una opción viable para la reducción del tiempo a la hora de calcular mapas de disparidad, y además este tipo de metodologías, son menos susceptibles a problemas como la iluminación de una escena, o a las oclusiones entre imágenes.

3. Para la mejora de los mapas de disparidad, se propone implementar un filtro de medianas a los mapas de disparidad, con el objetivo de reducir falsas correspondencias debido a ruido en las imágenes.
4. Se propone la estimación de la profundidad como método para la segmentación en imágenes, tomando como factor principal la profundidad de los puntos en la escena, y no depender de la segmentación por colores, o similitud entre vecinos.

Referencias

- [Abiel Aguilar González15] Abiel Aguilar González, M. A. E. J. L. C. A. H. R. H. d. L. A. S. A., Madaín Pérez Patricio. An fpga correlation-edge distance approach for disparity map. *Electronics, Communications and Computers (CONIELECOMP), 2015 International Conference on. IEEE, 2015. p. 21-28.*, 2015.
- [Anwar Hasni Abu Hasan09] Anwar Hasni Abu Hasan, R. A. H. y Johar, M. H. Region of interest in disparity mapping for navigation of stereo vision autonomous guided vehicle. *2009 International Conference on Computer Technology and Development*, 2009.
- [Aristide Chikando15] Aristide Chikando, J. K. Optimizing image segmentation using color model mixtures. *Proceedings of the 34th Applied Imagery and Pattern Recognition Workshop (AIPR05)*, 2015.
- [Baigen He11] Baigen He, Z. M. y Wei, Y. A matching algorithm on statistical properties of harris. *Proceeding of the IEEE International Conference on Information and Automation Shenzhen, China June 2011*, 2011.
- [Biadgie14] Biadgie, Y. y Sohn, K.-A. Feature detector using adaptive accelerated segment test. *Information Science and Applications (ICISA), 2014 International Conference on*, 2014.

- [Bradski08] Bradski, G. y Kaehler, A. *Learning OpenCV*. O'Reilly Media, Inc., 1ª ed^{ón}., 2008. ISBN 978-0-596-51613-0.
- [Coe78] Coe, B. *Colour Photography. The first hundred years 1840-1940*. Ash and Grant, 1ª ed^{ón}., 1978. ISBN 978-0-596-51613-0.
- [Compañ04] Compañ, P. *Estimación de la disparidad en visión estereoscópica. Tratamiento del color*. VISIÓN COMPUTACIONAL. Universidad de Alicante, Dpto. de Ciencia de la Computación e Inteligencia Artificial, 1ª ed^{ón}., 2004.
- [Dawson-Howe14] Dawson-Howe, K. *A PRACTICAL INTRODUCTION TO COMPUTER VISION WITH OPENCV*. John Wiley and Sons Ltd, 1ª ed^{ón}., 2014. ISBN 9781118848456.
- [Fuentes14a] Fuentes, V. Canion 3d, Noviembre 2014.
URL <http://recursos.cepindalo.es/mod/book/>
- [Fuentes14b] Fuentes, V. Gafas anaglifo, Noviembre 2014.
URL <https://www.ufopolis.com>
- [García99] García, L. J. T. *MÉTODOS ROBUSTOS DE RECONSTRUCCIÓN DE ESCENAS TRIDIMENSIONALES EMPLEANDO INFORMACIÓN BINOCULAR*. Universidad Politécnica de Madrid, 1ª ed^{ón}., 1999.
- [Harris88] Harris, C. y Stephens, M. A combined corner and edge detector. *En In Proc. of Fourth Alvey Vision Conference*, págs. 147–151. 1988.
- [Hartley04] Hartley, R. y Zisserman, A. *Multiple View Geometry in Computer Vision*. Cambridge University, 2ª ed^{ón}., 2004. ISBN 0-511-18618-5.

- [Heras] Heras, R. R. *DESARROLLO DE UN ROBOT DE TELEPRESENCIA EQUIPADO CON UN SISTEMA DE VISIÓN 3D Y TELÉMETRO LÁSER*.
- [Laganiere11] Laganiere, R. *OpenCV 2 Computer Vision Application Programming Cookbook*. Packt Publishing Ltd., 1ª ed^{ón}., 2011. ISBN 978-1-849513-24-1.
- [Lecumberry14] Lecumberry, F. Cálculo de disparidad en imágenes estéreo, una comparación. *III Workshop de Computación Gráfica, Imágenes y Visualización*, 2014.
- [Lourakis05] Lourakis, M. I. A. *A Brief Description of the Levenberg-Marquardt Algorithm Implemented*, Febrero 2005.
- [Lowe04] Lowe, D. G. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision* 60(2), 91–110, 2004, 2004.
- [Tippetts11] Tippetts, B. J. y Lee, D.-J. Dense disparity real-time stereo vision algorithm for resource-limited systems. *IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, VOL. 21, NO. 10, OCTOBER 2011*, 2011.
- [Ude10] Ude, A. *Robot Vision*. In-Tech, 1ª ed^{ón}., 2010. ISBN 978-953-307-077-3.
- [Valsan15] Valsan, V. N. y Patil, C. Y. A system on chip based stereo vision approach for disparity measurement. *2015 International Conference on Industrial Instrumentation and Control (ICIC) Colege of Engineering Pune, India*, 2015.