

TESIS DEFENDIDA POR

Cesar Augusto Puente Montejano

Y aprobada por el siguiente comité:

Dr. Gustavo Olague Caballero

Codirector del Comité

M.C. José Luis Briseño Cervantes

Codirector del Comité

Dr. César Cruz Hernández

Miembro del Comité

Dr. Alejandro Hinojosa Corona

Miembro del Comité

Dr. Eugenio Rafael Méndez Méndez

Miembro del Comité

Dr. Jorge Torres Rodríguez

Miembro del Comité

Dr. Pedro Gilberto López Mariscal

*Coordinador del Programa en
Ciencias de la Computación*

Dr. Raúl Ramón Castro Escamilla

*Director de Estudios
de Posgrado*

2 de Diciembre del 2005

TESIS

presentada por

Cesar Augusto Puente Montejano

para cubrir parcialmente los requisitos necesarios para
obtener el grado de **Maestro en Ciencias** en el

Centro de Investigación Científica y de Educación Superior de Ensenada

Especialidad: **Ciencias de la Computación**

**Reconstrucción tridimensional a partir dos imágenes basada
en el proceso de búsqueda de las abejas**

Fecha: Diciembre del 2005.

Miembros del comité: Dr. Gustavo Olague Caballero.
M.C. José Luis Briseño Cervantes.
Dr. César Cruz Hernández.
Dr. Alejandro Hinojosa Corona.
Dr. Eugenio Rafael Méndez Méndez.
Dr. Jorge Torres Rodriguez.

Tesis realizada en las instalaciones del Departamento de Ciencias de la Computación del
CICESE, bajo la dirección del Dr. Gustavo Olague.

RESUMEN de la tesis de **Cesar Augusto Puente Montejano**, presentada como requisito parcial para obtener el grado de MAESTRO EN CIENCIAS en CIENCIAS DE LA COMPUTACIÓN. Ensenada, B. C. Diciembre del 2005.

Reconstrucción tridimensional a partir dos imágenes basada en el proceso de búsqueda de las abejas

Resumen aprobado por:

Dr. Gustavo Olague Caballero

Codirector de Tesis

M.C. José Luis Briseño Cervantes

Codirector de Tesis

La reconstrucción tridimensional es un problema complejo que aún no ha sido resuelto satisfactoriamente, existen diversas técnicas y métodos propuestos por los especialistas en visión por computadora y en fotogrametría, pero ninguno de ellos resuelve el problema general, de manera que pueda afrontar cualquier circunstancia. La reconstrucción puede llevarse a cabo mediante la geometría de múltiples vistas. El caso base es la reconstrucción a partir dos imágenes. Una reconstrucción a partir de dos imágenes contiene muchas zonas ambiguas de las que no se puede determinar fácilmente su posición en el espacio. Estas zonas están sujetas a problemas de oclusión y de baja textura principalmente. Esto ocasiona modelos tridimensionales con demasiados puntos incoherentes con respecto a la escena original.

En este trabajo se desarrolla un algoritmo evolutivo inteligente basado en el comportamiento de las abejas para confrontar el problema de la reconstrucción, el cual, usando un enfoque diferente, reduce dramáticamente los resultados erróneos enfocándose en la búsqueda de información significativa en las imágenes: la textura y la semejanza. Una vez encontradas las zonas relevantes, el algoritmo se encarga de *propagar la precisión* mediante el proceso de reclutamiento, una forma de comunicación compleja que utilizan las abejas para transmitir información. De esta manera, abejas artificiales se agrupan alrededor de las zonas más significativas de la escena para generar el modelo tridimensional, conformando una reconstrucción quasi-densa.

Palabras clave: Reconstrucción 3D, visión estereo, algoritmos evolutivos, enfoque bio-inspirado, abejas.

ABSTRACT of the thesis presented by **Cesar Augusto Puente Montejano**, as a partial requirement to obtain the MASTER SCIENCE degree in COMPUTER SCIENCES. Ensenada, B. C. december 2005.

3D Reconstruction from two images based on honeybees foraging process

Abstract approved by:

Dr. Gustavo Olague Caballero

Thesis codirector

M.C. José Luis Briseño Cervantes

Thesis codirector

Three-dimensional reconstruction is a complex problem that has not been solved satisfactorily. Several techniques and methods has been proposed by the specialists in computer vision and photogrammetry. But none of them solves the general problem, so that it can confront any circumstance. The reconstruction can be carried out by means of the multiple view geometry. The base case is composed by two images. A reconstrucción from two images contains many ambiguous zones. The position in the space of this zones cannot be determined easily. These zones are mainly subject to problems of occlusion and low texture. This causes three-dimensional models with too many incoherent points with respect to the original scene.

In this work an intelligent evolutionary algorithm based on the honeybee behaviour is developed in order to confront the reconstruction problem. This algorithm uses a different approach to reduce dramatically erroneous results focusing in the relevant information of the images: texture and similarity. Once found relevant zones in texture, the algorithm spreads the precision by means of the recruitment process, a form of communication of the bees to transmit information. In this way, artificial bees are grouped around the most significant zones of the scene in order to generate a three-dimensional model, conforming a quasi-dense reconstruction of the scene.

Keywords: 3D Reconstruction, stereo vision, evolutionary algorithms, bio-inspired approach, honeybee.

Dedicado a mis padres

Juan Puente
y
Ma. de Jesús Montejano

Agradecimientos

Agradezco a mis hermanos Juan, Lucero, y Erick, por su apoyo a pesar de la distancia.

Al Dr. Gustavo Olague, por su colaboración y por compartir sus conocimientos a lo largo de la realización de este trabajo.

Al M.C. José Luis Briseño, por sus comentarios y consideraciones.

A los miembros del Comité de Tesis, por su interés y aportaciones a este trabajo.

Al Centro de Investigación Científica y de Educación Superior de Ensenada y al Consejo Nacional de Ciencia y Tecnología, por darme las facilidades para estudiar en esta institución.

A mis compañeros del laboratorio *Evovisión*, en especial a Cynthia, por ayudarme a comprender las bases de la investigación y por su apoyo durante este año y a Enrique, por sus comentarios y discusiones sobre el tema.

A mis compañeros del CICESE. En especial a Marcel, por animarme y motivarme a trabajar tenazmente.

A Angélica por su apoyo incondicional.

Ensenada, México
2 de Diciembre del 2005.

Cesar Augusto Puente Montejano

Tabla de Contenido

Capítulo	Página
Resumen	ii
Abstract	iii
Lista de Figuras	x
Lista de Tablas	xii
I Introducción.	1
I.1 Visión por computadora.	1
I.2 Motivación.	2
I.3 Antecedentes.	4
I.3.1 Reconstrucción basada en sensores activos.	4
I.3.2 Reconstrucción basada en sensores pasivos.	5
I.4 Alcances de esta tesis.	9
I.4.1 Objetivo General.	10
I.4.2 Objetivos Específicos.	10
I.5 Organización de esta tesis.	11
I Enfoque Geométrico	13
II La geometría aplicada a la visión por computadora.	14
II.1 Geometría proyectiva.	15
II.1.1 El plano proyectivo.	15
II.2 Geometría afín.	18
II.3 Geometría euclidiana.	19
II.4 Jerarquía de las transformaciones geométricas.	21
II.4.1 Isometrías.	22
II.4.2 Similaridades.	23
II.4.3 Afinidades.	24
II.4.4 Proyectividades.	26
II.4.5 Comparación entre transformaciones.	28
II.5 Modelo geométrico de una cámara.	30
II.5.1 Proceso de adquisición de la imagen.	33
II.6 Calibración de la cámara.	36
II.6.1 Solución utilizando la restricción $m_{34} = 1$	38

Tabla de Contenido (Continuación)

Capítulo	Página
II.6.2 Solución utilizando la restricción $\ m_3\ = 1$: Método Faugeras-Toscani.	39
II.7 Conclusión.	41
III La relación entre dos imágenes: Geometría Epipolar.	43
III.1 Geometría Epipolar.	43
III.2 La matriz Fundamental.	46
III.2.1 Derivación Geométrica.	48
III.2.2 Propiedades.	50
III.3 Las matrices de proyección a partir de la matriz fundamental.	52
III.3.1 Forma canónica de las matrices de proyección.	53
III.4 Calculo de la matriz fundamental.	54
III.4.1 Solución Lineal. (Algoritmo de los ocho puntos)	55
III.4.2 Algoritmo de los ocho puntos normalizado.	57
III.4.3 Algoritmo de Estimación de Máxima Verosimilitud. (MLE)	58
III.4.4 Algoritmo RANSAC.	60
III.4.5 La matriz fundamental a partir de las matrices de proyección.	62
III.5 Experimentación.	67
III.6 Conclusiones.	70
IV Reconstrucción 3D: Triangulación.	74
IV.1 Correspondencia Estéreo.	74
IV.1.1 Disparidad estéreo.	81
IV.2 Triangulación.	83
IV.2.1 Método del punto Medio.	84
IV.2.2 Método lineal.	85
IV.2.3 Método Polinomial.	87
IV.3 Experimentación.	91
IV.3.1 Correspondencia estéreo.	91
IV.3.2 Triangulación.	96
IV.3.3 Triangulación lineal.	97
IV.3.4 Triangulación polinomial.	98
IV.4 Conclusiones.	100

Tabla de Contenido (Continuación)

Capítulo	Página
II Enfoque Bio-inspirado.	107
V Introducción a los algoritmos evolutivos.	108
V.1 Métodos clásicos de búsqueda y optimización.	108
V.2 Principios Básicos.	110
V.2.1 La evolución natural.	111
V.3 Estructura General.	113
V.3.1 Generación de la población.	114
V.3.2 Evaluación.	115
V.3.3 Selección.	116
V.3.4 Reproducción.	118
V.3.5 Criterio de paro.	129
V.4 Técnicas Evolutivas.	130
V.4.1 Algoritmos Genéticos.	131
V.4.2 Estrategias de Evolución.	132
V.5 Conclusiones.	135
VI Reconstrucción 3D: Algoritmo de búsqueda basada en el comporta- miento de las abejas.	138
VI.1 Planteamiento de la reconstrucción 3D como problema de búsqueda. . .	140
VI.2 Comportamiento de las abejas.	142
VI.2.1 Organización social: castas.	143
VI.2.2 ¿Porqué abejas?	149
VI.2.3 El proceso de búsqueda.	152
VI.3 El proceso de búsqueda como paradigma de reconstrucción.	154
VI.3.1 Estructura del algoritmo.	155
VI.3.2 Funciones de aptitud.	161
VI.3.3 Los operadores evolutivos	167
VI.3.4 Reclutamiento.	169
VI.4 Experimentación.	172
VI.4.1 Pruebas con el operador de Mutación.	172
VI.4.2 Pruebas con el operador de Cruce.	176
VI.4.3 Pruebas con el operador de repartición.	179

Tabla de Contenido (Continuación)

Capítulo	Página
VI.5 Conclusiones.	180
VII Experimentación.	183
VII.1 Experimento I.	185
VII.1.1 Comparación con la triangulación.	186
VII.2 Experimento II.	188
VII.2.1 Comparación con la triangulación.	189
VII.3 Conclusiones.	190
VIII Conclusiones y trabajo futuro.	194

Lista de Figuras

Figura		Página
1	Etapas que conforman la visión por computadora.	3
2	Sensores Activos	6
3	Un modelo del plano proyectivo	18
4	Transformación Afín	25
5	Transformación proyectiva entre dos planos.	28
6	Comparación entre transformaciones.	29
7	Modelo de la cámara de ojo de aguja.	32
8	Modelo Geométrico de la adquisición de la imagen de una cámara.	34
9	Geometría Epipolar.	45
10	Par estéreo con cámaras convergentes.	46
11	Par estéreo con cámaras paralelas.	47
12	Geometría de F.	48
13	Derivación Geométrica de F.	50
14	Técnica RANSAC	61
15	Par estéreo.	64
16	Primer par de imágenes utilizadas para realizar pruebas.	68
17	Línea epipolar	71
18	Triangulación.	84
19	Método del punto medio.	85
20	Método polinomial.	88
21	Reformulación del problema de optimización.	89
22	El mapa de disparidad	95
23	El mapa de disparidad (cont.)	104
24	Criterios de aceptación en la correspondencia estéreo	105
25	Segundo par de imágenes utilizadas para realizar pruebas.	105
26	Aplicación del método polinomial	106
27	Triangulación con errores.	106
28	Métodos clásicos de búsqueda y optimización	109
29	Diagrama de flujo general de un algoritmo evolutivo.	114
30	Operadores de selección	118
31	Operadores de mutación	123
32	Operadores de cruzamiento	126

Lista de Figuras (Continuación)

Figura		Página
33	Operador de repartición	128
34	Estructura general de un algoritmo genético.	132
35	Estructura general de una estrategia de evolución	135
36	Significado de la danza con meneo.	154
37	Proceso de búsqueda de las abejas	156
38	Generación de la nueva población	159
39	El algoritmo basado en el proceso de búsqueda de las abejas	162
40	Función de aptitud de las abejas exploradoras.	165
41	Comparación entre mutación normal y polinomial basada en individuos exitosos.	174
42	Comparación entre mutación normal y polinomial basada en aptitud y diversidad.	175
43	Comparación entre cruzamiento SBX y lineal basada en individuos exitosos.	177
44	Comparación entre cruzamiento SBX y lineal.	178
45	Comparación entre diferentes coeficientes de repartición.	180
46	Experimento I.	192
47	Experimento II.	193

Lista de Tablas

Tabla		Página
I	Tipos de reconstrucción, según la información inicial.	8
II	Resumen de la jerarquía de transformaciones.	31
III	Resultados de experimentaciones con los métodos para calcular F. . .	69
IV	Evaluación de la universidad Middlebury.	93
V	Evaluación de la universidad Middlebury. Nueva versión.	94
VI	Medida de error en el modelo tridimensional de las imágenes de la figura 16.	98
VII	Medida de error en el modelo tridimensional de las imágenes de la figura 25.	99
VIII	Parámetros del algoritmo para la comparación entre operadores de mutación.	173
IX	Parámetros del algoritmo para la comparación entre operadores de cruzamiento.	177
X	Parámetros del algoritmo para generar un modelo 3D a partir de los pares estéreos de las figuras 16 y 25.	184
XI	Medida de error en el modelo tridimensional del par estéreo de la figura 16	188
XII	Medida de error en el modelo tridimensional del par estéreo de la figura 25	189

Capítulo I

Introducción.

En la década de los 60's, cuando la inteligencia artificial estaba en boga, los expertos en este campo enfrentaban el problema de lograr que una computadora *viera*. Han pasado más de cuarenta años y este problema sigue sin resolverse del todo, por el contrario, cada vez se representa como un reto mayor. Es por ello que surgió la visión por computadora, que es una disciplina que encara este problema con la ayuda principalmente de la geometría y de las ciencias de la computación, como la inteligencia artificial y de la tecnología como son las cámaras digitales. En esencia, el objetivo de la Visión por Computadora es: *Proveer del sentido de la vista a sistemas autónomos para que puedan interactuar de forma eficiente en ambientes complejos* (Forsyth y Ponce (2002)).

Uno de los problemas que encara la visión por computadora es la reconstrucción tridimensional de los objetos que se encuentran en una escena, a partir de una o más imágenes de ésta. Existen varias formas de abordar este problema, las cuales, como se verá durante este capítulo, dependen de los instrumentos y herramientas que se utilicen.

I.1 Visión por computadora.

La visión por computadora es una disciplina que puede dividirse en etapas, que van desde que una imagen es adquirida hasta el objetivo final, que es lograr una percepción del mundo (figura 1). Para adquirir la imagen se involucran principios de la geometría

proyectiva. Después, para eliminar el ruido de la imagen y restaurarla, está la etapa de procesamiento. Posteriormente, durante la etapa de análisis se identifican algunos de los componentes de la imagen mediante métodos como la detección de bordes y se modela tridimensionalmente la información contenida en la imagen, mediante la reconstrucción. Finalmente, en la etapa de visión artificial, también llamada visión de alto nivel, se comprende el contenido de la escena, el conocimiento de los objetos que la componen y la manipulación de este conocimiento a alto nivel para que un sistema autónomo lo utilice como guía para desenvolverse en su entorno.

I.2 Motivación.

La reconstrucción tridimensional juega un papel muy importante dentro de la visión por computadora, ya que se puede utilizar como herramienta en tareas de alto nivel como el reconocimiento de objetos y la comprensión de escenas, así como la generación de nuevas imágenes a partir del modelo tridimensional.

Además, por si sola, la reconstrucción tridimensional puede auxiliar a otros campos de investigación que necesitan la visualización digital como herramienta de trabajo. Por ejemplo en la arquitectura es capaz de generar modelos tridimensionales realistas de construcciones para su estudio detallado, en la arqueología funge como apoyo en excavaciones y otros estudios de campo que ésta realiza (Pollefeys *et al.* (1998)). La ingeniería industrial y mecánica encuentran un excelente apoyo para aplicaciones que tienen que ver con la inspección visual de procesos industriales, la aeronautica ha comenzado a utilizarla como herramienta de orientación para la navegación (Thakoor *et al.* (2004)). En la industria cinematográfica y de videojuegos se requieren modelos para simulaciones

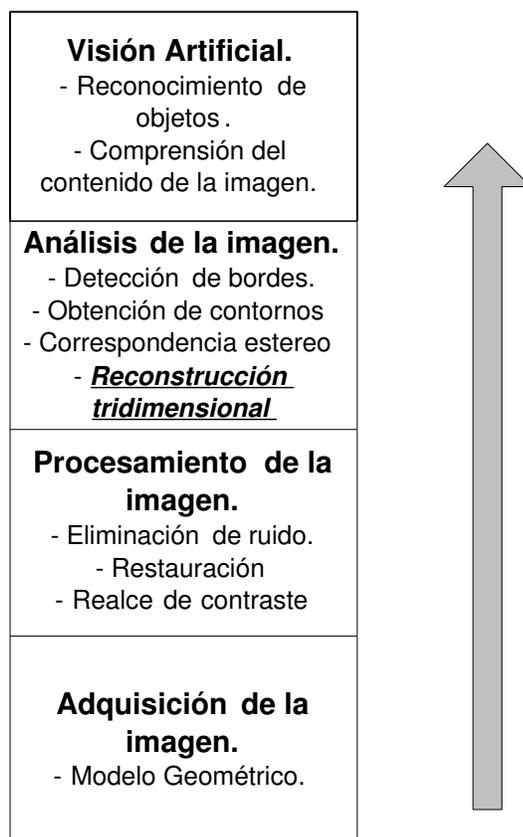


Figura 1: Etapas que conforman la visión por computadora.

gráficas más realistas, de la misma manera que en el mundo de las telecomunicaciones que incluyen imágenes (Koch y Good (1998)). La robótica requiere del conocimiento de su entorno de trabajo y la información visual es de gran importancia para lograr este objetivo (Boumaza y Louchet (2001)). Con éstas y otras aplicaciones que se van generando conforme madura este campo y se resuelve de manera satisfactoria su problemática, la reconstrucción tridimensional merece la atención que se le da dentro de la visión por computadora.

I.3 Antecedentes.

Una definición del problema de reconstrucción tridimensional de escenas la ha dado Marchand (Marchand y Chaumette (1999)):

La recuperación digital de la estructura espacial de una escena, sin el conocimiento previo de sus dimensiones; ni la localización, número y medidas de los objetos que la componen.

La recuperación de la estructura de una escena se puede llevar a cabo de dos maneras diferentes, de acuerdo al tipo de sensor que se utilice para obtener la información, y estos pueden ser pasivos o activos.

I.3.1 Reconstrucción basada en sensores activos.

Se le llama así al tipo de sensor que realiza una acción sobre la escena con el fin de obtener la información. Existen dos técnicas muy utilizadas basadas en sensores activos. El escáner por láser y la luz codificada (Davis *et al.* (2003)).

- **Escáner láser.** Un escáner láser típico se compone de una cámara y un apuntador láser que realiza un barrido de la escena. Este barrido posee una forma geométrica, la mas simple y común es la de un plano generado desde un punto de proyección que se mueve a lo largo de la escena que se debe reconstruir (figura 2, a). De esta manera la cámara registra las intersecciones de este plano con los objetos presentes. Existen sistemas comerciales de este tipo, como el IVP Ranger SAH5. La desventaja de este tipo de sensores es que son costosos y son propensos a ruido

al adquirir los datos, por ejemplo no son confiables en condiciones de mucha luz; y poseen poca resolución.

- **Luz codificada.** Este método determina la profundidad a la que se encuentra un objeto proyectando patrones de luz a la escena que cambian a través del tiempo, y se registran con una cámara (figura 2, b). Estos patrones están ordenados de manera que cada columna de píxeles que se proyecta en la imagen de la cámara puede ser identificada como única. Cada columna representa un vector que se compara con una base de datos donde se encuentran almacenados la forma en que se presentan los patrones a determinadas profundidades. Así, la profundidad a la que se encuentra lo que se proyecta en cada píxel se determina basándose en el patrón que se observe en él. El trabajo pesado para esta técnica lo representa el diseño del patrón a utilizar; el cual debe ser óptimo en términos de tiempo y robustez. La secuencia del patrón debe durar el menor tiempo posible y debe ser insensible a la intensidad de la luz propia de la escena. Precisamente una de las desventajas de esta técnica es que requiere que la luz de la escena sea poca o hasta nula.

I.3.2 Reconstrucción basada en sensores pasivos.

Los sensores pasivos son aquellos que no necesitan interactuar con la escena para obtener la información de la misma. El sensor pasivo por excelencia es la cámara; que es la herramienta principal de la visión por computadora. Así, la reconstrucción basada en cámaras se lleva a cabo a partir de las imágenes que éstas adquieren. El proceso de adquisición de una imagen por la cámara se explica en el capítulo 2.

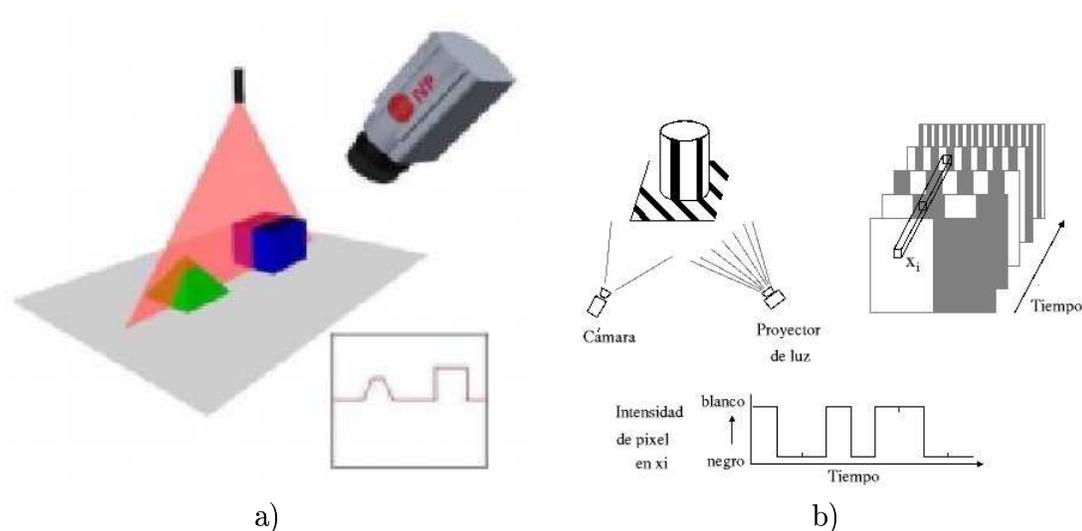


Figura 2: Sensores activos. a) Escáner láser. b) Luz codificada.

Es posible una reconstrucción basada en cámaras a partir de una sola imagen, o a partir de varias imágenes:

- **Una sola imagen.** Ésta solución es capaz de obtener las dimensiones espaciales de los objetos y su localización en la escena. Sin embargo, no consigue modelar a los objetos ocultos parcial o totalmente por otros; además de que no captura su volumen. El trabajo de Criminisi (Criminisi *et al.* (2000)) es un ejemplo de este tipo de reconstrucciones.
- **Dos o más imágenes.** Varias imágenes de la escena tomada desde diferentes puntos de vista pueden completar la información que una cámara puede obtener. Bajo este planteamiento surge un nuevo problema; definir el menor número posible de imágenes que obtengan la mayor cantidad de información de la escena. A esto

se le conoce como el problema de la siguiente mejor vista (en inglés *Next Best View*) y es un problema abierto, estudiado por la robótica, la fotogrametría y la visión por computadora. En este trabajo de tesis no se aborda dicho problema, ya que se trabaja sólo con dos imágenes y se asume que éstas poseen una buena localización como para obtener suficiente información y así generar una buena reconstrucción.

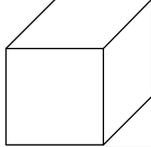
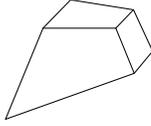
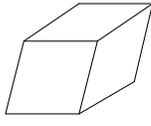
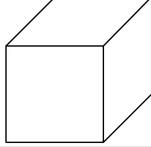
Que el resultado de una reconstrucción a partir de imágenes sea lo más parecido a la escena real, depende de la información inicial con la que se cuente. Así, cuando sólo se sabe qué pixel en una imagen corresponde a qué pixel en la otra, sin conocer nada sobre la localización y orientación de las cámaras, el resultado es una *reconstrucción proyectiva*. Si además de este conjunto de puntos se conoce la línea del horizonte de la escena dentro de las imágenes (llamada línea al infinito) el resultado es una *reconstrucción afín*. Finalmente, si además se conoce la localización, orientación y parámetros internos de las cámaras, así como algunos puntos de la escena, en el mundo real, se obtiene una *reconstrucción euclidiana*. Los tipos de reconstrucción se encuentran definidos en la tabla I.

En dicha tabla se muestra cómo se vería un cubo con los diferentes tipos de reconstrucción. En el capítulo 2 la información de la tabla se complementa.

Ahora bien, dependiendo del nivel de detalle que se obtenga, la reconstrucción puede ser dispersa o masiva:

- La reconstrucción dispersa obtiene las coordenadas tridimensionales de ciertas partes de la escena, cuya proyección en las imágenes se le conoce como puntos de interés, los cuales pueden ser bordes, puntos en esquinas u otros que se sabe

Tabla I: Tipos de reconstrucción, según la información inicial.

		Ejemplo
<i>Reconstrucción</i>	<i>Información</i>	
Proyectiva	- Puntos Correspondientes	
Afín	- Plano al infinito	
Euclidiana	- Calibración - Puntos de Control	

que delimitan un objeto. La reconstrucción dispersa se utiliza en aplicaciones que necesitan conocer su entorno rápidamente y sin mayor detalle, por ejemplo las aplicaciones en tiempo real, como la navegación de robots móviles.

- La reconstrucción masiva implica obtener la totalidad de los puntos proyectados en las imágenes de cada objeto de la escena. Principalmente se utiliza para aplicaciones relacionadas con la graficación, realidad virtual y cualquier otra cuyo objetivo sea modelar digitalmente de manera realista una escena del mundo.

I.4 Alcances de esta tesis.

Una vez expuesta la naturaleza del problema, se procede a definir los alcances y limitaciones del presente trabajo en forma de objetivos generales y específicos.

Se ha decidido por los sensores pasivos sobre los activos porque presentan una mayor robustez a las condiciones de la escena (intensidad de luz, etc), además dentro del grupo evovisión, al cual pertenece este trabajo, se cuenta con trabajos previos con cámaras; como son: algoritmos de calibración de cámaras y de correspondencia (Ramirez (2000); Pérez (2004); Hernández (2002)). Los cuales son antecedentes necesarios para este tipo de reconstrucción.

La reconstrucción que se hará en el presente trabajo es a partir de dos imágenes (par estéreo). El problema de la siguiente mejor vista, definido más arriba, no se aborda en este trabajo. Para obtener la localización y medidas reales de los objetos que la componen, se pretende realizar una reconstrucción euclidiana a partir del conocimiento de la localización y orientación de las cámaras que tomaron dichas imágenes, a trabajar con este conocimiento, se dice que se está en un contexto calibrado.

Para llevar a cabo la reconstrucción euclidiana, se plantea un enfoque inspirado en los procesos que ocurren en la naturaleza. Dentro de la visión por computadora existe una corriente de trabajo que ha modelado problemas complejos basándose en la similitud de éstos con algunos procesos vistos en la naturaleza.

La pretensión aquí es desarrollar un enfoque biológicamente inspirado, diferente al tradicional, que es un enfoque geométrico-matemático, el cual ha sido ampliamente estudiado y ha mostrado sus ventajas y desventajas. La razón de explorar un nuevo enfoque es la de encontrar formas diferentes de abordar el problema y complementar

puntos de vista a problemas que no han sido resueltos.

Habiendo explicado el problema a resolver y las motivaciones, se procede a definir los objetivos:

I.4.1 Objetivo General.

Dadas dos imágenes de una escena adquiridas mediante cámaras, diseñar un método basado en un enfoque bio-inspirado, que recupere las dimensiones de la escena, e implementar un sistema computacional que lo avale.

I.4.2 Objetivos Específicos.

Para alcanzar este objetivo, se han identificado varias etapas intermedias, las cuales se presentan como objetivos específicos que, al alcanzarse, gradualmente estarán contribuyendo al cumplimiento del objetivo general:

- *Conocer y analizar el enfoque geométrico tradicional, que lleva a cabo la reconstrucción tridimensional. Es necesario conocer el problema y la forma en que se ha abordado hasta el momento, conocer lo que se ha hecho y no se ha hecho, en base a esto proponer el nuevo enfoque.*
- *Encontrar un proceso natural que sirva como paradigma para proponer un método que aborde el problema de reconstrucción tridimensional. Es necesario encontrar dentro de la naturaleza un proceso cuya esencia sea similar al problema que se está abordando. Y así poder enfrentar con los recursos de uno, las dificultades del otro.*

- *Desarrollar un programa computacional que implemente ambos métodos.* Con el fin de comprobar el nuevo enfoque, se debe desarrollar un programa.
- *Comparar ambos métodos cualitativa y cuantitativamente.* La etapa final comprende la experimentación y la comparación entre ambos enfoques. El nuevo método debe arrojar resultados por lo menos comparables al anterior.

I.5 Organización de esta tesis.

Esta tesis se puede dividir en dos grandes partes. La primera explica a detalle el enfoque geométrico-matemático. Bajo este enfoque, cada imagen tiene asociada una matriz de parámetros (llamada matriz de proyección) de la cámara con la que fue tomada, \mathbf{M}_1 y \mathbf{M}_2 para imagen izquierda y derecha respectivamente. Un punto en el espacio tridimensional P se proyecta en la imagen izquierda como $p_1 = \mathbf{M}_1 P$ y en la derecha como $p_2 = \mathbf{M}_2 P$. A los puntos p_1 y p_2 se les llama *puntos correspondientes* porque representan al mismo punto del espacio. Basándose en esto, se han identificado tres cuestiones que deben ser resueltas para realizar la reconstrucción:

- **La geometría de la cámara.** ¿Cómo definir las matrices de proyección \mathbf{M}_1 y \mathbf{M}_2 para las dos imágenes? Esta cuestión se analiza en el capítulo II, junto con los conceptos geométricos en que la visión por computadora tiene su base.
- **La geometría epipolar.** ¿Dado un punto en la imagen izquierda p_1 , cómo restringir la posición del punto correspondiente p_2 en la imagen derecha? Esta pregunta depende de la llamada matriz fundamental, la cual se detalla en el capítulo III.

- **La geometría de la escena (reconstrucción).** ¿Dado un conjunto de puntos correspondientes $p_1 \longleftrightarrow p_2$ y las matrices de proyección M_1 y M_2 , cuál es la posición de P en el espacio tridimensional? Esta pregunta tiene que ver con la forma en que el enfoque geométrico-matemático aborda el problema de reconstrucción tridimensional a partir de dos imágenes: la triangulación, de la cual se habla en el capítulo IV.

La segunda parte comprende el desarrollo del nuevo enfoque bio-inspirado. En él se justifica el proceso natural tomado como base para desarrollar el nuevo enfoque, hasta su implementación y validación de resultados. El nuevo enfoque es radicalmente distinto al anterior, de hecho, exige que el planteamiento del problema sea el inverso del geométrico: en lugar de construir un modelo tridimensional a partir de las imágenes, la idea es buscar el modelo que sea más coherente con la información de las imágenes.

- En el capítulo V se da una introducción de la herramienta de programación que es pieza fundamental para implementar el nuevo enfoque: los algoritmos evolutivos.
- En el capítulo VI es donde se diseña la teoría y la implementación de un programa computacional, denominado *algoritmo de búsqueda basado en el comportamiento de las abejas*.

El capítulo 7 contiene la experimentación diseñada para comparar y evaluar el nuevo enfoque con el tradicional. Y finalmente, en el capítulo 8 se comentan las conclusiones a las que se ha llegado con este trabajo y se plantean ideas de lo que se puede realizar en un futuro sobre la misma línea de investigación.

Parte I
Enfoque Geométrico

Capítulo II

La geometría aplicada a la visión por computadora.

Este capítulo presenta los conceptos geométricos principales necesarios para comprender el material de este trabajo. Además, contiene el modelado geométrico del proceso de adquisición de la imagen por una cámara.

La geometría euclidiana es la más conocida, pues se utiliza para describir el mundo tridimensional. Según la geometría euclidiana, los objetos tienen ciertas dimensiones, las líneas que intersectan entre sí forman ángulos entre ellas y a dos líneas que caen sobre un mismo plano y nunca intersectan se les llama paralelas. Además, estas propiedades no cambian cuando se les aplica una *transformación* euclidiana (rotación y traslación). Sin embargo, al analizar el proceso de adquirir la imagen a través de una cámara, aplicar la geometría euclidiana no es suficiente, ya que las dimensiones y los ángulos pueden cambiar y las líneas paralelas pueden intersectarse.

La euclidiana, es de hecho, un subconjunto de lo que se conoce como *geometría proyectiva*. Además, hay entre ellas un nivel de de abstracción intermedio: la geometría *afín*. La geometría proyectiva modela muy bien el proceso de adquisición de una imagen, pues incluye una mayor diversidad de tipos de transformaciones que sólo la rotación y traslación; aunque menos propiedades que describen a un objeto permanecen sin cambiar (cambian dimensiones, ángulos y paralelismo). Sin embargo esto no representa

problema, pues la geometría proyectiva ha demostrado modelar fielmente la forma en que una cámara adquiere la imagen.

II.1 Geometría proyectiva.

La geometría proyectiva es aquella que trata las propiedades que se conservan al proyectar un objeto geométrico de una dimensión n , a una dimensión inferior $n-1$. Aplica para cualquier número de dimensiones, a los cuales se les llama *espacios proyectivos* (Hartley y Zisserman (2004)). Por ejemplo, la línea proyectiva se denota por \mathcal{P}^1 ; que es análoga al mundo euclidiano de una dimensión; el plano proyectivo \mathcal{P}^2 corresponde al plano euclidiano; y el espacio proyectivo \mathcal{P}^3 está relacionado con el espacio euclidiano tridimensional. Para la visión por computadora, es el plano proyectivo el más útil para los fines que persigue, ya que ofrece una manera simple de modelar la proyección del mundo tridimensional en el plano de la imagen.

II.1.1 El plano proyectivo.

Para representar el punto (x,y) del plano euclidiano en el plano proyectivo, se debe agregar una tercera coordenada, la cual representa un escalamiento.¹ El escalamiento no es muy importante para el resultado final, pues está presente en todo el plano, por lo general, el escalamiento se representa con la unidad: $(x, y, 1)$. Así, el punto $(x, y, 1)$ es el mismo punto que $(\alpha x, \alpha y, \alpha)$ para cualquier α no cero. En otras palabras

$$(X, Y, W) = (\alpha X, \alpha Y, \alpha W); \text{ para cualquier } \alpha \neq 0$$

¹En general, un punto en el espacio euclidiano n -dimensional se representa como un vector de $n+1$ componentes en el espacio proyectivo de la misma dimensión.

El punto $(0, 0, 0)$ no está permitido. Como la escala no es importante, a las coordenadas (X, Y, W) se les llama *coordenadas homogéneas* de un punto. Un punto \mathbf{p} es un vector de tres componentes, según la notación convencional de la visión por computadora, este punto puede escribirse como $p, (X, Y, W)^T$ o $\begin{pmatrix} X \\ Y \\ W \end{pmatrix}$. Esta notación se utilizará a lo largo de todo este documento.

Una línea en el plano euclidiano

$$ax + by + c = 0$$

también tiene su equivalente en el plano proyectivo. Usando el hecho de que la ecuación no se verá afectada por el escalamiento se puede llegar a:

$$\begin{aligned} aX + bY + cW &= 0, \\ u^T p &= p^T u = 0 \end{aligned}$$

donde $\mathbf{u}=(a, b, c)^T$ es la línea y $\mathbf{p}=(X, Y, W)^T$ es el punto en la línea. Como puede apreciarse, la línea y el punto tienen la misma representación en el plano proyectivo. Los parámetros de la línea son fáciles de interpretar: $-a/b$ es la pendiente, $-c/a$ es la intersección con el eje X y $-c/b$ es la intersección con el eje Y.

En la ecuación de arriba se aprecia que el punto \mathbf{p} y la línea \mathbf{u} se intercambian, dando el mismo resultado. Esto obedece a un principio de la geometría proyectiva llamado *principio de dualidad*, el cual se puede enunciar como sigue (Hartley y Zisserman (2004)):

Para cualquier teorema aplicable al plano proyectivo existe un teorema dual; el cual se deriva intercambiando los roles de los puntos y las líneas en

el teorema original.

Este principio es muy útil, ya que al probar un teorema aplicable a una línea, se puede crear su dual aplicable a un punto y no es necesario probarlo; pues la prueba sería la misma, sólo considerando a la línea como un punto.

Para transformar un punto en el plano proyectivo de regreso a coordenadas euclidianas, simplemente se divide la primera y la segunda entre la tercer coordenada: $(x, y) = (X/W, Y/W)$. La tercer coordenada puede ser cualquier número, incluso cero. Todo punto que posea como tercera coordenada un cero, al transformarlo será colocado en el origen $(0,0)$ del euclidiano. Así, el plano proyectivo cuenta con puntos que el euclidiano no tiene; a estos se les llama *puntos ideales o puntos al infinito*. Hay un punto ideal asociado con cada dirección en el plano; por ejemplo los puntos $(1, 0, 0)$ y $(0, 1, 0)$ están asociados con la dirección horizontal y vertical respectivamente. En \mathcal{P}^2 a los puntos ideales se les trata como a cualquier otro punto. Todos los puntos ideales caen en una línea, llamada *línea ideal o línea al infinito*, a la cual se le trata como a cualquier otra línea. La línea ideal se representa como $(0, 0, 1)$.

Una manera de representar el plano proyectivo \mathcal{P}^2 es como un conjunto de vectores en \mathcal{R}^3 (ya que \mathcal{P}^2 se conforma de vectores de tres coordenadas). El vector $k(x_1, x_2, x_3)^T$, cuando k varía forma un *rayo* que pasa por el origen (un rayo es un vector cuya magnitud varía de acuerdo a k). Este rayo representa a un punto en \mathcal{P}^2 . Por su parte, las líneas en \mathcal{P}^2 se representan por planos que parten del origen. Los puntos y líneas así representados se obtienen al intersectar los rayos y planos respectivamente con el plano $x_3 = 1$. Como se ve en la figura 3, los rayos que representan puntos ideales forman parte del plano $x_1 - x_2$, que representa la línea al infinito.

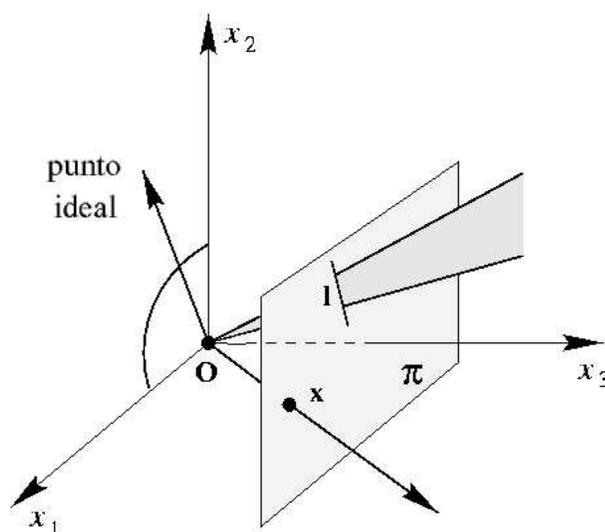


Figura 3: Un modelo del plano proyectivo. Los puntos y las líneas de \mathcal{P}^2 se representan por la intersección de rayos y planos en \mathcal{R}^3 con el plano $x_3 = 1$.

Supóngase ahora que se quiere encontrar el punto donde intersectan dos líneas. Por álgebra elemental, las dos líneas $u_1 = (a_1, b_1, c_1)^T$ y $u_2 = (a_2, b_2, c_2)^T$ se intersectan en el punto $p = (b_1c_2 - b_2c_1, a_2c_1 - a_1c_2, a_1b_2 - a_2b_1)$. Esta fórmula es más fácil de escribir como un producto cruz: $\mathbf{p} = \mathbf{u}_1 \times \mathbf{u}_2$. Si las dos líneas son paralelas, la tercera coordenada se hace cero, porque las pendientes serían iguales: $-a_1/b_1 = -a_2/b_2$. Aplicando el principio de dualidad, se puede afirmar que dados dos puntos \mathbf{p}_1 y \mathbf{p}_2 , la ecuación de la línea que pasa a través de ellos es $\mathbf{u} = \mathbf{p}_1 \times \mathbf{p}_2$.

II.2 Geometría afín.

El plano proyectivo es homogéneo, no hay una coordenada que reciba un trato preferencial sobre otra. No hay concepto de líneas paralelas porque éstas se cruzan en el

infinito, no pueden diferenciarse los puntos que caen en la línea al infinito porque no hay distinción al manejar los puntos. Si se quisiera hacer la diferencia entre aquellos puntos que caen en la línea al infinito y aquellos que no, se debe restringir el plano proyectivo.

Piénsese en una fotografía tomada de una región plana y despejada. Debido a la curvatura del mundo, se apreciaría una línea que separa el suelo del cielo: *la línea del horizonte*. Las líneas paralelas, como las vías de un tren, aparecerían como líneas que se cruzan en la línea del horizonte. Así, teniendo como referencia la línea del horizonte se puede diferenciar entre cualquier línea y líneas que son paralelas.

Este es el plano afín: el plano proyectivo mas una línea de referencia. Traduciéndola a términos geométricos, ésta es la línea al infinito $(0, 0, 1)$, por ser donde se intesectan las paralelas. Así, al distinguir una línea en el plano proyectivo, se introduce el concepto de paralelismo con él, la geometría afín. Ésta es entonces un subconjunto de la geometría proyectiva, y como tal, es más específica y menos flexible.

Gracias al paralelismo hay otras propiedades que forman parte de la geometría afín. Por ejemplo, ahora es posible definir igualdades de intervalos entre dos puntos que caen en líneas paralelas, es decir, se consigue la noción de *distancia relativa*. Por ejemplo, siendo A, B, C, D puntos, de las líneas paralelas AB y CD , éstas tendrán la misma longitud si las líneas formadas por AC y BD también son paralelas.

II.3 Geometría euclidiana.

La geometría afín es producto de una restricción de la geometría proyectiva por la línea al infinito, del mismo modo, la geometría euclidiana surge al aplicar una restricción a

la geometría afín.

En el plano afín es posible distinguir entre líneas paralelas y aquellas que no lo son, sin embargo, el círculo no es un concepto válido, ya que el estiramiento del plano que conserva la línea al infinito lo convertiría en elipse. Así, la geometría afín no distingue entre círculos y elipses. En la geometría euclidiana, sin embargo, sí se distinguen; además en ella se cumple que dos elipses al intersectarse, se cruzan hasta en cuatro puntos; y que dos círculos no pueden hacerlo en más de dos puntos. Algebraicamente, se están intersectando dos curvas de segundo grado, o dicho de otra forma, se está resolviendo un sistema de dos ecuaciones cuadráticas. De este sistema se esperarían cuatro soluciones. La pregunta es, qué hace que un círculo intersecte sólo en dos puntos. La respuesta es que existen otras dos soluciones además de las reales; sólo que son complejas.

La ecuación analítica del círculo en coordenadas homogéneas (x, y, w) está dada por

$$(x - aw)^2 + (y - bw)^2 = r^2w^2$$

Este es un círculo con centro en las coordenadas $(a, b, 1)^T$. Según Hartley y Zisserman (2004), los puntos $(1, \pm i, 0)$ caen en todos los círculos del plano euclidiano, es decir, cada círculo pasa sobre estos puntos, y además, éstos caen en la intersección de dos círculos cualquiera. Ya que su coordenada final es cero, ambos puntos deben pertenecer a la línea al infinito. Por obvias razones, se les llama *puntos circulares* del plano.

Nótese que a pesar de ser complejos, los puntos circulares satisfacen un par de ecuaciones reales: $x^2 + y^2 = 0$ y $w = 0$. Esta observación da una pista de cómo la geometría euclidiana se deriva de la proyectiva, diferenciando primero la línea al infinito,

y consecuentemente, los puntos circulares que caen en ella (Hartley y Zisserman (2004)).

Ahora bien, un círculo puede definirse como una cónica (una curva definida por una ecuación de segundo grado) que pasa por dos puntos circulares. Una cónica está definida por cinco puntos arbitrarios en el plano, como se puede apreciar si se cuenta el número de coeficientes en la ecuación general de la cónica

$$ax^2 + bxy + cy^2 + dxw + eyw + fw^2$$

Generalizando el término de círculo, los puntos circulares que se deben definir pertenecen a una cónica que atraviesa la línea al infinito, llamada *cónica absoluta*. De esta manera, recapitulando, se puede decir que la geometría euclidiana es un subconjunto de la geometría proyectiva, si se le restringe por la definición de la línea al infinito y de los puntos que forman parte de la cónica absoluta que intersecta dicha línea.

II.4 Jerarquía de las transformaciones geométricas.

Según Felix Klein (Hartley y Zisserman (2004)), la geometría es el estudio de propiedades invariantes bajo un grupo de transformaciones. De aquí se puede derivar que la geometría proyectiva bidimensional, es el estudio de las propiedades del plano proyectivo \mathcal{P}^2 que son invariantes bajo un grupo de transformaciones conocidas como *proyectividades*. Una proyectividad es el mapeo invertible h de \mathcal{P}^2 a él mismo, tal que tres puntos x_1, x_2 y x_3 caen en una misma línea si y solo si $h(x_1), h(x_2)$ y $h(x_3)$ también lo hacen. Se considera invariante a aquellas propiedades que no cambian al aplicarles alguna transformación en particular.

El grupo de proyectividades posee subgrupos de transformaciones que se distinguen

unos de otros por especializarse en ciertas invariantes, los cuales están relacionados con los subconjuntos de geometrías que forman la geometría proyectiva. Estos subgrupos incluyen al *grupo afín*, el cual consiste en matrices cuyo último renglón es $(0, 0, 1)$ y el *grupo euclidiano*; el cual es un subgrupo del afín el cual, tiene además, la submatriz superior izquierda de 2×2 ortogonal. También puede identificarse el *grupo euclidiano orientado* cuya submatriz de 2×2 posee determinante de 1.

A continuación se presentan estas transformaciones comenzando por las más especializadas, avanzando progresivamente hasta las más generales, y así se define una jerarquía de transformaciones.

II.4.1 Isometrías.

Las isometrías, del grupo euclidiano orientado, son transformaciones que preservan la distancia euclidiana, como su nombre lo dice *iso*=igual, *metría*=medida; y se representa como

$$\begin{pmatrix} x' \\ y' \\ 1 \end{pmatrix} = \begin{pmatrix} \epsilon \cos\theta & -\text{sen}\theta & t_x \\ \epsilon \text{sen}\theta & \cos\theta & t_y \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}$$

Donde $\epsilon = \pm 1$. Si $\epsilon = 1$ entonces la transformación preserva la orientación original; de lo contrario, el resultado tiene la orientación inversa, es decir, se vería como el reflejo en un espejo del original.

Las transformaciones euclidianas modelan el movimiento de un cuerpo rígido, pues preservan las dimensiones de éste. Ésta es la aplicación más importante que desempeñan. En forma de bloque las isometrías pueden escribirse en forma concisa

$$x' = H_E x = \begin{pmatrix} R & t \\ 0^T & 1 \end{pmatrix} x$$

donde \mathbf{R} es una matriz de rotación de 2×2 (matriz ortogonal, tal que $R^T R = I$), \mathbf{t} es un 2-vector (vector con dos componentes) de translación y 0^T representa un 2-vector renglón de ceros. Los casos especiales son cuando $\mathbf{t}=0$, una rotación pura y cuando $\mathbf{R} = I$, conocida como translación pura. Estas transformaciones cuentan con tres grados de libertad (3 dof); esto es, los parámetros que deben especificarse para definir la transformación. Y son uno por el ángulo de rotación y dos por el movimiento de translación.

Invariantes.

En este grupo, las invariantes son la longitud (distancia entre dos puntos), el ángulo (entre dos líneas) y el área.

II.4.2 Similaridades.

Una similaridad, del grupo euclidiano, es una isometría, más un escalamiento isotrópico (uniforme en todas sus direcciones). Su representación matricial es la siguiente

$$\begin{pmatrix} x' \\ y' \\ 1 \end{pmatrix} = \begin{pmatrix} s \cos \theta & -s \sin \theta & t_x \\ s \sin \theta & s \cos \theta & t_y \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}$$

En forma de bloque queda

$$x' = H_S x = \begin{pmatrix} sR & t \\ 0^T & 1 \end{pmatrix} x$$

Donde s representa el escalamiento uniforme. Las similaridades preservan la forma de los objetos en los que se aplica la transformación, aunque no el tamaño. Poseen cuatro grados de libertad: los tres de las isometrías más el escalado.

Invariantes.

Los ángulos no se ven afectados por la rotación, traslación y escalamiento; consecuentemente, las líneas paralelas siguen siendo paralelas después de aplicar la transformación, pero la distancia entre dos puntos no es una invariante. Sin embargo, la razón d_i/d_j entre dos distancias cualquiera permanece invariante, pues el factor de escalamiento se cancela al dividirse. Lo mismo ocurre con la razón entre dos áreas a_i/a_j .

II.4.3 Afinidades.

Una transformación afín o afinidad es una transformación lineal seguida de una traslación. Su forma matricial es

$$\begin{pmatrix} x' \\ y' \\ 1 \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & t_x \\ a_{21} & a_{22} & t_y \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}$$

O en forma de bloque

$$x' = H_A x = \begin{pmatrix} A & t \\ 0^T & 1 \end{pmatrix} x$$

donde \mathbf{A} es una matriz no singular, y se compone de dos rotaciones y un escalamiento no isotrópico

$$A = R(\theta)R(-\phi)DR(\phi) \tag{1}$$

donde $R(\theta)$ y $R(-\phi)$ son rotaciones por θ y ϕ respectivamente, y D es una matriz diagonal

$$D = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$$

En pocas palabras, se puede decir que \mathbf{A} es la concatenación de una rotación (por ϕ) un escalamiento en diferentes direcciones por λ_1 y λ_2 , una rotación de regreso (por $-\phi$) y finalmente otra rotación (por θ). El ángulo ϕ especifica la dirección del escalamiento, y la proporción entre λ_1 y λ_2 . Véase la figura 4.

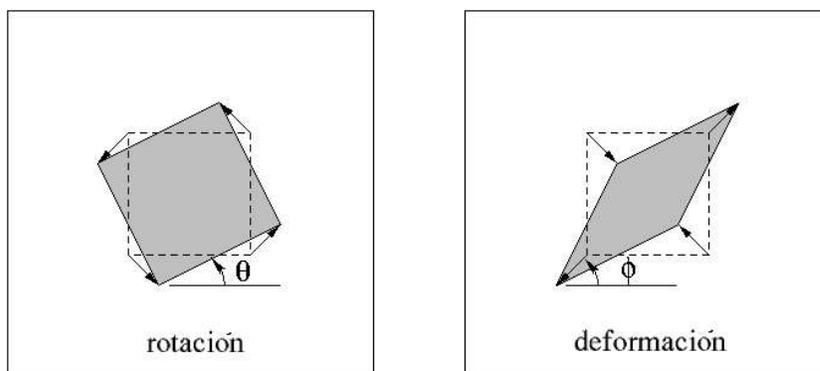


Figura 4: Transformación Afín. a) Rotación por θ . b) Deformación causada por el escalamiento no isotrópico.

Esta transformación tiene seis grados de libertad, correspondientes a sus seis elementos: dos ángulos, dos escalamientos y dos traslaciones.

Invariantes.

Debido al escalamiento no isotrópico, las invariantes de la similaridad, como los ángulos, la razón entre distancias y áreas no se conservan en la afinidad. Tres de las invariantes

más importantes para esta transformación son:

- **Líneas paralelas.** Dos líneas paralelas intersectan en un punto al infinito. Bajo una transformación este punto se mapea a otro punto al infinito; en consecuencia, dos líneas paralelas se mapean a dos líneas paralelas.
- **Razón de distancias entre segmentos de líneas paralelas.** El escalamiento de un segmento de línea depende del ángulo entre la dirección de la línea y la dirección del escalamiento. Este escalamiento es común a todas las líneas con la misma dirección; así, al aplicar la transformación la razón entre estas líneas cancelaría el escalamiento, en un caso parecido al escalamiento de la similaridad.
- **Razón entre áreas.** Esta invariante viene de la ecuación (1). Las rotaciones y las traslaciones no afectan un área, los escalamientos si lo hacen. Al aplicar la transformación afín, el área queda escalada por el producto $\lambda_1\lambda_2$, el cual es igual al determinante de \mathbf{A} ($\det \mathbf{A}$). Así es que el escalamiento se cancela al aplicar la razón a_i/a_j

II.4.4 Proyectividades.

La transformación proyectiva, también llamada *colineación*, o bien *homografía*, es una transformación lineal de coordenadas homogéneas representada por una matriz no singular de 3×3

$$\begin{pmatrix} x'_1 \\ x'_2 \\ x'_3 \end{pmatrix} = \begin{pmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$

Ésta pertenece al grupo que generaliza y engloba a todas las demás transformaciones. En forma de bloque la transformación es

$$x' = H_P x = \begin{pmatrix} A & t \\ \mathbf{v}^T & v \end{pmatrix} x$$

Donde $\mathbf{v} = (v_1, v_2)^T$ es un vector renglón. La matriz H_P puede ser cambiada multiplicándola por un factor escalar arbitrario diferente de cero, sin alterar la transformación proyectiva. Esto se debe a que es una matriz homogénea, y sólo las razones entre los elementos de la matriz tienen sentido. En consecuencia, de los nueve elementos de la matriz, hay ocho razones diferentes, y por lo tanto, son ocho los grados de libertad que tienen estas transformaciones.

Como ejemplo, véase la figura 5, que utiliza el modelo del plano proyectivo que se vió anteriormente, en la figura 3. La proyección de un rayo que pasa por un punto común (el origen del sistema coordenado) define el mapeo de un punto que pertenece a un plano, en otro plano. De la misma manera, se puede deducir que una línea en el plano π se mapea en una línea del plano π' si se considerara un plano que pasa por el origen e intersectara ambos planos.

Invariantes.

La invariante fundamental en las transformaciones proyectivas es la *razón doble* (*cross-ratio* en inglés) de cuatro puntos colineales. La razón de distancias en una línea es invariante para una afinidad, pero no para una proyectividad. La razón doble o razón de razones se define como

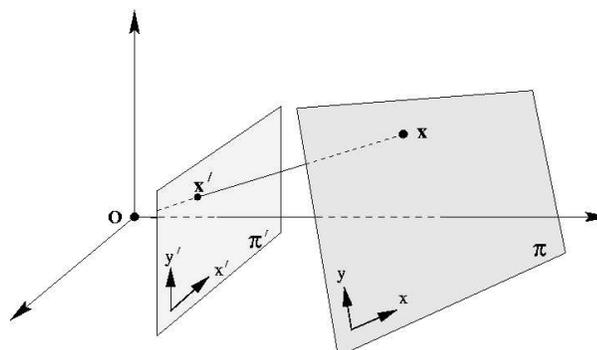


Figura 5: Transformación proyectiva entre dos planos.

$$Cross(x_1, x_2, x_3, x_4) = \frac{|x_1x_2||x_3x_4|}{|x_1x_3||x_2x_4|}$$

Esta definición también es válida para el caso en que uno de los puntos sea un punto al infinito.

II.4.5 Comparación entre transformaciones.

Las afinidades (6 dof), ocupan un lugar intermedio entre similaridades (4 dof) y proyectividades (8 dof). Las primeras generalizan a las similaridades en que los ángulos no se conservan, es por eso que los objetos se deforman. Por otro lado, el escalamiento del área de un objeto es la misma al aplicar la afinidad, y la orientación de una línea depende de su orientación inicial, no de su posición en el plano. En contraste, para una proyectividad, el escalamiento de un área varía con la posición en la que está, por ejemplo, un cuadrado que está mas lejano en el plano proyectivo aparece más pequeño que uno que está más cerca. Véase la figura 6, aplicando una transformación simliar a una imagen, en a), los círculos y los cuadrados aparecen como lo que son. En b),

aplicando una afinidad, los círculos aparecen como elipses, las líneas ortogonales no lo son más. Sin embargo, los lados de los cuadrados, que son paralelos en el mundo real, permanecen paralelos. Al aplicar una proyectividad en c), las líneas paralelas convergen en un punto fuera de la imagen. Las figuras cercanas a la cámara son más grandes que las más alejadas.

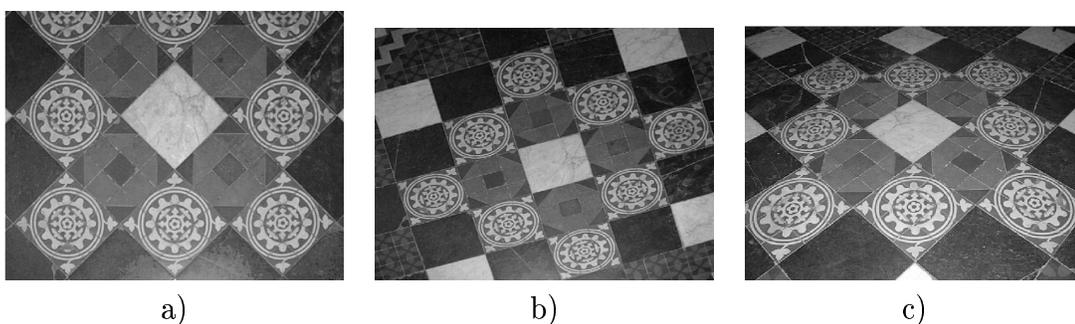


Figura 6: Comparación entre transformaciones. a) Similaridad. b) Afinidad. c) Proyectividad.

La principal diferencia entre transformaciones afines y proyectivas es que el vector \mathbf{v} no es nulo en la segunda. Este vector es responsable de los efectos no lineales de la proyectividad. Esto se demuestra al mapear un punto al infinito $(x_1, x_2, 0)^T$. Primero con una transformación afín

$$\begin{pmatrix} A & t \\ 0^T & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ 0 \end{pmatrix} = \begin{pmatrix} A \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \\ 0 \end{pmatrix}$$

Ahora aplicando la proyectividad

$$\begin{pmatrix} A & t \\ \mathbf{v}^T & v \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ 0 \end{pmatrix} = \begin{pmatrix} A \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \\ v_1 x_1 + v_2 x_2 \end{pmatrix}$$

En el primer caso el punto al infinito sigue siendo un punto al infinito. En el segundo el punto se transforma en un punto finito. Esto es lo que permite a las transformaciones proyectivas modelar los puntos de fuga, que no es otra cosa más que la representación de los puntos al infinito sobre una imagen, pintura o dibujo.

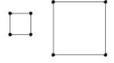
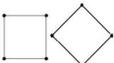
En la tabla II se muestra, en resumen, las características de los grupos de transformaciones. En la columna *Matriz* se muestra la representación matricial de cada transformación. La columna *Invariantes* tiene las propiedades invariantes de cada uno; y en *Distorsión* se muestra un ejemplo de los efectos típicos al aplicar estas transformaciones sobre un cuadrado. Una transformación en un renglón más alto, engloba a las de renglones bajos y por lo tanto es capaz de producir los mismos efectos que ellas.

Esta tabla representa un complemento de información de la tabla I del capítulo 1. Comparando ambas tablas, se puede intuir que existe una relación entre la jerarquía de transformaciones y los tipos de reconstrucción que se pueden obtener. Aunque en este apartado se han presentado las transformaciones en el plano proyectivo (2D), éstas aplican en el espacio proyectivo (3D) y pueden extenderse a éste (Hartley y Zisserman (2004)).

II.5 Modelo geométrico de una cámara.

La formación de una imagen se lleva a cabo por un sensor que registra la luz reflejada con los objetos que se encuentran en el rango de visión del sensor. El sensor que se

Tabla II: Resumen de la jerarquía de transformaciones.

			Distorsión
<i>Transformación</i>	<i>Matriz</i>	<i>Propiedades Invariantes</i>	
Proyectiva (8 dof)	$\begin{pmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{pmatrix}$	- Razón doble - Colinearidad	
Afin (6 dof)	$\begin{pmatrix} a_{11} & a_{12} & t_x \\ a_{21} & a_{22} & t_y \\ 0 & 0 & 1 \end{pmatrix}$	-Razón de áreas y líneas - Paralelismo	
Similar (4 dof)	$\begin{pmatrix} sr_{11} & sr_{12} & t_x \\ sr_{21} & sr_{22} & t_y \\ 0 & 0 & 1 \end{pmatrix}$	- Razón de líneas - Ángulos	
Isometría (3 dof)	$\begin{pmatrix} r_{11} & r_{12} & t_x \\ r_{21} & r_{22} & t_y \\ 0 & 0 & 1 \end{pmatrix}$	- Distancia y áreas absolutas - Ángulos	

utiliza a lo largo de este trabajo es una cámara CCD (Charge-Coupled Device) con la que se obtienen tablas de 2 dimensiones cuadrículadas donde a cada elemento se le llama píxel (abreviación de *picture element*). Estas tablas, a las que comúnmente se les llama imágenes, contienen una señal resultante de un proceso complejo de adquisición de datos donde intervienen 3 aspectos: el óptico, el electrónico, y geométrico. Aquí se explicará el aspecto geométrico de la formación de una imagen.

Una cámara utiliza el mismo modelo de adquisición de la imagen que un ojo humano; donde lo fundamental es la proyección de un punto sobre un plano. En una aproximación a primer orden, la visión por computadora utiliza el modelo de la cámara de ojo de aguja (en inglés *pin-hole model*) en la que la imagen resulta de la proyección de los puntos de la escena a través de un solo punto sobre el plano de la imagen.

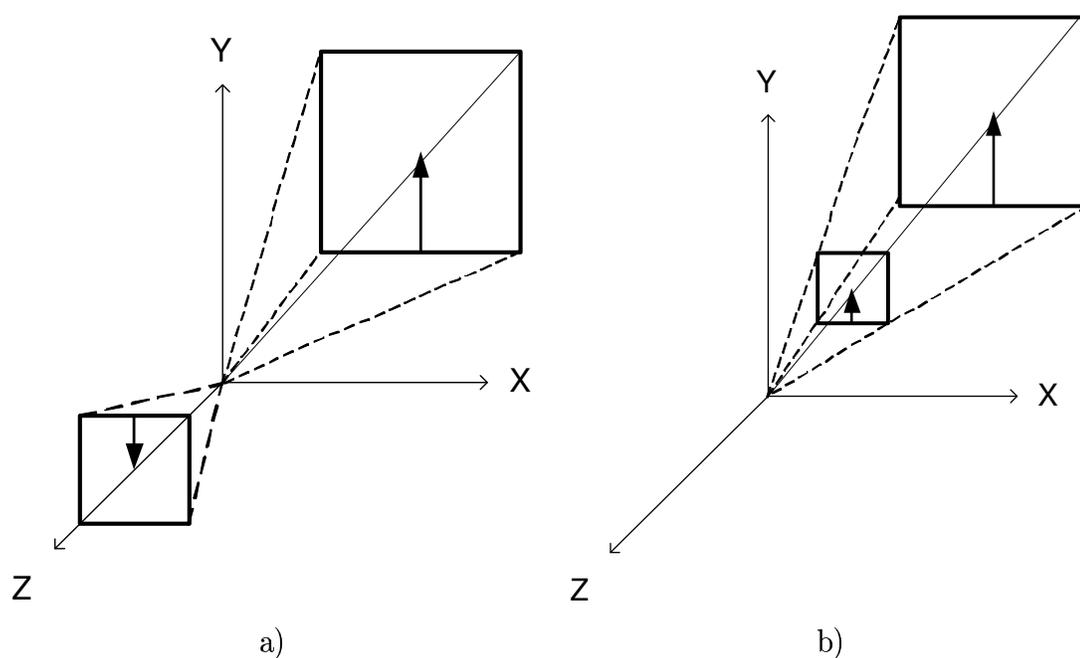


Figura 7: Modelo de la cámara de ojo de aguja. a)Original. b) Corrección en la orientación de la imagen

En a) de la figura 7 se ilustra el modelo de la cámara de ojo de aguja; en ella la imagen se forma detrás del punto de proyección y además parece invertida. Sin embargo, se puede modificar la geometría de manera que el plano de proyección aparezca frente al

punto de proyección como se aprecia en b) de la misma figura; esto permite que la imagen se forme perfectamente orientada y la modificación no afecta al modelo.

El modelo geométrico asociado a la adquisición de una imagen por una cámara, posee cierto número de parámetros que se deben estimar para *calibrar* la cámara, a los cuales se les conoce como parámetros intrínsecos y extrínsecos. Los primeros son las características internas de la cámara (distancia focal, tipo de pixel, etc.); mientras que los segundos son transformaciones que relacionan el sistema coordenado de la cámara con el de la escena (transformaciones de rotación y traslación).

Para comprender cómo una cámara adquiere la información puede verse la figura 8. En ella se esquematiza un punto \mathbf{P} en el espacio tridimensional, que se encuentra bajo un sistema de coordenadas de referencia de la escena (X_w, Y_w, Z_w) . El primer paso es conocer la transformación de este sistema de referencia al sistema de coordenadas de la cámara (X_c, Y_c, Z_c) , esta transformación consiste en una rotación \mathbf{R} y una traslación \mathbf{t} . Después viene la transformación del sistema coordenado de la cámara al sistema de la imagen (u,v) , durante este proceso se pierde una dimensión al proyectar \mathbf{P} del espacio (3D), al plano de la imagen en \mathbf{p} (2D); donde intervienen los parámetros intrínsecos de la cámara (I_C) . Ambas transformaciones pueden agruparse en una sola matriz \mathbf{M} , la cual representa la relación entre el punto 3D y el punto proyectado 2D: $\mathbf{p} = \mathbf{MP}$.

II.5.1 Proceso de adquisición de la imagen.

Durante el proceso de adquisición de la imagen, cada punto del espacio tridimensional se proyecta a la imagen; para pasar de las coordenadas tridimensionales del marco de referencia de la escena, expresadas en milímetros, a las coordenadas bidimensionales de

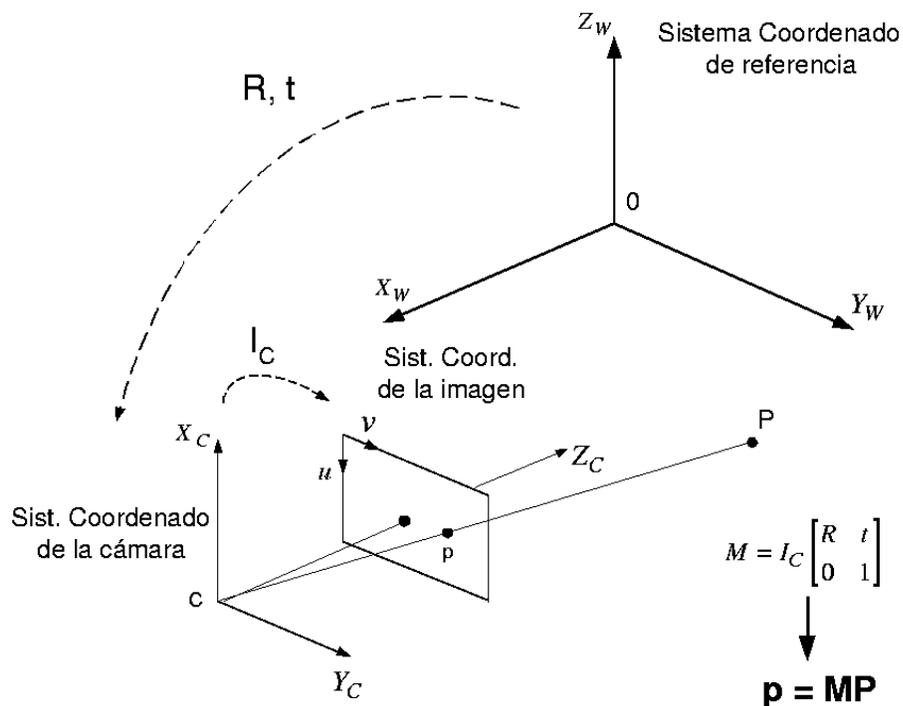


Figura 8: Modelo Geométrico de la adquisición de la imagen de una cámara.

la imagen, expresadas en píxeles, son necesarios tres pasos (Ramirez (2000)): la transformación de la escena a la cámara; la proyección 3D-2D y finalmente la transformación de la cámara a la imagen. A continuación se detalla cada uno de estos pasos.

Transformación Escena-Cámara.

El modelado de los parámetros extrínsecos se realiza mediante la transformación escena-cámara, la cual se compone de tres rotaciones y tres translaciones que transforman el punto de las coordenadas de la escena a las coordenadas de la cámara

$$A = \begin{pmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{32} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \end{pmatrix} \quad (2)$$

Proyección 3D - 2D.

Ahora a los puntos expresados en el marco de referencia de la cámara se les somete a una transformación proyectiva, que no es otra cosa más que proyectar el punto del espacio tridimensional sobre el plano de la imagen de la siguiente manera

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

Transformación Cámara - Imagen.

Para pasar a coordenadas expresadas en píxeles, las coordenadas obtenidas anteriormente se someten a una transformación afín del plano. Ésta se compone de los llamados parámetros intrínsecos, expresados en la siguiente matriz

$$C = \begin{pmatrix} \alpha_u & 0 & u_0 \\ 0 & \alpha_v & v_0 \\ 0 & 0 & 1 \end{pmatrix} \quad (3)$$

Donde α_u y α_v dan la distancia focal medida en píxeles y u_0, v_0 son las coordenadas del pixel principal, llamado origen de la imagen. Finalmente, el conjunto de transformaciones se puede expresar en una sola matriz; $\mathbf{M} = \mathbf{C} \mathbf{P} \mathbf{A}$; o bien

$$M = I_C A \quad (4)$$

Donde I_C representa la matriz de parámetros intrínsecos y \mathbf{A} representa la matriz de parámetros extrínsecos. De esta manera, la expresión que relaciona un punto del espacio tridimensional con su proyección en la imagen que se dio al principio de ese capítulo queda como sigue

$$\begin{pmatrix} s_u \\ s_v \\ s \end{pmatrix} = \begin{pmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & m_{34} \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \longrightarrow p = MP \quad (5)$$

Las coordenadas del píxel $p = (s_u, s_v, s)^T$ presentan un factor de escala con respecto a las dimensiones de la escena real. Como este factor de escala está presente en toda la imagen, las coordenadas de un píxel pueden escribirse como $p = (u, v, 1)^T$ al dividir todos los componentes del vector entre el tercero: $p = (s_u/s, s_v/s, s/s)^T$, siendo u y v las coordenadas del píxel en la imagen. Esta simplificación no afecta al modelo geométrico.

II.6 Calibración de la cámara.

Se pueden calcular los parámetros intrínsecos y extrínsecos en función de los coeficientes de la matriz \mathbf{M} (Ramirez (2000)). La matriz \mathbf{M} se puede descomponer en 3 vectores renglón y tres coeficientes:

$$\mathbf{M} = \begin{pmatrix} \mathbf{m}_1 & m_{14} \\ \mathbf{m}_2 & m_{24} \\ \mathbf{m}_3 & m_{34} \end{pmatrix},$$

donde $\mathbf{m}_i = (m_{i1}, m_{i2}, m_{i3})$. Así, tenemos que:

$$\left\{ \begin{array}{l} r_3 = m_3, \\ u_0 = m_1 \cdot m_3, \\ v_0 = m_2 \cdot m_3, \\ \alpha_u = -\|m_1 \wedge m_3\| \\ \alpha_v = \|m_2 \wedge m_3\| \\ r_1 = 1/\alpha_u(m_1 - u_0 m_3), \\ r_2 = 1/\alpha_v(m_2 - v_0 m_3), \\ t_x = 1/\alpha_u(m_{14} - u_0 m_{34}), \\ t_y = 1/\alpha_v(m_{14} - v_0 m_{34}), \\ t_z = m_{34} \end{array} \right. \quad (6)$$

donde el operador (\cdot) representa el producto punto, y el operador (\wedge) representa el producto cruz entre dos vectores. Para determinar los parámetros de la cámara, es necesario tener un objeto en la escena que funcione como referencia, el cual se le conoce como mira de calibración. Esta mira se compone de un conjunto de puntos cuyas coordenadas son perfectamente conocidas en el sistema coordenado de la escena; en el par de imágenes de la figura 16 la mira de calibración es el objeto cuadrículado. Dependiendo del número de puntos y de la precisión de la mira de calibración, los parámetros serán más o menos fidedignos. Derivándolo de (5) cada uno de estos puntos genera dos ecuaciones de la siguiente manera:

$$u = \frac{m_{11}X + m_{12}Y + m_{13}Z + m_{14}}{m_{31}X + m_{32}Y + m_{33}Z + m_{34}}, \quad (7)$$

$$v = \frac{m_{21}X + m_{22}Y + m_{23}Z + m_{24}}{m_{31}X + m_{32}Y + m_{33}Z + m_{34}} \quad (8)$$

II.6.1 Solución utilizando la restricción $m_{34} = 1$.

Las ecuaciones (7) y (8) representan la línea recta que pasa por el centro óptico C y el punto p (figura 8). A esta línea también se le llama línea de visión asociada a un punto de la imagen con coordenadas (u, v). Los coeficientes de la matriz M son 12, pero como se aprecia en la ecuación (6), m_{34} , es la componente z del vector de translación de los parámetros extrínsecos y dentro del modelo representa la escala de la imagen con respecto a la escena. Se puede establecer $m_{34} = 1$, lo cual implica dividir todos los coeficientes de **M** entre m_{34} y determinar los parámetros a un factor de escala, de esta manera se tienen 11 incógnitas, y como cada pixel posee dos coordenadas (u, v), se necesitan al menos 6 pixeles de la mira de calibración para obtener una solución mediante este sistema de ecuaciones lineales. Entonces despejando (7) y (8), dejando los términos independientes a la derecha:

$$X_i m_{11} + Y_i m_{12} + Z_i m_{13} + m_{14} - u_i X_i m_{31} - u_i Y_i m_{32} - u_i Z_i m_{33} = u_i m_{34},$$

$$X_i m_{21} + Y_i m_{22} + Z_i m_{23} + m_{24} - v_i X_i m_{31} - v_i Y_i m_{32} - v_i Z_i m_{33} = v_i m_{34}.$$

De esta forma se tienen 2n ecuaciones para n puntos; y en forma matricial:

el eje Z. Mediante identidades trigonométricas se puede verificar que $r_{31}^2 + r_{32}^2 + r_{33}^2 = 1$.

Obteniendo así

$$\|m_3\|^2 = m_{31}^2 + m_{32}^2 + m_{33}^2$$

Entonces, la ecuación (9) se puede dividir en dos partes

$$Bx_9 + Cx_3 = 0 \quad (11)$$

donde \mathbf{B} es una matriz de $2n \times 9$

$$B = \begin{pmatrix} & & & & : & & & & \\ X_i & Y_i & Z_i & 1 & 0 & 0 & 0 & 0 & -u_i \\ 0 & 0 & 0 & 0 & X_i & Y_i & Z_i & 1 & -v_i \\ & & & & : & & & & \end{pmatrix}$$

\mathbf{C} es una matriz de $2n \times 3$

$$C = \begin{pmatrix} & & : & \\ -u_i X_i & -u_i Y_i & -u_i Z_i & \\ -v_i X_i & -v_i Y_i & -v_i Z_i & \\ & & : & \end{pmatrix},$$

$$x_9 = (\mathbf{m}_1 \ m_{14} \ \mathbf{m}_2 \ m_{24} \ m_{34})^T$$

y

$$x_3 = (\mathbf{m}_3)^T$$

Así, la solución se obtiene minimizando el siguiente sistema:

$$Q = \|Bx_9 + Cx_3\|^2, \quad (12)$$

$$\text{sujeto a: } \|x_3\|^2$$

Al final, los parámetros de la cámara están dados por x_3 y x_9 . Como el signo del vector x_3 no está definido, se tienen dos soluciones: \mathbf{M} y $-\mathbf{M}$. Se escoge \mathbf{M} sabiendo que la escena se encuentra delante de la cámara y no detrás de ésta.

II.7 Conclusión.

En este capítulo se han sentado las bases geométricas para llevar a cabo la reconstrucción euclidiana a partir de dos vistas. De las tres cuestiones que se plantearon en la sección I.5 se ha abordado la primera: La geometría de la cámara. Las dos siguientes se dejan para el capítulo 3 y 4 respectivamente.

Las primeras secciones se dedicaron a mostrar la forma en que se representa y manipula lo que es una de las partes fundamentales del marco teórico de la visión por computadora: la geometría proyectiva. Sus principales aportaciones son en aquellas áreas de la visión en las cuales la descripción de imágenes y la propia adquisición de éstas es importante, como en la calibración de la cámara, pares estéreo, síntesis de imágenes, reconstrucción tridimensional, reconocimiento de objetos y análisis de sombras.

Una vez definido el marco teórico, se describió el proceso de adquisición de la imagen por una cámara. Encapsulado, gracias a la geometría proyectiva, en la llamada matriz de proyección. Esta matriz se compone por los parámetros intrínsecos y extrínsecos de

una cámara, es necesaria para llevar a cabo una reconstrucción euclidiana. Pues permite pasar del sistema coordenado del mundo al de la imagen y viceversa. Se analizaron dos formas de calcular los parámetros intrínsecos y extrínsecos de una cámara (calibración) a partir de su modelo geométrico del proceso de adquisición de imágenes. El primer método (restricción $m_{34} = 1$) tiene la desventaja de que los parámetros están a un factor de escala que depende de la translación en z del sistema coordenado de la cámara, porque $t_z = m_{34}$; esto implica que una reconstrucción tridimensional a partir de estos parámetros también resultaría a escala. El segundo método (Faugeras-Toscani) genera los parámetros sin factor de escala. Por lo tanto se ha elegido e implementado éste último para este trabajo de tesis.

Capítulo III

La relación entre dos imágenes: Geometría Epipolar.

En este capítulo se habla acerca de la relación geométrica que existe entre dos imágenes tomadas de una misma escena. Esta relación se encuentra definida por la geometría epipolar. Las imágenes pueden adquirirse simultáneamente, con un par estéreo (dos cámaras montadas una al lado de la otra) o secuencialmente; por una sola cámara que se mueve alrededor de la escena. Estas dos situaciones son geoméricamente equivalentes y no serán diferenciadas en este trabajo.

Primero se describe la geometría epipolar, y de ahí se deriva su pieza clave: la matriz fundamental, junto con las propiedades de ésta. La matriz fundamental es independiente de la estructura de la escena que se proyecta en las imágenes, y se puede calcular de dos maneras; a partir de pares de píxeles de ambas imágenes, sin conocimiento alguno de las cámaras; o bien calcularla a partir de ambas matrices de proyección. Este cálculo se mostrará también.

III.1 Geometría Epipolar.

La llamada geometría epipolar es una parte de la geometría proyectiva que modela los principios que relacionan dos imágenes (que son denominadas *par estéreo*) diferentes de una misma escena (Hartley y Zisserman (2004)); y depende de los parámetros intrínsecos

de las cámaras y de la posición relativa de una con respecto a la otra, además, es independiente de la escena. La matriz fundamental $\mathbf{F}_{3 \times 3}$ encapsula esta geometría. Si un punto \mathbf{P} en el espacio tridimensional se proyecta en el píxel p_1 en la primera imagen, y en p_2 en la segunda, entonces estos píxeles (llamados a partir de ahora *puntos correspondientes*) satisfacen la siguiente relación

$$p_2^T \mathbf{F} p_1 = 0 \quad (13)$$

La geometría epipolar es en esencia, la intersección de los planos de ambas imágenes con un conjunto de planos que tienen como eje la línea base (ésta es la línea que une los centros ópticos de ambas cámaras). Las entidades geométricas que componen la geometría epipolar pueden verse en la figura 9. La terminología es:

- El **epipolo** es el punto de intersección de la línea base con el plano de la imagen. También representa la proyección en una imagen del centro óptico de la cámara de la otra imagen.
- Un **plano epipolar** es aquél que contiene la línea base. Al conjunto de ellos se les llama *lápiz* de planos epipolares.
- Una **línea epipolar** es la intersección de un plano epipolar con el plano de la imagen. Todas las líneas epipolares intersectan con el epipolo de dicha imagen.

Un par estéreo puede tener dos configuraciones, las cuales se modelan de manera indistinta por la geometría epipolar. La primera se da cuando las cámaras son convergentes. Un ejemplo de esta configuración, con varias líneas epipolares, se muestra en la

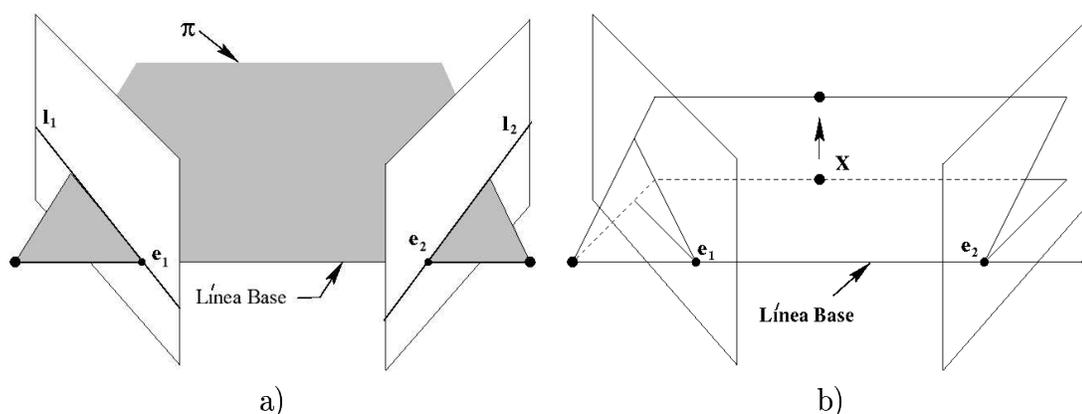


Figura 9: Geometría Epipolar. a) La línea base intersecta ambas imágenes en los epipolos (e). Un plano epipolar (π) intersecta a una imagen en la línea epipolar (l). b) Conforme el punto P varía, el plano epipolar va rotando sobre la línea base

figura 10. En ella se puede apreciar que un objeto proyectado en una imagen, aparece en la otra con cierto desplazamiento en ambas direcciones (arriba-abajo e izquierda-derecha); es decir, el movimiento aparente entre ambas imágenes es una rotación y una traslación. Las líneas epipolares se intersectan en el epipolo, el cual en este caso cae afuera del campo de visión de las imágenes.

En la segunda configuración el plano de la imagen de una cámara es paralelo al de la otra. En este caso, el movimiento aparente es una traslación horizontal, en la figura 11 se puede apreciar que un punto en la imagen derecha aparece desplazado hacia la izquierda con respecto a la imagen izquierda sobre el mismo renglón. En este caso, las líneas epipolares no se intersectan; o formalmente hablando, intersectan en el infinito. En consecuencia, los epipolos están en el infinito y las líneas epipolares son paralelas.

La geometría epipolar es muy importante para llevar a cabo la reconstrucción, como se verá en el siguiente capítulo; ya que un plano epipolar relaciona a los dos planos de

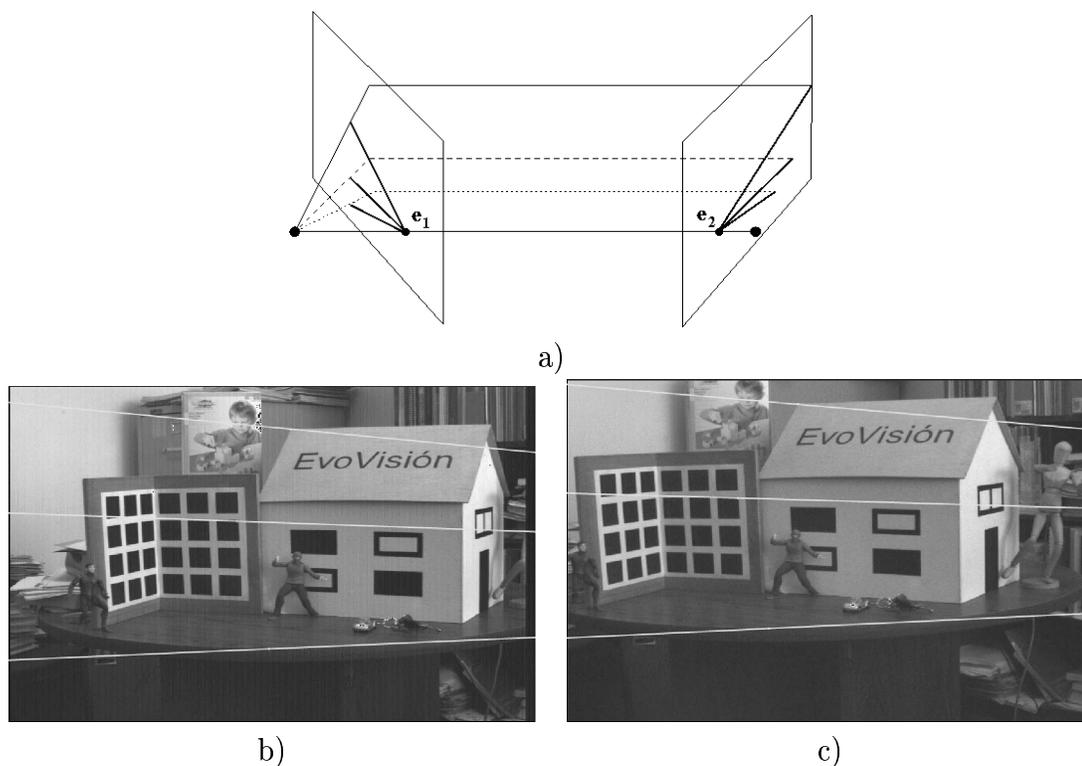


Figura 10: Par estéreo con cámaras convergentes. a) Esquema de la geometría epipolar. b) Imagen Izquierda. c) Imagen Derecha.

las imágenes, pues los intersecta en sendas líneas epipolares; y todos los píxeles que caen en una línea corresponden a algún píxel en la otra línea.

III.2 La matriz Fundamental.

La matriz fundamental es la representación algebraica de la geometría epipolar. En esta sección se deriva la matriz fundamental del mapeo entre un punto y su línea epipolar, inmediatamente después se especifican las propiedades de ésta. Dado un par estéreo, como el de la figura 12, a cada punto p_1 en la primer imagen le corresponde una línea

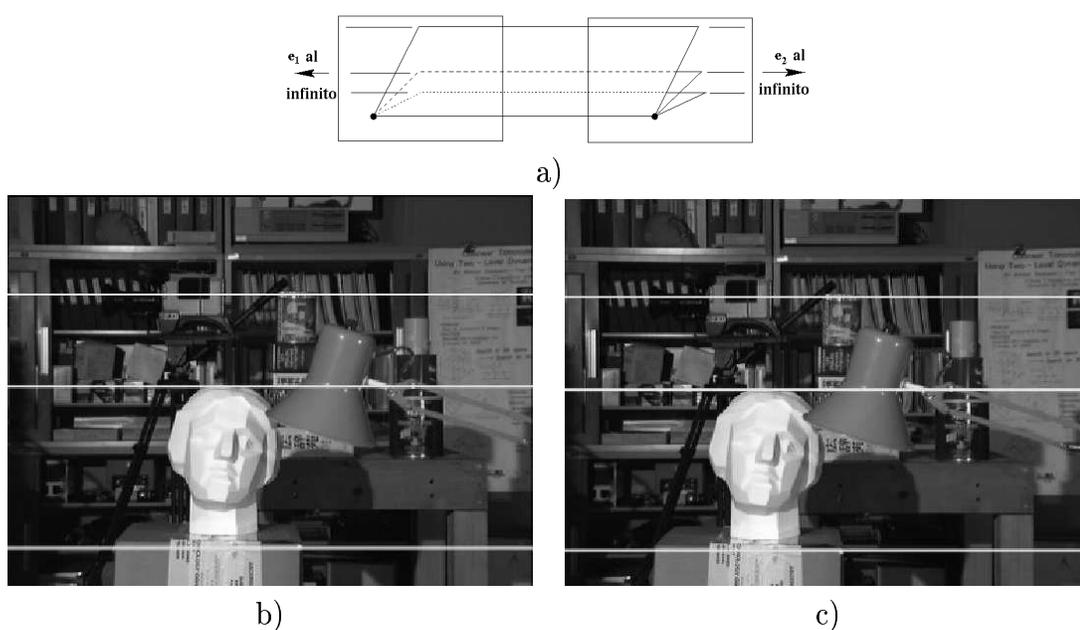


Figura 11: Par estéreo con cámaras paralelas. a) Esquema de la geometría epipolar. b) Imagen Izquierda. c) Imagen Derecha.

epipolar l_2 en la otra imagen. Debido a la geometría epipolar, algún punto p_2 en la segunda imagen correspondiente al punto p_1 de la primera, debe encontrarse a lo largo de la línea epipolar l_2 .

Se puede apreciar en la figura que la línea epipolar l_2 es la proyección en la segunda imagen del rayo que sale del centro óptico de la primer cámara a p_1 . De esta manera se tiene un mapeo $p_1 \mapsto l_2$ desde un punto en una imagen a su correspondiente línea epipolar en la otra imagen.

Este mapeo es una transformación proyectiva de puntos a líneas, el cual está representado en forma algebraica por la matriz F , la matriz fundamental.

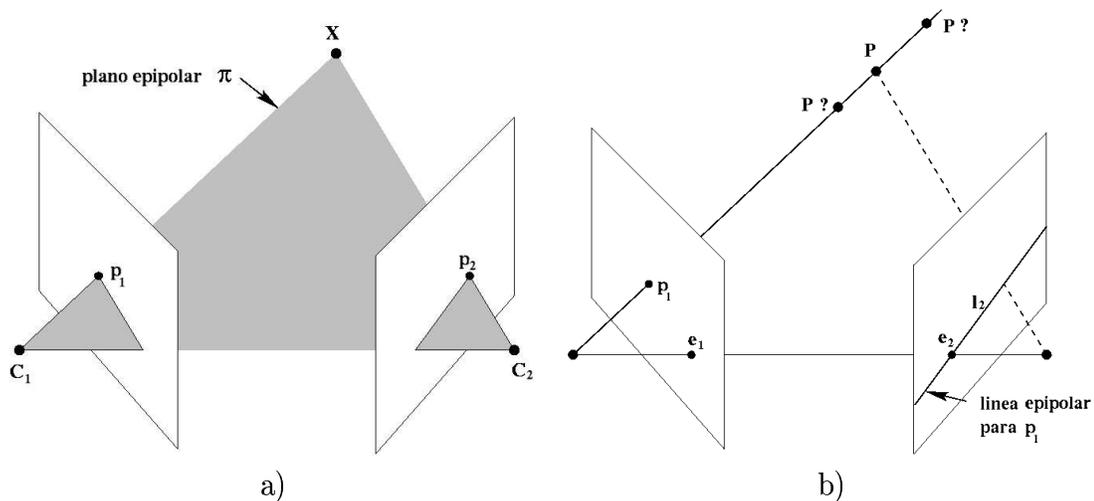


Figura 12: Geometría de F. a) Los centros ópticos, p_1 y p_2 y P caen sobre el mismo plano π . b) El punto P debe estar en alguna parte del rayo formado por C_1 y p_1 , la proyección de éste sobre la segunda imagen es l_2

III.2.1 Derivación Geométrica.

El mapeo de un punto en una imagen a una línea epipolar correspondiente en la otra imagen se puede describir en dos pasos. En el primer paso, un punto p_1 en la primera imagen se mapea a algún punto p_2 de la otra; y en el segundo paso, se obtiene la línea epipolar l_2 de la unión entre p_2 y el epipolo e_2 .

1. **Transferencia del punto vía un plano.** Véase la figura 13. Considérese un plano π en el espacio, que no pase por los centros ópticos de ambas cámaras; considérese también un rayo proyectado a través del centro de la primera cámara, que pasa por p_1 e intersecta al punto P sobre el plano π . Este punto P , se proyecta entonces a un punto p_2 en la segunda imagen. Este procedimiento es conocido como *transferencia vía el plano π* . Por otro lado, debido a que P está sobre el rayo

de p_1 , el punto p_2 debe encontrarse sobre la línea epipolar l_2 (según los principios de geometría epipolar). Los puntos p_1 y p_2 son las proyecciones en la primera y segunda imagen respectivamente, del punto en el espacio tridimensional P ubicado sobre el plano π . El conjunto de todos los puntos p_{1i} en la primera imagen y los puntos correspondientes p_{2i} en la segunda, son proyectivamente equivalentes, debido a que son proyectivamente equivalentes al conjunto de puntos planares P_i . Y así, se puede concluir que existe una homografía H_π que mapea cada punto p_{1i} a p_{2i} .

2. **Construcción de la línea epipolar.** Dado un punto p_2 , la línea epipolar l_2 es aquella que pasa por p_2 y e_2 , y puede expresarse mediante el producto cruz

$$l_2 = e_2 \times p_2 = [e_2]_\times p_2, \quad (14)$$

donde $[e_2]_\times$ es la representación matricial del primer elemento del producto cruz entre dos vectores v_1 y v_2

$$[v]_\times = \begin{pmatrix} 0 & -v_3 & v_2 \\ v_3 & 0 & -v_1 \\ -v_2 & v_1 & 0 \end{pmatrix}$$

Ya que H_π resultó ser una homografía que relaciona p_1 con p_2 , se puede aceptar que $p_2 = H_\pi p_1$. Sustituyendo esta expresión en (14) se obtiene lo siguiente

$$l_2 = [e_2]_\times H_\pi p_1 = F p_1,$$

donde $H_\pi = \mathbf{F}$, que es la matriz fundamental.

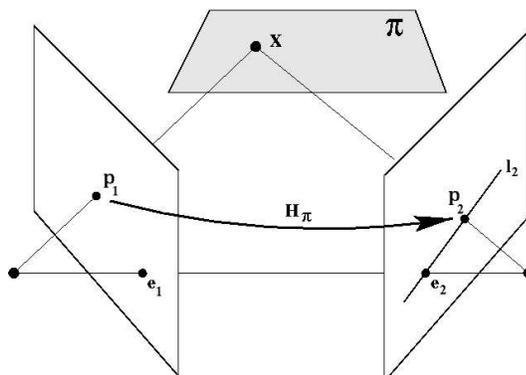


Figura 13: Derivación Geométrica de F .

Geoméricamente, F representa un mapeo del plano proyectivo \mathcal{P}^2 de la primer imagen al conjunto de todas las líneas epipolares posibles en la segunda. Lo cual significa que F realiza un mapeo de 2D a 1D. Nótese que en la derivación que se acaba de dar, se incluye un plano π , pero éste no se requiere para que F exista. Aquí sólo se usa como ilustración del mapeo de una imagen a otra.

III.2.2 Propiedades.

La matriz fundamental es una matriz de 3×3 , homogénea (por ser una transformación proyectiva) y de rango 2 (debido a que son dos el número de columnas linealmente independientes que posee) que satisface la ecuación (13). En la siguiente lista se explican brevemente las propiedades más importantes y útiles de F para la visión por computadora:

- **Transpuesta.** Si F es la matriz fundamental de las cámaras que conforman el par estéreo (M_1, M_2) , entonces F^T es la matriz fundamental de este par en el

orden inverso: (M_2, M_1) .

- **Líneas epipolares.** Para cualquier punto p_1 en la primera imagen, su línea epipolar correspondiente se representa por $l_2 = Fp_1$. De manera similar, y por la primer propiedad, $l_2 = F^T p_2$ representa la línea epipolar correspondiente al punto p_2 en la segunda imagen.
- **El epipolo.** Para cualquier punto p_1 (diferente de e_1), la línea epipolar $l_2 = Fp_1$ contiene al epipolo e_2 . Entonces, e_2 satisface $e_2^T (Fp_1) = (e_2^T F)p_1 = 0$ para todo p_1 . Esto se debe a que $e_2^T F = 0$, es decir, e_2 es el espacio nulo de la izquierda de la matriz F . De igual forma $Fe_1 = 0$, esto es, e_1 es el espacio nulo de la derecha de F .
- **F tiene siete grados de libertad.** Por ser una matriz homogénea de 3×3 tiene ocho razones independientes (son nueve elementos, pero el factor de escala que tienen en común no es significativo); sin embargo, como también satisface la condición de que $\det F = 0$, esto le resta un grado de libertad más.
- **F es una homografía.** Es decir, es una transformación proyectiva que mapea un punto a una línea. En este caso, un punto p_1 de la primera imagen define una línea $l_2 = Fp_1$ en la segunda, la cual es la línea epipolar de p_1 . Sin embargo, como l_1 y l_2 son líneas epipolares correspondientes, entonces cualquier punto p_1 sobre l_1 se mapea sobre la misma línea l_2 . Esto significa, que no hay un mapeo inverso (de una línea a un punto) y que F no es una matriz de rango completo.
- **Cálculo de las matrices de proyección.** Es posible calcular las matrices de

proyección de las cámaras del par estéreo. Esta propiedad se analiza a detalle en la siguiente sección.

III.3 Las matrices de proyección a partir de la matriz fundamental.

Hasta este punto, se han examinado las propiedades de \mathbf{F} y su inferencia en la relación entre un par de puntos correspondientes $p_1 \longleftrightarrow p_2$. Ahora se verá una de las propiedades más significativas de la matriz Fundamental: determinar a partir de ella las matrices de proyección de las dos imágenes.

Como se dijo en un principio, la ecuación (13) es un sistema homogéneo, lo que significa que es un sistema proyectivo, ya que se deriva a partir de relaciones propias de la geometría proyectiva, como la intersección de planos y líneas; además de que depende solamente de las coordenadas de los puntos (en píxeles) de la imagen. No se encuentran involucradas medidas de geometría euclidiana como podría ser el ángulo entre los dos rayos. Esto significa que la relación entre dos imágenes es proyectivamente invariante, es decir, que al aplicar una transformación proyectiva H a ambos puntos correspondientes: $p'_1 = H_1 p_1$ y $p'_2 = H_2 p_2$; se cumple que $F' = H_1^{-T} F H^{-1}$.

Las matrices de proyección son las que relacionan un punto en el espacio con un punto en el plano de la imagen, en este proceso tiene que ver un sistema coordenado euclidiano en el mundo real, del que estas matrices depende. Por ejemplo una rotación de este sistema coordenado cambiaría los coeficientes de M_1 y M_2 . Sin embargo, la matriz fundamental no depende del mundo real, sólo de los puntos de correspondencia en las imágenes. Así, si $p_1 \longleftrightarrow p_2$ son puntos de correspondencia con respecto a M_1 y

M_2 , de un punto P , $p_1 \longleftrightarrow p_2$ también lo son para M_1H y M_2H , de un punto H^1P . De esto último se puede inferir que la matriz fundamental para M_1 y M_2 es la misma que para M_1H y M_2H .

III.3.1 Forma canónica de las matrices de proyección.

Dada esta ambigüedad, es común especificar una *forma canónica* del par de matrices, en la cual, $M_1 = [I \mid 0]$, donde I es la matriz identidad de 3×3 , y 0 es un vector columna de ceros de 3×1 ; y $M_2 = [SF \mid e_2]$, donde e_2 es el epipolo de la imagen derecha y S es una matriz antisimétrica.

Las matrices de proyección deben tener rango 3. M_1 lo tiene. M_2 también lo tiene, de acuerdo a los siguientes argumentos. S puede escribirse en términos de su vector nulo como $S = [s]_{\times}$. Debido a que $e_2F = 0$, el *espacio columna*¹ de \mathbf{F} es perpendicular a e_2 . Así, $s^T e_2 \neq 0$, entonces s no es perpendicular a e_2 , y así, no está en el espacio columna de \mathbf{F} . El espacio columna de $[s]_{\times}F$ se define por el producto cruz de s con las columnas de \mathbf{F} ; y es igual al plano perpendicular a s , por lo que tiene rango 2. Como e_2 no es perpendicular a s , no cae en este plano; y por lo tanto $[[s]_{\times}F \mid e_2]$ tiene rango 3, como se requiere.

Como sugieren Luong y Vieville (Hartley y Zisserman (2004)), una buena opción para definir S es $S = [e_2]_{\times}$; ya que $e_2^T e_2 \neq 0$. Quedando así las dos matrices definidas como:

$$M_1 = [I \mid 0], \tag{15}$$

¹El espacio columna de una matriz, es el espacio vectorial generado por las columnas de esta matriz, vistas como vectores.

$$M_2 = [[e_2] \times F \mid e_2] \quad (16)$$

Hay que recordar la ambigüedad de que la matriz fundamental es la misma para M_1 y M_2 que para M_1H_1 y M_2H_2 ; donde H es una transformación proyectiva. Así, las matrices M_1 y M_2 resultantes serán también proyectivas, lo que al llevar a cabo una reconstrucción tridimensional a partir de ellas, resultaría en una reconstrucción proyectiva.

III.4 Cálculo de la matriz fundamental.

Existen varios métodos numéricos que estiman \mathbf{F} a partir de un conjunto de puntos correspondientes de un par estéreo (Hartley y Zisserman (2004)); en esta sección se analizan varios de ellos: el método lineal, que es el más sencillo de implementar, junto con su mejora; otro método basado en la distancia geométrica, y finalmente, un método que selecciona automáticamente los mejores puntos de correspondencia para obtener un mejor resultado.

La matriz fundamental se define por $p_2^T \mathbf{F} p_1 = 0$, que es la ecuación (13), para cualquier conjunto de puntos correspondientes en las imágenes $p_1 \longleftrightarrow p_2$. Conociendo suficientes pares de puntos de correspondencia, se puede usar esta ecuación para obtener \mathbf{F} . Sean $p_1 = (u_1, v_1, 1)$ y $p_2 = (u_2, v_2, 1)$, al desarrollar de forma algebraica la ecuación (13), un par de puntos cualquiera genera una ecuación lineal cuyas incógnitas son los coeficientes de \mathbf{F}

$$u_2 u_1 f_{11} + u_2 v_1 f_{12} + u_2 f_{13} + v_2 u_1 f_{21} + v_2 v_1 f_{22} + v_2 f_{23} + u_1 f_{31} + v_1 f_{32} + f_{33} = 0 \quad (17)$$

Con n pares de puntos correspondientes se obtiene el siguiente conjunto de ecuaciones lineales, en forma matricial

$$\mathbf{A}\mathbf{f} = \begin{pmatrix} u_{21}u_{11} & u_{21}v_{11} & u_{21} & v_{21}u_{11} & v_{21}v_{11} & v_{21} & u_{11} & v_{11} & 1 \\ & & & & : & & & & \\ u_{2i}u_{1i} & u_{2i}v_{1i} & u_{2i} & v_{2i}u_{1i} & v_{2i}v_{1i} & v_{2i} & u_{1i} & v_{1i} & 1 \\ & & & & : & & & & \\ u_{2n}u_{1n} & u_{2n}v_{1n} & u_{2n} & v_{2n}u_{1n} & v_{2n}v_{1n} & v_{2n} & u_{1n} & v_{1n} & 1 \end{pmatrix} \mathbf{f} = 0 \quad (18)$$

La matriz fundamental se compone de 9 coeficientes, pero uno de ellos es un término independiente, como puede verse en (17). La matriz \mathbf{A} de (18) debe tener un rango de al menos 8; si es exactamente ocho se obtendrá una solución única. Por lo tanto, para resolver este sistema, se necesitan al menos 8 pares de puntos correspondientes. Se han desarrollado distintos métodos para obtener la matriz fundamental, cada uno tiene sus ventajas y desventajas. A continuación se muestra una descripción de los métodos que se tomaron en cuenta durante este trabajo.

III.4.1 Solución Lineal. (Algoritmo de los ocho puntos)

Este método, desarrollado por Longuet-Higgins (H.C. (1981)), es el más simple, pues propone una solución de mínimos cuadrados al conjunto de ecuaciones lineales de (18). La solución de mínimos cuadrados que él propone para \mathbf{f} , es el vector singular (eigenvector) que corresponde al valor singular (eigenvalor) más pequeño de la matriz \mathbf{A} , esto es, la última columna de \mathbf{V} de método numérico para la descomposición en valores singulares (SVD) de una matriz: $\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{V}^T$ (Press *et al.* (1994)). El vector solución \mathbf{f} encontrado de esta manera minimiza:

$$\| \mathbf{A} \mathbf{f} \| \text{ sujeto a } \| \mathbf{f} \| = 1$$

Una propiedad muy importante de \mathbf{F} es que es singular, de rango 2. Muchas de las aplicaciones de \mathbf{F} dependen de este hecho. Por ejemplo, el cálculo de las líneas epipolares. Generalmente, la matriz \mathbf{F} calculada con este método no es de rango 2. Sin embargo esta propiedad se puede forzar. La forma más conveniente de hacerlo es reemplazar \mathbf{F} con la matriz \mathbf{F}' que minimice la norma de Frobenius:

$$\| \mathbf{F} - \mathbf{F}' \|$$

sujeto a: $\det \mathbf{F}' = 0$.

La manera de hacerlo es, otra vez, con SVD. Sea $\mathbf{F} = \mathbf{U} \mathbf{D} \mathbf{V}^T$, con $\mathbf{D} = \text{diag}(r, s, t)$ y $r \geq s \geq t$. Entonces $\mathbf{F}' = \mathbf{U} \text{diag}(r, s, 0) \mathbf{V}^T$ minimiza la norma de Frobenius definida arriba.

Así el algoritmo de los ocho puntos para calcular la matriz fundamental, puede resumirse en el siguiente algoritmo:

Algoritmo de 8 puntos:

1. **Solución Lineal:** La matriz fundamental se construye con los coeficientes de \mathbf{f} de la ecuación (18), que corresponde al eigenvalor más pequeño de \mathbf{A} .
2. **Forzar la restricción:** Reemplazar \mathbf{F} con \mathbf{F}' ; que es la matriz singular más cercana a \mathbf{F} bajo la norma de Frobenius

Existe un algoritmo que utiliza solo 7 puntos pero se utiliza en casos especiales cuando la matriz \mathbf{A} tiene rango siete y esto, generalmente arroja como resultado dos o tres matrices fundamentales. Esta característica se usa en los dos últimos algoritmos que se analizan aquí.

III.4.2 Algoritmo de los ocho puntos normalizado.

El algoritmo anterior es el más simple, y si se realiza procurando un buen conjunto de puntos correspondientes, puede generar una solución extremadamente buena. La clave para lograrlo consiste en normalizar las coordenadas del conjunto de pares de puntos correspondientes antes de desarrollar el sistema de ecuaciones. Es posible que los puntos que se elijan estén muy alejados entre sí, dentro del sistema de coordenadas de la imagen. Hartley (Hartley y Zisserman (2004)) sugiere que se lleve a cabo una transformación simple de los puntos para darle estabilidad al sistema de ecuaciones. El costo de este pre-procesamiento de los puntos, es relativamente insignificante, y mejora sustancialmente el resultado.

Esta transformación consiste en una traslación y un escalamiento de la imagen, de manera que el centroide de los puntos que se van a usar en el método quede en el origen del sistema coordenado (traslación) y la distancia euclidiana del origen a los puntos sea $\sqrt{2}$ (escalamiento). Después de este preprocesamiento, el algoritmo es muy similar al anterior. A continuación se muestra:

Algoritmo de 8 puntos normalizado:

1. **Normalización:** Transformar las coordenadas de la imagen de la siguiente manera: $\hat{u}_{1i} = T_1 u_{1i}$ y de la misma manera para v con T_2 ; donde T_1 y T_2 son transformaciones de normalización que consiste en una traslación y un escalamiento. A partir de ahora \hat{u}_i y \hat{v}_i son las coordenadas de los puntos normalizados.
2. Encontrar la matriz Fundamental $\hat{\mathbf{F}}'$ que corresponde a los puntos correspondientes $\hat{p}_1 \longleftrightarrow \hat{p}_2$ de la siguiente manera:

- (a) **Solución lineal:** Determinar $\hat{\mathbf{F}}$ a partir del eigenvector correspondiente al eigenvalor más pequeño de la matriz $\hat{\mathbf{A}}$. Donde $\hat{\mathbf{A}}$ se compone del conjunto de puntos $\hat{p}_1 \longleftrightarrow \hat{p}_2$.
- (b) **Cumplimiento de la restricción forzoso:** Reemplazar $\hat{\mathbf{F}}$ con la matriz $\hat{\mathbf{F}}'$ que minimize la norma de Frobenius usando SVD.
3. Realizar el proceso inverso a la normalización para obtener la matriz fundamental, esto es: $\mathbf{F} = T_2^T \hat{\mathbf{F}}' T_1$

El factor de escala que se elige depende de las condiciones del par de imágenes y del grado de dispersión en el que se encuentren los puntos. Sin embargo, debe cumplir la restricción de que la distancia media del conjunto de puntos al origen sea de $\sqrt{2}$; esto es, que el promedio de las coordenadas de los puntos sea: (1, 1, 1).

III.4.3 Algoritmo de Estimación de Máxima Verosimilitud. (MLE)

También se le conoce como método *Gold Standard*. Los puntos de correspondencia son propensos a tener ruido, esto es, por la limitada resolución de la cámara, o por errores de redondeo en cualquier método numérico que se utilice para adquirir estos puntos de la imagen. El método gold standard hace la suposición de que el ruido presente en los puntos de correspondencia es gaussiano, su objetivo entonces, es resolver un modelo de este error mediante una estimación de máxima verosimilitud. El modelo de error consiste en la distancia geométrica

$$\sum d(p_{1i}, \hat{p}_{1i})^2 + d(p_{2i}, \hat{p}_{2i})^2, \quad (19)$$

donde $p_1 \longleftrightarrow p_2$ es el conjunto de puntos de correspondencia y \hat{p}_1 y \hat{p}_2 son las correspondencias estimadas que satisfacen (13) exactamente.

La manera de minimizar este modelo se muestra en el siguiente algoritmo:

Algoritmo MLE:

1. Calcular \hat{F} inicial de rango 2 usando el algoritmo lineal.
2. Calcular las correspondencias estimadas \hat{p}_{1i} y \hat{p}_{2i} de la siguiente manera:
 - (a) Definir las matrices de proyección como $M_1 = [I \mid 0]$ y $M_2 = [[e_2] \times \hat{F} \mid e_2]$
 - (b) Con los puntos $p_{1i} \longleftrightarrow p_{2i}$ y \hat{F} calcular el punto en el espacio \hat{P}_i usando un método de triangulación (ver capítulo siguiente)
 - (c) Los puntos estimados quedan definidos como: $\hat{p}_{1i} = M_1 \hat{P}_i$ y $\hat{p}_{2i} = M_2 \hat{P}_i$
3. Minimizar (19) sobre \hat{F} y \hat{P}_i , $i = 1..n$.

Nótese que este método utiliza el método lineal como punto de partida para calcular la matriz fundamental; pues es necesario tener \mathbf{F} para calcular las matrices de proyección y estimar los puntos correspondientes, los cuales satisfarán la ecuación (13). Las matrices de proyección se obtienen a partir de la matriz fundamental, como se explica en la sección III.3.

La minimización se lleva a cabo mediante el algoritmo **Levenberg-Marquardt**. Éste es un método de minimización de parámetros iterativo y deriva del método más general de *Iteración de Newton*. Una aplicación particular de estos métodos es la de encontrar los ceros en una función de una sola variable. Su generalización a varias variables y su aplicación a encontrar soluciones de mínimos cuadrados en lugar de

soluciones exactas es relativamente directa (Hartley y Zisserman (2004)). El método Levenberg-Marquardt es una variación del de Newton que permite una convergencia a la solución más rápida. Levenberg-Marquardt se puede considerar como una combinación de iteración de Newton y el método de *descenso gradiente*.

Este método es muy costoso computacionalmente, debido las operaciones que realiza (calcula de las matrices de proyección, triangulación y minimización). Sin embargo es ampliamente recomendado (Hartley y Zisserman (2004)) por ser estable, robusto y preciso.

III.4.4 Algoritmo RANSAC.

Este algoritmo se basa en una técnica conocida como *RANdom SAmple Concensus* (RANSAC), que significa *Consenso de una muestra escogida al azar*. RANSAC es una técnica para mejorar un grupo de datos según cierto modelo que éstos deben seguir. Logra su objetivo seleccionando iterativamente un subconjunto aleatorio del grupo original de datos. A este subconjunto se le somete a prueba contra el modelo, es decir, se determina qué tan bien un dato encaja en las especificaciones del modelo. Después de k iteraciones, solo se conserva el subconjunto de los que mejor encajan; a éstos se les llama *inliers* (ver figura 14).

Considérese un conjunto de puntos correspondientes $p_1 \longleftrightarrow p_2$ cuya única fuente de error posible es la medida de sus posiciones en las imágenes, la cual sigue una distribución gaussiana. En este caso, el modelo al que se deben ajustar los datos es a cumplir con la ecuación 13; por consiguiente, se debe tener una matriz fundamental inicial. Así, el objetivo del método de estimación de la matriz fundamental basado

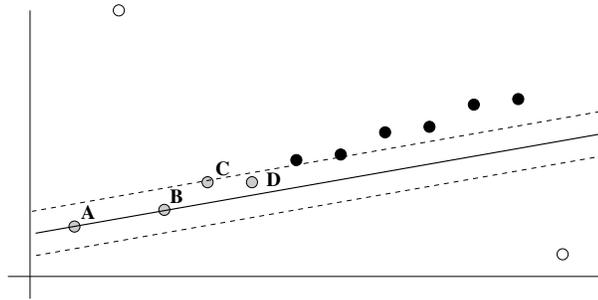


Figura 14: Técnica RANSAC. Los puntos grises representan *inliers*, en este caso el modelo al que deben ajustarse los puntos es una línea recta, en líneas punteadas se aprecia el margen de holgura.

en RANSAC es calcular la matriz Fundamental con los datos disponibles, eliminar los puntos imprecisos a partir ella y calcular una nueva matriz \mathbf{F} . A continuación se describe el algoritmo.

Algoritmo RANSAC:

1. Repetir lo siguiente hasta que se asegure cierto porcentaje de inliers.
 - (a) Seleccionar 7 pares de puntos de manera aleatoria y obtener una matriz \mathbf{F} con el metodo de 7 puntos, lo que dará de una a tres soluciones diferentes para \mathbf{F} .
 - (b) Calcular la distancia d_{\perp} que cada par de puntos del conjunto completo tiene con respecto a lo que indique \mathbf{F} .
 - (c) Calcular el numero de puntos que sean consistentes con \mathbf{F} (aquellos cuya distancia sea menor a cierto número de pixeles).
 - (d) Si existen varias soluciones para \mathbf{F} el paso anterior se calcula por cada una

de ellas y se escoge aquella con el mayor número de puntos que cumplan. En caso de empate se escoge la que tenga la desviación estándar más pequeña.

2. Re-estimar \mathbf{F} con solo los *inliers*, minimizando (19) con el algoritmo Levenberg-Marquardt

Este método también utiliza un método lineal para calcular una matriz fundamental inicial, pero prefiere el de 7 puntos sobre el de 8 porque para la técnica RANSAC, el número de iteraciones para asegurar una alta probabilidad de *inliers* es exponencial según el tamaño de la muestra. Así para asegurar un 99% de *inliers* en el subconjunto final se necesitan dos veces más iteraciones con ocho puntos que con siete. Sin embargo, la desventaja es que puede arrojar hasta tres soluciones, y todas deben tomarse en cuenta.

La distancia d_{\perp} mide qué tan cerca está un par de puntos de correspondencia con respecto a la geometría epipolar que indique \mathbf{F} . Esto se hace de la misma manera que en el paso 2 del método MLE. Es decir, se calculan las matrices de proyección, el punto en el espacio 3D y se genera \hat{p}_{1i} y \hat{p}_{2i} para obtener sus distancias a p_{1i} y p_{2i} . Un par de puntos se considera un *inlier* si su distancia es menor a un umbral que se deja de holgura. Hartley (Hartley y Zisserman (2004)) aconseja que éste sea de 1.25 píxeles.

III.4.5 La matriz fundamental a partir de las matrices de proyección.

Si cada una de las cámaras cuenta ya con su matriz de proyección, calculadas mediante el proceso de calibración (ver capítulo 2), es posible calcular a partir de ellas la matriz fundamental que las relacione a ambas. El proceso de calibración implica el cálculo de

la matriz de proyección a partir de coordenadas conocidas de puntos del mundo real, es decir, del espacio euclidiano. Por lo tanto, la matriz fundamental estará también en un contexto euclidiano, a diferencia de los métodos anteriores, que calculan \mathbf{F} a partir de puntos de correspondencia en las imágenes y por lo tanto la matriz está en un contexto proyectivo.

La matriz fundamental relaciona las proyecciones del punto del espacio 3D en cada una de las imágenes; entonces el primer paso es expresar el punto de la imagen derecha en función del mismo punto en la imagen izquierda. A este proceso se le conoce como *calibración estereoscópica* (figura 15) de ahí que a dos cámaras tomando imágenes de una misma escena se le llama par estéreo.

La matriz de transformación A_S convierte un punto P del sistema coordenado de la cámara izquierda $P = (X_1, Y_1, Z_1)$ al de la derecha $P = (X_2, Y_2, Z_2)$. Para calcularla es necesario extraer los parámetros extrínsecos \mathbf{A} de cada cámara. Con A_1 y A_2 se calcula A_S de la siguiente manera

$$A_S = A_2 A_1^{-1} \quad (20)$$

A_S se compone de una transformación de rotación y una de translación

$$A_S = \begin{pmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Ahora se puede expresar un punto P al mismo tiempo en ambos sistemas coordenados

$$\begin{pmatrix} X_2 \\ Y_2 \\ Z_2 \\ 1 \end{pmatrix} = A_S \begin{pmatrix} X_1 \\ Y_1 \\ Z_1 \\ 1 \end{pmatrix}. \quad (21)$$

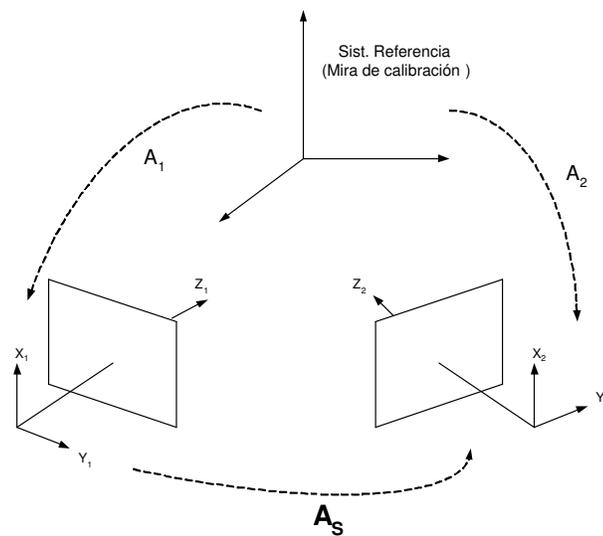


Figura 15: Par Estéreo y la matriz que relaciona ambos sistemas coordenados: A_S .

Las coordenadas de los pixeles del plano de la imagen tienen su equivalente en el sistema coordenado de la cámara, y están dadas por

$$\begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = C^{-1} \begin{pmatrix} u \\ v \\ 1 \end{pmatrix}, \quad (22)$$

donde la matriz C es (3). Aplíquese (22) de la misma manera para la imagen

izquierda y derecha.

Así, las coordenadas en el sistema coordenado de la cámara para p_1 son $(x_1, y_1, 1)$, donde $x_1 = X_1/Z_1$ y $y_1 = Y_1/Z_1$ y de manera similar para p_2 , x_2 y y_2 pueden escribirse de la siguiente manera, derivandola a partir de (21):

$$x_2 = \frac{X_2}{Z_2} = \frac{r_{11}X_1 + r_{12}Y_1 + r_{13}Z_1 + t_x}{r_{31}X_1 + r_{32}Y_1 + r_{33}Z_1 + t_z}$$

$$y_2 = \frac{Y_2}{Z_2} = \frac{r_{21}X_1 + r_{22}Y_1 + r_{23}Z_1 + t_y}{r_{31}X_1 + r_{32}Y_1 + r_{33}Z_1 + t_z}$$

Tomando en cuenta que $X = xZ$ y $Y = yZ$, además que $p_1 = (x_1, y_1, 1)$ y que r_i representa el i -ésimo renglón de la matriz de rotación de A_S , lo de arriba se reescribe como:

$$x_2 = \frac{Zr_1p_1 + t_x}{Zr_3p_1 + t_z}$$

$$y_2 = \frac{Zr_2p_1 + t_y}{Zr_3p_1 + t_z}$$

Y de esta manera ya se tiene el punto p_2 en función del punto p_1 . Ahora se procede a obtener la matriz \mathbf{F} a partir de estas ecuaciones. Primero que nada, si en las ecuaciones de arriba se despeja Z y se igualan una con otra, queda la relación lineal

$$(t_z r_2 p_1 - t_y r_3 p_1) x_2 + (t_x r_3 p_1 - t_z r_1 p_1) y_2 = t_x r_2 p_1 - t_y r_1 p_1$$

La cual, describe un conjunto de puntos de la imagen derecha que podrían corresponder al punto en la izquierda. Estos puntos forman una línea ($ax + by + c$): *la línea epipolar derecha*. de la forma $a_2 x_2 + b_2 y_2 + c_2 = 0$; donde:

$$a_2 = t_z r_2 p_1 - t_y r_3 p_1 = (t_y r_{31} - t_z r_{21}) x_1 + (t_y r_{32} - t_z r_{22}) y_1 + (t_y r_{33} - t_z r_{23})$$

$$b_2 = t_x r_3 p_1 - t_z r_1 p_1 = (t_z r_{11} - t_x r_{31})x_1 + (t_z r_{12} - t_x r_{32})y_1 + (t_z r_{13} - t_x r_{33})$$

$$c_2 = t_x r_2 p_1 - t_y r_1 p_1 = (t_x r_{21} - t_y r_{11})x_1 + (t_x r_{22} - t_y r_{12})y_1 + (t_x r_{23} - t_y r_{13})$$

Y en forma matricial queda

$$\begin{pmatrix} a_2 \\ b_2 \\ c_2 \end{pmatrix} = \begin{pmatrix} 0 & -t_z & t_y \\ t_z & 0 & -t_x \\ -t_y & t_x & 0 \end{pmatrix} \begin{pmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{pmatrix} \begin{pmatrix} x_1 \\ y_1 \\ 1 \end{pmatrix}$$

O bien, agrupando las dos matrices en una sola

$$\begin{pmatrix} a_2 \\ b_2 \\ c_2 \end{pmatrix} = E \begin{pmatrix} x_1 \\ y_1 \\ 1 \end{pmatrix} \quad (23)$$

Donde la matriz E es la llamada matriz esencial. Esta es la transformación epipolar en la que a un punto de la imagen izquierda $(x_1, y_1, 1)$ corresponde una recta en la imagen derecha.

Se sabe que la ecuación de la recta $(ax+by+c)$ se puede expresar como multiplicación de vectores así

$$\begin{pmatrix} x & y & 1 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = 0$$

Aplicándolo a 23

$$\begin{pmatrix} x_2 & y_2 & 1 \end{pmatrix} E \begin{pmatrix} x_1 \\ y_1 \\ 1 \end{pmatrix} = 0$$

Obteniendo así la relación de la imagen izquierda con la derecha

$$p_2^T E p_1 = 0 \quad (24)$$

Del mismo modo la derecha con la izquierda $p_1^T E^T p_2 = 0$

Con la matriz esencial, la relación entre ambas imágenes es a nivel de coordenadas de la cámara; pero la matriz fundamental se define en coordenadas de la imagen. Para extender la relación al nivel de coordenadas de la imagen (píxeles) se aplica la siguiente transformación

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \begin{pmatrix} \alpha_u & 0 & u_0 \\ 0 & \alpha_v & v_0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}$$

Que matricialmente se expresa $p_{im} = C p_c$. Valido para ambas cámaras. Así, al despejar p_c y sustituir en (24), tenemos que $p_2^T (C_2^{-1})^T E C_1^{-1} p_1$

Si se asume que $F = (C_2^{-1})^T E C_1^{-1}$, la ecuación (24) queda de la siguiente manera

$$p_2^T \mathbf{F} p_1 = 0,$$

donde \mathbf{F} es la matriz fundamental que describe la geometría epipolar entre ambas imágenes a nivel de píxel, como en (13).

III.5 Experimentación.

En esta sección se lleva a cabo una comparación de los cuatro algoritmos citados arriba para obtener la matriz Fundamental. También se muestra un ejemplo ilustrativo de la geometría epipolar. No hubo necesidad de implementar la calibración de la cámara

ya que este trabajo fue hecho anteriormente, ver (Pérez (2004); Ramirez (2000)), como parte del proyecto *EvoVisión*, al que esta tesis pertenece.

El par de imágenes que se utilizó para esto se muestra en la figura 16. Las pruebas se hicieron tomando pares de puntos manualmente en ambas imágenes, se compararon dos muestras diferentes, esto con el fin de tener dos escenarios diferentes: el primero consiste en un conjunto de 12 pares de puntos, tomados con ayuda de un detector de esquinas con precisión subpíxel de Hernández (Hernández (2002)), La segunda consiste en un conjunto de 64 pares de puntos tomados con la ayuda del detector de esquinas de Harris con precisión a píxel (Harris y Stephens (1988)). Se ha desarrollado el algoritmo, se ha corrido 10 veces cada uno y se ha tomado el error promedio para cada algoritmo.

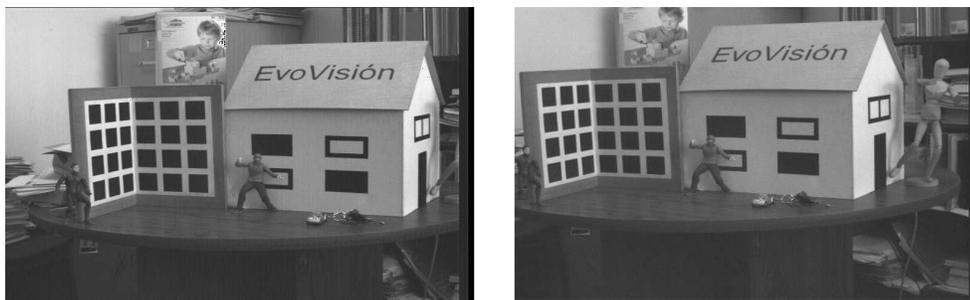


Figura 16: Primer par de imágenes utilizadas para realizar pruebas.

El error esta definido como sigue

$$e = \frac{1}{N} \sum_{i=1}^N d(p_{2i}, Fp_{1i})^2 + (p_{1i}, F^T p_{2i})^2 \quad (25)$$

Donde $d(x, l)$ es la distancia (medida en pixeles) entre el punto x y la línea l . El error

Tabla III: Resultados de experimentaciones con los métodos para calcular F.

	Lineal	Lineal Normalizado	MLE	RANSAC
12 Puntos	1.06055	0.104263	0.00557306	0.00714011
64 Puntos	108.217	0.390625	0.263012	0.394947

consiste en la distancia al cuadrado entre el punto correspondiente y la línea epipolar en la cual este punto debería estar (esto va por cada punto correspondiente). N es el número de puntos que componen la muestra. Los resultados se muestran en la tabla III.

El error está medido en píxeles. Éstos se reflejan directamente en el cálculo de líneas epipolares, como se ve en la figura 17. Esta figura muestra al programa `xcv` que se desarrolló durante este trabajo, que entre otras cosas, es capaz de visualizar las líneas epipolares entre un par de imágenes. La ventana de la aplicación se encuentra dividida en dos, cada parte contiene una de las dos imágenes de la figura 16. Cada imagen tiene un punto que corresponde a la línea epipolar de la otra imagen; el punto de la imagen izquierda se encuentra en un extremo de la primera letra "v" de *Evovisión*, mientras que en la imagen derecha el punto se encuentra en la ventana inferior de la casa de madera. En la figura, ambos puntos aparecen resaltados con cruces, tanto en el inciso (a) como en el inciso (b). Para demostrar que la magnitud del error está relacionada a la exactitud de los puntos que se emplean para calcular la matriz fundamental, se ilustra la prueba con los 64 puntos con precisión a píxel. Claramente se puede ver que el método lineal usado en el inciso a) de la figura 17 la líneas epipolares no pasan

por donde lo indica el punto de la otra imagen. Mientras que en b), usando MLE, un método menos susceptible a los errores en los datos de entrada, si se cumple con esta condición. En el caso del método lineal puede apreciarse que para el punto de la izquierda, la correspondiente línea epipolar en la imagen derecha pasa por el otro extremo de la letra "v", mientras que para el punto de la ventana en la imagen derecha, la línea epipolar correspondiente en la imagen izquierda pasa varios píxeles por debajo de donde debería pasar. Los resultados mejoran notablemente usando MLE, como se aprecia en el inciso b).

III.6 Conclusiones.

En este capítulo se han sentado las bases geométricas para llevar a cabo la reconstrucción euclidiana a partir de dos vistas. De las tres cuestiones que se plantearon en I.5, la segunda se abordó aquí; la cual representa la antesala de la reconstrucción: La geometría de la cámara y la geometría epipolar. La tercera, que representa a la reconstrucción en si, se deja para el siguiente capítulo.

La geometría epipolar relaciona dos imágenes de una misma escena mediante la matriz fundamental \mathbf{F} . Se describieron aquí 4 métodos diferentes para obtener \mathbf{F} con un conjunto de puntos correspondientes conocidos de antemano, y otro método que la obtiene a partir de las matrices de proyección de ambas cámaras. De los cuales se puede decir lo siguiente:

- **Lineal.** Es un método bastante simple y fácil de implementar, pues implica no más que la solución de mínimos cuadrados de un conjunto de ecuaciones lineales; sin embargo resultó ser el más impreciso en ambas muestras utilizadas. No es



a)



b)

Figura 17: Línea epipolar usando muestra de 64 puntos. a) Lineal. b) MLE.

muy recomendable, en todo caso, es mejor usar la versión normalizada.

- **Lineal Normalizado.** La versión normalizada del método lineal muestra un buen rendimiento. Proporciona resultados muy aceptables, además se utiliza como base para los métodos más complejos. Si lo que se busca es rapidez antes que precisión se recomienda utilizar este método.

- **MLE.** Los mejores resultados obtenidos en este trabajo se han dado aplicando este método, son también los mejores reportados en la literatura (Hartley y Zisserman (2004)), siempre y cuando la suposición de que el ruido presente es gaussiano sea viable. La desventaja es el tiempo que consume la minimización en el paso 3 de su algoritmo.
- **RANSAC.** El método RANSAC ha demostrado ser al menos tan bueno como el método lineal. Es un método considerado como automático, ya que por sí solo descarta ciertos pares de puntos y conserva los mejores para un mejor resultado al calcular \mathbf{F} . La desventaja que presenta es que consume demasiado tiempo al elegir a los mejores pares de puntos en el paso 1, por no mencionar el tiempo en la minimización del paso 2.
- **\mathbf{F} a partir de \mathbf{M}_1 y \mathbf{M}_2 .** Es un método muy sencillo que consiste sólo de multiplicación de matrices, una vez que se tienen las matrices de proyección. Además, si las matrices de proyección son resultado de un proceso de calibración, \mathbf{F} estará en un contexto euclidiano, así, el resultado numérico de la matriz fundamental es muy diferente al que se obtiene en los cuatro métodos anteriores.

Las matrices de proyección pueden calcularse a partir de \mathbf{F} . Sin embargo, si se utilizan estas matrices para llevar a cabo la reconstrucción no se obtendría una reconstrucción euclidiana, sino proyectiva. Por lo cual, esta propiedad de la matriz fundamental no se utiliza para llevar a cabo la reconstrucción. Sin embargo, representa un avance significativo para el trabajo futuro, cuando se desee llevar a cabo una reconstrucción proyectiva.

El cálculo de \mathbf{F} se utiliza para resolver el problema de correspondencia estéreo, que es el paso inmediato anterior a la reconstrucción euclidiana bajo un enfoque geométrico. La precisión de \mathbf{F} es de vital importancia para la precisión de la correspondencia, que a su vez es esencial para un buen resultado en la reconstrucción. Ambos se analizan en el siguiente capítulo .

Capítulo IV

Reconstrucción 3D: Triangulación.

Hasta ahora se ha definido la matriz fundamental \mathbf{F} a partir de unos pocos píxeles de correspondencia conocidos de antemano. Una vez que se tiene \mathbf{F} el siguiente paso es encontrar el píxel correspondiente de cada píxel de la imagen izquierda en la imagen derecha. Esto es lo que se conoce como el problema de *correspondencia estéreo* y el resolverlo de manera precisa es esencial para llevar a cabo una buena reconstrucción. La matriz fundamental relaciona un píxel de la imagen izquierda con una línea de la imagen derecha, lo que reduce el espacio de búsqueda, del plano de la imagen a la línea epipolar. Sin embargo aún hay que realizar la búsqueda por esta línea epipolar.

Una vez que se tienen los puntos de correspondencia de ambas imágenes se puede llevar a cabo la reconstrucción tridimensional. A la obtención de puntos en el espacio tridimensional a partir de sus proyecciones en dos imágenes se le conoce como **Triangulación**.

IV.1 Correspondencia Estéreo.

La correspondencia estéreo es el punto de partida para la reconstrucción tridimensional. Consiste en determinar qué par de puntos, proyectados en sendas imágenes, pertenecen al mismo punto en el espacio tridimensional. Conociendo los puntos de correspondencia en las imágenes es como se lleva a cabo la reconstrucción en el enfoque geométrico, y de

la precisión de estos puntos depende la precisión que se obtenga en la reconstrucción. La correspondencia estéreo es un problema considerado entre los más complejos, y ha sido ampliamente estudiado en visión por computadora. La dificultad de éste problema radica en las ambigüedades producidas durante el proceso de adquisición de las imágenes, por ejemplo:

- *Geometría*: Se refiere a las relaciones espaciales entre la escena tridimensional y su proyección en la imagen.
- *Ruido*: Son las inevitables variaciones de luminosidad que producen errores durante la formación de la imagen dentro de la cámara.
- *Falta de textura*: La textura es la propiedad que caracteriza y permite diferenciar una superficie u objeto de otro diferente. La falta de textura se presenta en zonas de superficies u objetos cuyos valores de intensidad de gris o de color (según sea el caso) permanece casi invariante.
- *Oclusión*: Son aquellas áreas que aparecen en una imagen, pero no en la otra, debido a la diferencia de desplazamiento y orientación entre ambas cámaras.

Existen una gran cantidad de técnicas y métodos que abordan este problema, algunos lo resuelven con mejor precisión que otros; entre ellos se pueden citar (Sun (2002); Sun *et al.* (2003); Gruen (1985); Fielding y Kam (2000)), pero no se ha llegado aún a la solución general. De acuerdo a la forma en que abordan el problema, a estos métodos se les puede dividir en dos grandes grupos; locales y globales. A continuación se describirán brevemente los tipos de métodos de correspondencia que existen:

- **Métodos globales.** Éstos asumen ciertas características presentes en las imágenes y resuelven un problema de optimización. Buscan un valor por cada par de puntos de correspondencia que minimice una función de costo global. La función de costo depende de cada método. Este tipo de métodos es menos sensitivo a los problemas de los métodos locales, ya que su modelado proporciona ayuda adicional a regiones que son difíciles de evaluar con una correspondencia local. La principal desventaja de éstos, es que son computacionalmente muy costosos. El costo aumenta drásticamente de acuerdo al tamaño de la imagen.
- **Métodos locales.** Estos utilizan restricciones que se aplican a los píxeles alrededor del píxel (los vecinos) que se desea evaluar. La desventaja de utilizar métodos locales es que generalmente obtienen algunos errores en regiones de la imagen donde la variabilidad de la intensidad de gris de los píxeles es casi nula (baja textura). También ocurre esto en regiones con oclusiones; esto es, píxeles que aparecen en la imagen izquierda, pero que no aparecen en la derecha y viceversa. Finalmente, también ocurre en las regiones fronterizas entre dos objetos proyectados en la imagen. Con todo y esto, los métodos locales son muy utilizados por su rapidez y facilidad de implementación. Dentro de este grupo existe otra división; los métodos basados en características y los métodos basados en áreas.
 - *Características.* Aquí las imágenes se procesan por un operador que extrae las características (bordes, esquinas, segmentos de línea y de curva, segmentos de elipses, círculos y poligonales) que aparecen en ambas imágenes. Después, el proceso de correspondencia se aplica a los atributos asociados con

las características detectadas. Los bordes y esquinas son fáciles de detectar, pero presentan fallos en su correspondencia cuando existen oclusiones. Las demás características requieren gran consumo de recursos computacionales.

- *Áreas.* Estos métodos utilizan una medida de correlación para evaluar la similitud entre un píxel en la imagen izquierda y el posible correspondiente en la derecha. Esta medida se encuentra en función de los valores de intensidad de gris o color, según sea el caso, de un vecindario del píxel en una imagen con un vecindario en la otra imagen.

En el trabajo de Ramírez (Ramirez (2000)) se presenta un método local basado en áreas para el problema de correspondencia, el cual es robusto y consiste en la búsqueda del píxel correspondiente al de la imagen izquierda, a lo largo de la línea epipolar de la imagen derecha. La medida de correlación entre dos píxeles de dos imágenes que este método utiliza se le llama correlación cruzada de norma cero (Zero Normalized Cross Correlation) ZNCC, la cual se define como sigue

$$\frac{\sum_{i=-n}^n \sum_{j=-n}^n (I(x_1 + i, y_1 + j) - \overline{I(x_1, y_1)}) \cdot (I(x_2 + i, y_2 + j) - \overline{I(x_2, y_2)})}{\sqrt{\sum \sum (I(x_1 + i, y_1 + j) - \overline{I(x_1, y_1)})^2 \cdot \sum \sum (I(x_2 + i, y_2 + j) - \overline{I(x_2, y_2)})^2}} \quad (26)$$

Donde $I(x_1, y_1)$ y $I(x_2, y_2)$ representa la intensidad de gris del píxel de coordenadas (x, y) en la imagen izquierda y derecha respectivamente, $\overline{I(x_1, y_1)}$ y $\overline{I(x_2, y_2)}$ es el valor de intensidad promedio del vecindario del píxel (x, y) de la imagen izquierda y derecha. n es el tamaño del vecindario menos 1 entre dos. Por ejemplo, si el tamaño del vecindario es 9×9 , n es igual a 4. Ramírez recomienda una ventana de 11×11 .

El resultado de la ecuación 26 está entre 0 y 1. Así, el píxel correspondiente será aquel que obtenga el resultado más cercano a 1.

En el trabajo de tesis de Pérez (Pérez (2004)) se presenta un estudio detallado del problema de correspondencia, también propone un nuevo método para resolverlo llamado *algoritmo de infección*. Éste combina las propiedades de los dos grupos de métodos, locales y globales. Localmente utiliza el método de ZNCC definido mas arriba, y para evitar las desventajas que este tipo de métodos presenta, desarrolla un novedoso enfoque basado en vida artificial imitando el proceso de la propagación de la infección de una enfermedad, este proceso corresponde al método global.

Mediante un autómata celular, decide cuándo aplicar ZNCC y cuándo establecer el valor de la correspondencia en función de los vecinos ya procesados. Esta decisión permite un 0%, 47%, 70%, 78% o hasta un 99% de ahorro en cálculo de la correspondencia, según se decida. Para mas detalles sobre el funcionamiento de este método véase (Pérez (2004)).

Para obtener los puntos de correspondencia, durante este trabajo de tesis se implementó un algoritmo basado en el algoritmo de infección con un ahorro de procesamiento del 0%. Esto es, todos los píxeles se calcularon aplicando ZNCC. Con respecto al algoritmo original se adicionaron algunos criterios que permiten obtener la correspondencia con mayor precisión. El algoritmo es el siguiente:

Algoritmo 3.1: Realizar lo siguiente por cada píxel de la imagen izquierda $p_1 = (u_1, v_1)$:

1. Calcular la línea epipolar en la imagen derecha: Fp_1

2. Búsqueda del píxel correspondiente en la imagen derecha: Para no buscar sobre toda la línea epipolar en la imagen derecha, se define como píxel de partida (u_1, v_1) , y se define un rango de búsqueda unidireccional r , es decir, comenzar la búsqueda r píxeles antes del de partida aplicando la ecuación 26 a cada uno de éstos, y terminar de buscar en r .
3. Sea p_2 el píxel con el valor de correlación mayor, si éste es mas grande que el umbral de aceptación hacer lo siguiente:
 - (a) Calcular la línea epipolar en la imagen izquierda: Fp_2
 - (b) Búsqueda del píxel correspondiente en la imagen izquierda: Para no buscar sobre toda la línea epipolar en la imagen izquierda, se define como píxel de partida $p_2 = (u_2, v_2)$, y se define un rango de búsqueda unidireccional r , es decir, comenzar la búsqueda desde p_2 y terminar r píxeles después, aplicando la ecuación 26 a cada uno de éstos.
 - (c) Sea p'_1 el píxel con el valor de correlación mayor, si éste cumple con los criterios de aceptación, entonces considerar p_2 como el correspondiente de p_1 .

Los criterios de aceptación de este algoritmo son los siguientes:

- *Rango de Búsqueda.* Éste es unidireccional, no se hace la búsqueda en ambos sentidos de la línea epipolar. De acuerdo a la imagen en la que se esté buscando el píxel correspondiente, la búsqueda es antes o después del píxel de partida. Ya que se trabaja con dos imágenes, siempre se presenta el caso de que las coordenadas

de los píxeles correspondientes de la imagen izquierda son menores en la imagen derecha. Este criterio elimina la mitad de los candidatos, sabiendo de antemano que entre ellos no se encuentra el correspondiente.

- *Umbral de Correlación.* El valor que arroja la ecuación 26 se encuentra entre 0 y 1. Para imágenes de baja textura, como las de la figura 16 se recomienda un valor alto, por ejemplo 0.90. De esta manera se asegura que el correspondiente sea casi idéntico. Aunque es posible que algunos píxeles buenos no sean escogidos por las diferencias de iluminación en las imágenes.
- *Verificación Derecha \leftrightarrow Izquierda.* Al píxel p_2 que cumple con los criterios anteriores, se le somete al mismo proceso que a p_1 para verificar si realmente es su correspondiente. Sólo si el píxel correspondiente de p_2 (p'_1) resulta ser p_1 o un píxel muy cercano a él se acepta a p_2 como el píxel correspondiente de p_1 . Fuera de este rango el píxel correspondiente se rechaza, aunque cumpla con los dos criterios anteriores.

Estos criterios de aceptación son muy importantes, ya que entre más rigurosos sean, la certeza de que los puntos correspondientes sean los correctos aumenta, pero la cantidad de puntos que se obtiene disminuye. Debe establecerse entonces un compromiso entre cantidad y certeza; es decir, buscar un equilibrio de manera que se obtenga la mayor cantidad de puntos correspondientes correctos. Cada par de puntos correspondientes es almacenado, para después calcular a partir de ellos el punto en el espacio 3D mediante un método de triangulación.

Es difícil evaluar cuantitativamente los métodos de correspondencia y poder comparar la efectividad de unos con otros, ya que el resultado es dependiente del tipo y calidad del par estéreo. En la literatura de visión por computadora se han hecho algunos esfuerzos por llevar a cabo la evaluación. El más reciente es el de Scharstein y Szelinski (Scharstein y Szelinski (2002)). Ellos proponen un conjunto de imágenes, que a su juicio, reúnen las características necesarias para hacer una buena evaluación. Esta evaluación consiste en generar un *mapa de disparidad* del par estéreo y obtener sus errores estadísticos con respecto al mapa de disparidad considerado como verdadero (llamado *ground truth*). Para comprender a que se refiere el término mapa de disparidad es necesario conocer el problema de *disparidad estéreo* el cual se desarrolla a continuación.

IV.1.1 Disparidad estéreo.

El problema de disparidad estéreo es una simplificación del problema de correspondencia estéreo; en la cual, el par de imágenes está rectificado; esto es, las líneas epipolares son paralelas a la línea base (ver figura 11). De esta manera, la exactitud de los puntos de correspondencia que se obtienen se puede evaluar a través de un *valor de disparidad*.

El término *disparidad* pertenece al área de investigación de la visión humana (Marr (1982)), y se refiere a la diferencia en la ubicación de los objetos que se están viendo con el ojo izquierdo y derecho. En visión por computadora, la disparidad es el término para referirse al inverso de la profundidad; la cual se puede medir como la distancia que existe entre la coordenada en que un punto del espacio 3D se proyecta en la primera imagen y la coordenada de la otra imagen.

El mapa de disparidad es una imagen artificial que se forma a partir del par estéreo (ver figura 22). Las coordenadas (u_d, v_d) de los píxeles de esta imagen coinciden con las coordenadas de una de las dos imágenes del par, que se toma como referencia; por lo general la referencia es la imagen izquierda del par. Así, $u_d = u_1$ y $v_d = v_1$. El valor de disparidad, que será la intensidad del nivel de gris que se coloque en cada píxel de esta imagen, se define como la distancia que existe entre un píxel de la imagen izquierda y su correspondiente en la derecha

$$d_{(u_d, v_d)} = v_1 - v_2 \quad (27)$$

La disparidad refleja el inverso de la profundidad. Así, para un punto que se encuentre a menor profundidad (esto es, que está más cercano a las cámaras) el valor de disparidad es mayor, y viceversa. De esta manera, en el mapa de disparidad se distinguen los objetos por su profundidad. Si un píxel en el mapa, que forma parte de la proyección de un objeto fue producto de una equivocación al asignar los puntos de correspondencia, su valor de disparidad será muy diferente al de los demás píxeles de ese objeto. Esta característica ha motivado a Scharstein y Szelinski a utilizar el mapa de disparidad como medida de evaluación en su trabajo para clasificar y calificar a los diferentes métodos de correspondencia estéreo. Sin embargo, la evaluación sólo es aplicable en imágenes rectificadas, por lo que no aplica para las imágenes de la figura 16. Sin embargo, en la sección IV.3 se hicieron pruebas para generar un mapa de disparidad con otras imágenes que cumplen con esta restricción, estas imágenes son las que utilizan Scharstein y Szelinsky en su evaluación y están disponibles en www.middlebury.edu/stereo. En la figura 22 se muestran algunas de estas imágenes

junto con el mapa de disparidad obtenido a partir de ellas.

Es necesario decir que la evaluación de los métodos de correspondencia estéreo mediante el mapa de disparidad es incompleta, pues sólo se ocupa de la evaluación específica del problema de disparidad estéreo. Otra forma de evaluar la exactitud de los puntos de correspondencia es generar el modelo tridimensional euclidiano, y percatarse de qué tan fidedigno resulta, con respecto a la escena real. Éste también se lleva a cabo en la sección IV.3.

IV.2 Triangulación.

La triangulación consiste en obtener un punto a partir de su proyección en las dos imágenes de una escena; es decir, a partir de los puntos de correspondencia. Formalmente se puede definir como sigue (ver figura 18):

Sean $P=(X,Y,Z,1)$ un punto proyectado en dos imágenes $p_1 = (u_1, v_1, w_1)$ y $p_2 = (u_2, v_2, w_2)$. La triangulación consiste en encontrar la intersección de los dos rayos que parten de los centros de proyección de las cámaras C_1 y C_2 y pasan por p_1 y p_2 hacia el espacio.

Para llevar a cabo la triangulación y conseguir una reconstrucción euclidiana, es necesario conocer los puntos de correspondencia, su matriz fundamental y los parámetros de calibración de la cámara. Conociendo todo esto, la intersección parecería trivial. Sin embargo debido a ruido y a errores numéricos tanto en la resolución de las cámaras como en los métodos de calibración y de correspondencia, la intersección de ambos rayos no siempre se da. Existen varios métodos que resuelven este problema y a continuación se presentan:

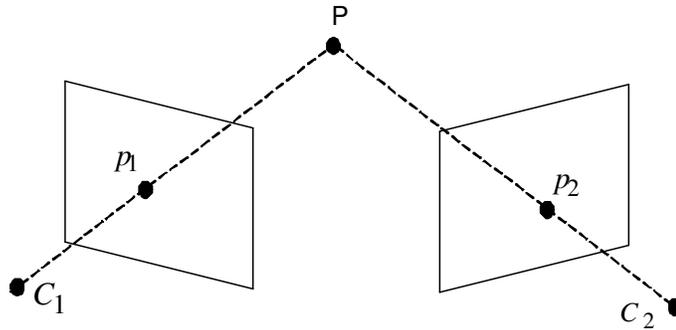


Figura 18: Triangulación.

IV.2.1 Método del punto Medio.

Este método busca el punto medio geométrico entre ambos rayos (figura 19). Los píxeles p_1 y p_2 deben estar expresados en coordenadas de la cámara: (x, y, z) . El primer rayo que pasa por p_1 se define como αp_1 . Y el que pasa por p_2 como $t + \beta R^T p_2$; esto es, el punto p_2 expresado en el sistema coordenado de p_1 .

Donde t es un vector de translación y R^T es una matriz de rotación que representan la transformación rígida de la cámara dos a la cámara uno. De esta manera se busca un vector w que intersekte a ambos rayos. Se busca que este vector sea lo mas pequeño posible, es decir que intersekte a los rayos en la distancia más corta que exista entre uno y otro. Por lo tanto el vector w debe ser ortogonal a ambos: $w = (p_1 \times R^T p_2)$. Así el método del punto medio se define como

$$\alpha p_1 + \gamma (p_1 \times R^T p_2) = t + \beta R^T p_2. \quad (28)$$

Este es un sistema de tres ecuaciones lineales (ya que p_1 y p_2 son vectores de tres

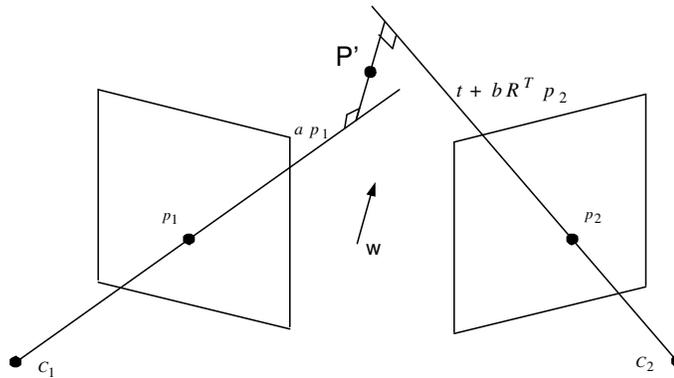


Figura 19: Método del punto medio.

coordenadas) y tres incógnitas (α, β, γ) el cual puede ser resuelto con cualquier método numérico que resuelva sistema de ecuaciones lineales, como el método de Gauss. Una vez obtenida la solución numérica de las incógnitas se sustituyen en la ecuación y el resultado se divide entre dos para obtener el punto medio; y éste será el punto triangulado P.

IV.2.2 Método lineal.

Este método se basa en las matrices de los parámetros de las cámaras (ecuación 5)

$$\begin{pmatrix} s_u \\ s_v \\ s \end{pmatrix} = \begin{pmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & m_{34} \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \rightarrow p = MX$$

Recuérdese que s representa el factor de escala sobre las coordenadas de la imagen y que $(s_u, s_v, s)^T = (u, v, 1)^T$ representa la simplificación de las coordenadas de un píxel

que no toma en cuenta el factor de escala ya que éste se encuentra presente en toda la imagen.

Para explicar el método lineal se tomará en cuenta el factor de escala sustituyendo $(u, v, 1)^T$ por $(wu, wv, w)^T$, donde w representa un factor de escala que actúa sobre las coordenadas de la imagen u y v . Quedando la ecuación (5) de la siguiente manera

$$\begin{pmatrix} wu \\ wv \\ w \end{pmatrix} = \begin{pmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & m_{34} \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}$$

Ahora bien, cada renglón de la matriz M puede representarse como vectores por separado como m_i^T , donde i representa al i -ésimo renglón. Así es posible manejar por separado cada ecuación lineal

$$wu = m_1^T X \quad (29)$$

$$wv = m_2^T X \quad (30)$$

$$w = m_3^T X \quad (31)$$

Por cada cámara se tiene un sistema de 3 ecuaciones lineales. Ya que se tienen dos cámaras, el sistema se compone de 6 ecuaciones lineales con 3 incógnitas, pero en realidad, debido a w , la última ecuación de cada sistema individual es linealmente dependiente de los otros dos. Sustituyendo la ecuación (31) en las ecuaciones (29) y (30) se obtiene

$$\begin{aligned} um_3^T X &= m_1^T X \\ vm_3^T X &= m_2^T X \end{aligned}$$

Por lo tanto se tienen 4 ecuaciones linealmente independientes con 3 incógnitas de la forma $Ax = 0$

$$\begin{pmatrix} u_1 m_3^{1T} - m_1^{1T} \\ v_1 m_3^{1T} - m_2^{1T} \\ u_2 m_3^{2T} - m_1^{2T} \\ v_2 m_3^{2T} - m_2^{2T} \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} = 0 \quad (32)$$

Este sistema puede ser resuelto mediante descomposición en valores singulares SVD: $A = UDV^T$ (Press *et al.* (1994)) siendo el resultado x igual a la última columna de V .

IV.2.3 Método Polinomial.

En realidad este método no es un método de triangulación *per se*, pues no realiza la reproyección de las imágenes al espacio. Más bien realiza una corrección en los píxeles de manera que los rayos intersecten en el espacio. Este método define el píxel \hat{p}_1 en la imagen izquierda más cercano a p_1 , que cumpla con la restricción epipolar (ecuación 13), de la misma manera define un píxel \hat{p}_2 para p_2 . Una vez definidos estos píxeles, los autores del método aconsejan aplicar el método lineal (Hartley y Sturm (1997)) para triangular los nuevos píxeles \hat{p}_1 y \hat{p}_2 y así, debido a la geometría epipolar, garantizar que los rayos se intersectan en el espacio. El método polinomial utiliza un criterio de optimización, pues partiendo de la idea de que los píxeles originales no intersectarán en el espacio, busca el píxel más cercano a p_1 en la imagen izquierda y a p_2 en la imagen

derecha que cumplan con la restricción epipolar. Matemáticamente se define como:

$$MIN d(p_1, \hat{p}_1)^2 + d(p_2, \hat{p}_2)^2,$$

$$s.a. p_2^T F p_1 = 0$$

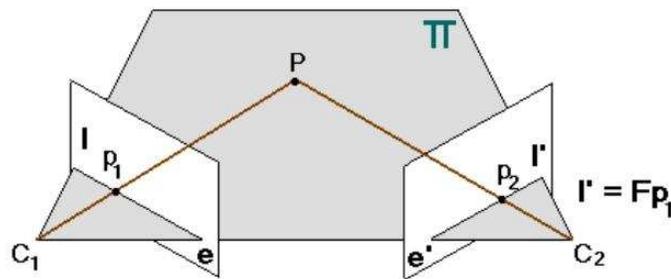


Figura 20: Método polinomial.

La restricción epipolar dicta que los píxeles \hat{p}_1 y \hat{p}_2 se encuentran en un mismo plano y así se garantizaría la intersección de los rayos que salen de los píxeles (figura 20), ver (Hartley y Sturm (1997)) para mas detalles. Este criterio puede reducirse aún mas tomando en cuenta que, según la geometría epipolar, un par de puntos correspondientes, deben estar dentro de un par de líneas epipolares correspondientes.

Así, el píxel \hat{p}_1 cae dentro de la línea epipolar $l = F^T \hat{p}_2$ y \hat{p}_2 en la línea $l_2 = F \hat{p}_1$ (figura 21) y el problema de optimización definido arriba queda de la siguiente manera

$$\min C = d(p_1, l(t))^2 + d(p_2, l_2(t))^2 \quad (33)$$

Esto es: *Encontrar la línea epipolar $l_1(t)$ más cercana a p_1 y $l_2(t)$ mas cercana a p_2 .*

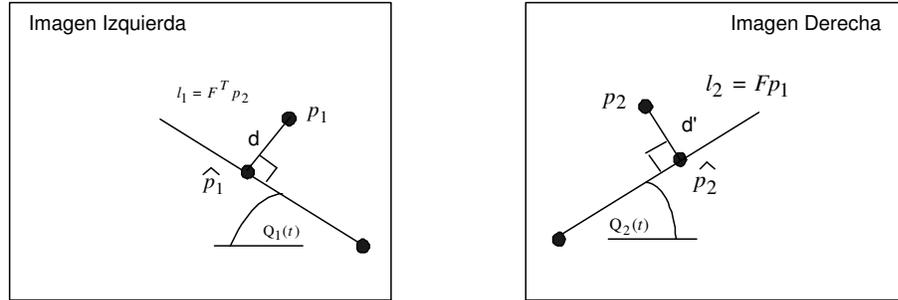


Figura 21: Reformulación del problema de optimización.

Detalles de la minimización.

$l_1(t)$ y $l_2(t)$ representan el conjunto de líneas epipolares que atraviesan ambas imágenes, las cuales se encuentran parametrizadas por t (figura 21). Este parámetro, se utiliza para diferenciar entre las líneas epipolares candidatas a ser la más cercana a los puntos correspondientes. Con el problema de optimización planteado de esta manera, la búsqueda de la mejor línea epipolar se puede convertir en un polinomio en función de t y así evitar el proceso de minimizado.

Para simplificar el trabajo, se debe trasladar p_1 y p_2 al origen del sistema coordenado de cada imagen y los epipolos de cada imagen e_1 y e_2 al eje equis; estos es, $p_1 = p_2 = (0, 0, 1)^T$, $e_1 = (1, 0, f)^T$, $e_2 = (1, 0, f')^T$. Si cambian los epipolos, la matriz fundamental también cambia para cumplir con la condición $F e_1 = e_2^T F = 0$

$$F = \begin{pmatrix} f f' d & -f' c & -f' d \\ -f b & a & b \\ -f d & c & d \end{pmatrix}$$

Ahora bien, la línea epipolar $l_1(t)$ está definida por dos puntos: el pixel $(0, t, 1)^T$ y

el epipolo $(1, 0, f)^T$ mediante su producto cruz: $(0, t, 1) \times (1, 0, f) = (tf, 1, -t)$. De aquí se puede inferir que la distancia al cuadrado de p_1 a $l_1(t)$ es

$$d(p_1, l_1(t))^2 = \frac{t^2}{1 + (tf)^2}$$

Usando la geometría epipolar para encontrar la línea epipolar $l_2(t)$ correspondiente en la otra imagen, se puede ver que

$$l_2(t) = F\hat{p}_1 = F(0, t, 1)^T = (-f'(ct + d), at + b, ct + d)^T$$

Esta es la representación de $l_2(t)$ como un vector homogéneo. La distancia al cuadrado de esta línea a p_2 es

$$d(p_2, l_2(t))^2 = \frac{(ct + d)^2}{(at + b)^2 + f'^2(ct + d)^2}$$

De esta manera el criterio de optimización se ha convertido en un polinomio en función de t (Hartley y Sturm (1997))

$$d(p_1, l_1(t))^2 + d(p_2, l_2(t))^2 = s(t) = \frac{t^2}{1 + t^2 f^2} + \frac{(ct + d)^2}{(at + b)^2 + f'^2(ct + d)^2}$$

Para encontrar el mínimo, se pueden usar técnicas de cálculo elemental, comenzando con la obtención de la primera derivada del polinomio

$$s'(t) = \frac{2t}{(1 + t^2 f^2)^2} + \frac{2(ad - bc)(at + b)(ct + d)}{((at + b)^2 + f'^2(ct + d)^2)^2}$$

Los máximos y mínimos del polinomio se obtienen igualando la primera derivada a cero. Juntando los dos términos en $s'(t)$ bajo un común denominador e igualando el

numerador a 0 se obtiene lo siguiente

$$g(t) = t((at + b)^2 + f'^2(ct + d)^2) - (ad + bc)(1 + f^2t^2)^2(at + b)(ct + d) = 0$$

Si se desarrolla $s(t)$ se verá que es un polinomio de grado 6, el cual tiene a lo más 6 raíces reales, correspondiendo a 3 máximos y 3 mínimos de la función polinomial. El mínimo global puede encontrarse sacando las raíces en $g(t)$ y evaluándolas en $s(t)$.

El mínimo global representa el valor que t debe tomar para definir las líneas epipolares $l_1(t)$ y $l_2(t)$ que serán tomadas en cuenta. Teniendo estas líneas, se define como \hat{p}_1 y \hat{p}_2 al píxel dentro de cada línea más cercano a p_1 y p_2 respectivamente para llevar a cabo la triangulación con el método lineal.

IV.3 Experimentación.

Se procede ahora a mostrar en la práctica la teoría que se planteó en este capítulo sobre correspondencia estéreo y triangulación.

IV.3.1 Correspondencia estéreo.

Se han desarrollado dos pruebas diferentes para comprobar la efectividad del algoritmo 3.1. Uno es generar los mapas de disparidad y la otra es generar el modelo tridimensional.

Mapa de Disparidad.

Se realizaron pruebas con cuatro pares de imágenes propuestas por Scharstein y Szelinsky: *Tsukuba*, *Sawtooth*, *Venus*, *Map* (ver figuras 22 y 23), de las cuales se obtuvieron

sus mapas de disparidad y se sometieron a la evaluación disponible en el sitio web de la universidad Middlebury www.middlebury.edu/stereo. En esta evaluación, la calidad del mapa se determina obteniendo el porcentaje de píxeles incorrectos, comparados con un mapa de referencia que contiene las disparidades reales, de acuerdo la siguiente expresión

$$B = \frac{1}{N} \sum_{(x,y)} (|d_C(x,y) - d_R(x,y)| > \delta)$$

Donde $d_C(x,y)$ representa el valor de disparidad calculado para la coordenada (x,y) , y $d_R(x,y)$ representa su valor de disparidad real; δ es la tolerancia, que para estas pruebas fue de 1 píxel. Finalmente, N es el número de píxeles que se evalúan. Los resultados de las pruebas se muestran en la tabla IV. En dicha tabla se evalúan dos de los problemas más comunes para un algoritmo de correspondencias:

- *Regiones sin textura (S.T.)*. Regiones cuya variabilidad en el nivel de gris es menor a cierto umbral.
- *Regiones con discontinuidad (D)*. Píxeles cuyos vecinos presentan una disparidad que difiere de la suya en más de la permitida por un umbral. Estas regiones se presentan en las fronteras de los objetos, donde se da un cambio drástico entre la disparidad del objeto más cercano con respecto al que está más alejado.

También se presenta el resultado de evaluar el total de la imagen (\mathcal{T}), pero sin tomar en cuenta las regiones con oclusión (aquellas que aparecen en una de las imágenes, pero no aparecen en la otra), ya que los autores de la prueba decidieron no tomarlas en cuenta, porque la mayoría de los algoritmos de correspondencia no las contemplan.

Tabla IV: Resultados que se obtuvieron al evaluar el algoritmo 3.1 con la prueba de la universidad Middlebury. Nótese que *Map* no tiene la columna $\mathcal{S.T.}$ porque esta imagen no contiene regiones sin textura.

	<i>Tsukuba</i>			<i>Sawtooth</i>			<i>Venus</i>			<i>Map</i>	
	\mathcal{T}	$\mathcal{S.T.}$	\mathcal{D}	\mathcal{T}	$\mathcal{S.T.}$	\mathcal{D}	\mathcal{T}	$\mathcal{S.T.}$	\mathcal{D}	\mathcal{T}	\mathcal{D}
Algorit. 3.1	8.90	7.64	42.48	5.79	4.87	34.20	5.33	6.60	41.30	3.33	32.43

Los números representan el porcentaje de error del total de píxeles de cada región. En cuanto al total de la imagen, puede verse que el algoritmo presenta un buen rendimiento, pues alcanza niveles bajos de error, siendo la de menor porcentaje *Map*, con 3.33% y la mayor porcentaje *Tsukuba*, con 8.90%. Algo similar sucede con las regiones sin textura, donde el menor porcentaje se obtiene en *Sawtooth*, con 4.87%; mientras que el mayor es para *Tsukuba*, con 7.64%. Sin embargo para las regiones con discontinuidad el rendimiento es regular, pues como se dijo anteriormente, para los métodos locales basados en áreas es difícil lidiar con las fronteras de los objetos. *Map* de nueva cuenta es la de menor porcentaje con 32.43%, y *Tsukuba* la de mayor, con 42.48%.

Recientemente, Middlebury publicó una evaluación más estricta, donde elimina *Sawtooth* y *Map* y coloca dos nuevos pares de imágenes: *Cones* y *Teddy* (ver figura 23). Además contiene nuevos criterios de evaluación. La columna \mathcal{T} la han renombrado para describir mejor su función: *regiones sin oclusión* $\mathcal{S.O.}$. La columna $\mathcal{S.T.}$ desaparece, ocupando su lugar la nueva \mathcal{T} , que evalúa toda la imagen, incluyendo regiones con poca oclusión. La columna \mathcal{D} sigue ahí, pero además de evaluar las regiones con discontinuidad, evalúa regiones con oclusión. En la tabla V se muestran los resultados con la nueva evaluación. Nótese que los valores de *Tsukuba* y *Venus* no son necesariamente los mismos que los de la tabla IV. Para la primera columna el menor porcentaje de

Tabla V: Resultados que se obtuvieron al evaluar el algoritmo 3.1 con la nueva versión de la prueba de Middlebury.

	<i>Tsukuba</i>			<i>Venus</i>			<i>Teddy</i>			<i>Cones</i>		
	<i>S.O.</i>	<i>T</i>	<i>D</i>	<i>S.O.</i>	<i>T</i>	<i>D</i>	<i>S.O.</i>	<i>T</i>	<i>D</i>	<i>S.O.</i>	<i>T</i>	<i>D</i>
Algorit. 3.1	8.90	10.5	41.6	5.33	6.42	42.5	19.0	25.3	45.6	15.2	21.9	39.3

error fue para *Venus*, con 5.33%, mientras que el mayor porcentaje lo obtuvo *Teddy*, con 19%. Lo mismo ocurre para la segunda columna, *Venus* con 6.42% de error y *Teddy* con 25.3%. La tercera es la más difícil, pues evalúa la discontinuidad y la oclusión, aquí el menor porcentaje lo obtuvo *Cones*, con 39.3%, mientras que el mayor fue para *Teddy* 45.6%.

Modelo tridimensional.

Otra manera de evaluar la efectividad del algoritmo 3.1 es triangular el conjunto de puntos de correspondencia que arroja y comprobar que el modelo tridimensional de la escena es coherente. La prueba consiste en guardar el conjunto de puntos de correspondencia que arroja este algoritmo en un archivo, para luego introducirlos como entrada al método de triangulación lineal. Ambos fueron implementados durante este trabajo de tesis en el programa *xcv*. La salida del método lineal es un modelo tridimensional obtenido a partir de las imágenes originales, en un formato que permite visualizarlo en tres dimensiones. Este formato es el conocido como VRML (Virtual Reality Modeling Language). En la figura 24 se muestran cuatro imágenes tomadas de cuatro modelos con diferentes valores en su criterio de aceptación. **R** corresponde al rango de búsqueda en píxeles, **U** es el umbral de correlación y **V** representa el número de píxeles que se dió de holgura en la verificación *Derecha* \leftrightarrow *Izquierda*.

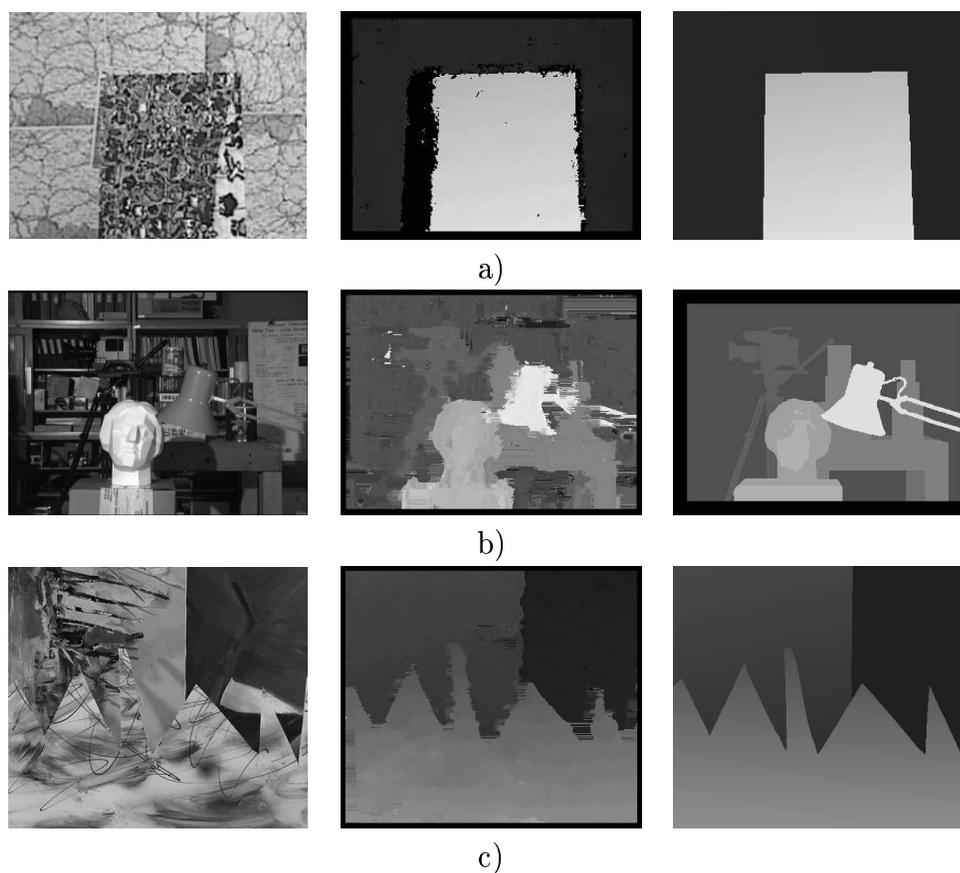


Figura 22: Imágenes de prueba de la universidad de Middlebury, sus mapas de disparidad obtenidos aplicando el algoritmo 3.1 y el mapa de disparidad con los valores reales. a) Map. b) Tsukuba. c) Sawtooth.

En la imagen, los huecos que aparecen representan puntos descartados por los criterios de aceptación del método de correspondencia. Nótese cómo en a) aparecen más puntos que en las demás imágenes; pero al girar el modelo tridimensional, se puede apreciar que existen muchos puntos fuera de lugar, principalmente en la parte de la mesa y en la pared del fondo, izquierda. Aparecen más cerca o más lejos de lo que

deberían. La imagen del inciso b) presenta una cantidad menor de puntos triangulados, Se puede apreciar que con respecto al inciso a), muchos puntos que pertenecen al tejado de la casa y de la mesa ya no aparecen. Es decir, se descartan puntos en regiones de color uniforme (baja textura), que es donde se presentan los errores de correspondencia. En el inciso c), continúa este proceso quitando puntos en zonas de baja textura como el interior de los cuadrados en la mira de calibración y las ventanas de la casa. Ya en el inciso d), desaparece por completo la mesa y de la mira de calibración y la casa solo quedan los bordes. Al girar el modelo tridimensional, en formato VRML se aprecia que la gran mayoría de los puntos que quedaron, aunque son pocos, están correctamente triangulados.

IV.3.2 Triangulación.

Durante el proceso de experimentación para triangulación se eligió un método de entre los dos propuestos anteriormente: punto medio y lineal. El método del punto medio es muy simple, sin embargo realiza la reconstrucción bajo el eje coordenado de la cámara izquierda, es decir, las coordenadas de la escena estarían dadas de acuerdo al eje coordenado de esta cámara; y no de acuerdo a las coordenadas de la mira de calibración, que es el sistema de referencia de la escena. Este problema podría resolverse mediante una transformación euclidiana (rotación y translación) de todos los puntos triangulados de un eje coordenado al otro; sin embargo, el método del punto medio fue diseñado para reconstrucciones euclidianas. Por esta razón se prefiere el método lineal, que es general y también es sencillo de implementar; pues en trabajos futuros se pretende realizar reconstrucciones afines y proyectivas.

IV.3.3 Triangulación lineal.

Se han realizado pruebas con el programa desarrollado `xcv`, aplicando el método lineal con y sin el método polinomial sobre las imágenes de la figura 16 y 25, siendo la salida un modelo tridimensional en formato VRML. En la figura 24 se pueden apreciar los resultados. En general, los resultados son cualitativamente buenos, pues una buena parte de los puntos triangulados son coherentes con la realidad. Se aprecia en el modelo que aparecen todos los objetos que conforman la escena y que además guardan la proporción que se muestra en las imágenes. Aunque en el inciso d), donde la gran mayoría de los puntos son correctos la información que refleja es pobre; difícilmente se distinguen los objetos que forman la escena. En cuanto a la evaluación cuantitativa del modelo tridimensional, es difícil medirla con respecto a la escena original si no se tienen las dimensiones y coordenadas exactas de ésta última. En los dos pares de imágenes que se utilizaron como prueba se conocen las coordenadas de un objeto: la mira de calibración. La cual actúa precisamente como sistema coordenado de referencia en la escena. Así, se conoce la dimensión y coordenadas de cada uno de los cuadrados que aparecen dentro de la mira.

Para fines prácticos se eligieron los puntos de las esquinas de estos cuadrados para medir el grado de error del modelo con respecto a la escena. En las tablas VI y VII se muestran 12 de estas esquinas, aparecen sus coordenadas dentro de la escena y las coordenadas obtenidas por el modelo. Dicho modelo fue obtenido triangulando el conjunto de puntos de correspondencia del inciso b) de la figura 24. Todas las unidades están expresadas en milímetros. Como medida de error se usa la distancia euclidiana entre las coordenadas del modelo M y las coordenadas de la escena E

Tabla VI: Medida de error en el modelo tridimensional de las imágenes de la figura 16.

X Escena	Y Escena	Z Escena	X Modelo	Y Modelo	Z Modelo	Dif. X	Dif. Y	Dif. Z	Dist. Euclidiana
175	0	0	175.59	-2.11	1.83	0.59	2.11	1.83	2.86
175	0	105	171.32	-0.53	104.15	3.68	0.53	0.85	3.81
175	0	225	168.12	-5.55	224.04	6.88	5.55	0.96	8.89
70	0	0	69.78	-1.56	0.75	0.22	1.56	0.75	1.75
70	0	105	69.41	-3.2	103.32	0.59	3.2	1.68	3.66
70	0	225	67.57	-5.72	225.04	2.43	5.72	0.04	6.21
0	70	0	5.83	76.7	0.05	5.83	6.7	0.05	8.88
0	70	105	-3.89	66.1	104.35	3.89	3.9	0.65	5.55
0	70	225	5.37	74.81	225.56	5.37	4.81	0.56	7.23
0	175	0	14.76	179.95	-0.3	14.76	4.94	0.3	15.57
0	175	105	14.5	180.3	105.42	14.5	5.3	0.42	15.44
0	175	225	12.18	178.89	225.07	12.18	3.89	0.07	12.78

$$E = \sqrt{(X_E - X_M)^2 + (Y_E - Y_M)^2 + (Z_E - Z_M)^2}$$

Nótese que este error es muy pequeño, el mayor es de 15.57 milímetros en la primera tabla y de 16.52 milímetros en la segunda.

IV.3.4 Triangulación polinomial.

Otro aspecto a probar de lo que se ha mostrado durante este capítulo es lo concerniente al método polinomial. Como se explicó, el método polinomial no es un método de triangulación propiamente, más bien es un método de corrección de los puntos que se van a triangular. La prueba consiste en aplicar sobre el conjunto de puntos correspondientes el método polinomial antes de introducirlos en el método lineal. Una vez más, como medida cuantitativa, se utilizó la mira de calibración. En la figura 26 se ilustra una

Tabla VII: Medida de error en el modelo tridimensional de las imágenes de la figura 25.

X Escena	Y Escena	Z Escena	X Modelo	Y Modelo	Z Modelo	Dif. X	Dif. Y	Dif. Z	Dist. Euclidiana
175	0	0	174.27	-4.71	0.13	0.73	4.71	0.13	4.76
175	0	105	179.51	3.39	104.58	4.51	3.39	0.42	5.66
175	0	225	164.43	-12.58	225.09	10.57	12.58	0.09	16.43
70	0	0	70.95	0.3	-0.93	0.95	0.3	0.93	1.36
70	0	105	78.62	5.76	105.35	8.62	5.76	0.35	10.37
70	0	225	56.32	-9.25	224.52	13.68	9.25	0.48	16.52
0	70	0	0.5	70.85	-0.54	0.5	0.85	0.54	1.13
0	70	105	9.83	75.92	105.25	9.83	5.92	0.25	11.48
0	70	225	-11.22	58.56	225.44	11.22	11.44	0.44	16.03
0	175	0	2.74	176.98	-0.03	2.74	1.98	0.03	3.38
0	175	105	9.01	185.7	102.32	9.01	10.7	2.68	14.25
0	175	225	-8.19	169.37	225.21	8.19	5.63	0.21	9.94

comparación entre el error (la distancia euclidiana) de los mismos 12 puntos aplicando el método polynomial y no aplicándolo. En el inciso a) se presenta la gráfica de la prueba con el par estéreo de la figura 16 y en el inciso b) con el par estéreo de la figura 25.

En ambas gráficas, puede verse que el error obtenido es casi el mismo para el caso de aplicar el método polinomial y para el caso en que no se aplica. La corrección que realiza el método polinomial es ínfima. Se esperaba un resultado así, pues habiendo obtenido la matriz fundamental a partir del proceso de calibración (ver capítulo 2) el margen de error en las matrices de proyección de cada cámara y en consecuencia de la matriz fundamental debería ser muy pequeño.

En cuestiones de resultados, el aplicar el método polinomial en una reconstrucción euclidiana pudiera parecer irrelevante, no así en cuestiones de recursos computacionales.

En las pruebas realizadas el proceso de reconstrucción, que incluye la correspondencia estéreo y la triangulación se realiza en 21 minutos, 4 segundos sin aplicar el método polinomial, mientras que aplicandolo se lleva a cabo en 25 minutos, 48 segundos; en una computadora con procesador Pentium III, a 800 MHZ, con 64 MB de memoria RAM.

IV.4 Conclusiones.

En este capítulo se mostró el desarrollo del enfoque geométrico para resolver el problema de reconstrucción tridimensional a partir de dos imágenes: la triangulación. También se habló del problema del cual depende la triangulación: la correspondencia estéreo.

De la correspondencia estéreo se puede decir que es muy susceptible a errores, principalmente en regiones de baja textura, en regiones de una imagen que no aparece en otra (oclusión) y en los bordes de los objetos (discontinuidades), esto provoca que se asigne un píxel correspondiente que en realidad no es el correcto. Ya que los errores en estas regiones no se pueden evitar, los criterios de aceptación descartan aquellos puntos que caen dentro de estas regiones. El criterio de *umbral de aceptación* es muy útil en las regiones de baja textura y el criterio de *verificación izquierda \longleftrightarrow derecha* lo es en las regiones con oclusión y discontinuidad. El criterio de rango de búsqueda sólo acota el espacio donde se busca el píxel correspondiente y evita que la búsqueda se prolongue demasiado.

Así, entre más rigurosos sean los criterios de aceptación los puntos de correspondencia que se obtengan serán exactos. Esto es por demás importante, ya que se ha demostrado que de ello depende en gran medida el éxito de la triangulación. Pues al proyectar los rayos hacia el espacio, estos se cruzan en un lugar erróneo, dando como

resultado lo que se ve en la figura 27, en donde los criterios de aceptación fueron menos rigurosos. La imagen del modelo tridimensional fue hecha desde un punto lejano, para ver toda la escena. Obsérvese que ésta se ve *estirada* hacia atrás, llega un punto en el que converge y vuelve a salir. Esto se debe a que un error tan pequeño, como la asignación del píxel contiguo al verdadero correspondiente, puede significar un gran error en el modelo tridimensional. Ese píxel podría ser la proyección de un objeto más lejano o más cercano al que corresponde. De ahí la importancia de que los criterios de aceptación en el algoritmo de los puntos correspondientes se apliquen rigurosamente, aunque al final se genere un modelo tridimensional con algunos huecos, como en la figura 24. Con los resultados en la evaluación externa del método de correspondencia que se está usando por parte de la universidad de Middlebury, se puede concluir que este método es competitivo y puede usarse como base para realizar la triangulación. Sin embargo, el método tiene problemas en zonas de baja textura.

Sobre la decisión acerca de qué método de triangulación utilizar, se ha elegido el método lineal sobre el del punto medio, ya que este último realiza la reconstrucción bajo el eje coordenado de la cámara izquierda, es decir, las coordenadas de la escena estarían dadas de acuerdo al sistema coordenado de esta cámara; y no de acuerdo al sistema coordenado de la mira de calibración. Y aunque este problema se podría resolver aplicando una rotación y translación a cada punto triangulado para convertirlo de un sistema coordenado al otro; el método del punto medio fue diseñado para reconstrucciones euclidianas. Se prefiere entonces, usar el método lineal, que es general; pues en trabajos futuros se pretende realizar reconstrucciones afines y proyectivas.

El tema sobre el que falta concluir en este capítulo es la efectividad del método

polinomial. Según la literatura (Hartley y Zisserman (2004); Hartley y Sturm (1997)), éste método es muy utilizado para realizar la triangulación y es sumamente efectivo si se aplica antes del método lineal en reconstrucciones proyectivas, pero en dichos trabajos anteriores no se habla acerca de su efectividad en reconstrucciones euclidianas. Por los resultados obtenidos durante este trabajo se puede decir que no representa mejoras significativas al utilizarlo en reconstrucciones euclidianas, pero no por deficiencia del método, mas bien, se debe a que no es posible corregir los puntos correspondientes, pues se tiene la certeza de que cumplen con la geometría epipolar (que es en lo que se basa la corrección del método polinomial), ya que fueron obtenidos bajo un enfoque calibrado. Además, si se toma en cuenta que aplicar el método polinomial implica mayor consumo de tiempo, se puede concluir que es mejor no utilizarlo para llevar a cabo una reconstrucción euclidiana.

En este capítulo se ha analizado el método geométrico para resolver el problema de reconstrucción, el cual se compone de dos procesos: correspondencia estéreo y triangulación. Ambos son importantes, pero la correspondencia es pieza fundamental para el éxito de la triangulación. Si existen errores de correspondencia, por muy pequeños que sean, éstos se propagan en grado superlativo. La opción que se presentó hasta ahora fue eliminar los errores aumentando la rigidez de los criterios de aceptación, sin embargo el modelo tridimensional que resulta posee poca información de la escena y ésta, no necesariamente es información de interés. Con esta idea en mente, en los siguientes capítulos se propone un enfoque diferente, el cual no lleva a cabo la correspondencia ni la triangulación, en cambio, utiliza como herramienta los algoritmos evolutivos y reconoce de manera inteligente la información de la escena que puede ser relevante,

concentrándose en ella para arrojar un modelo tridimensional.

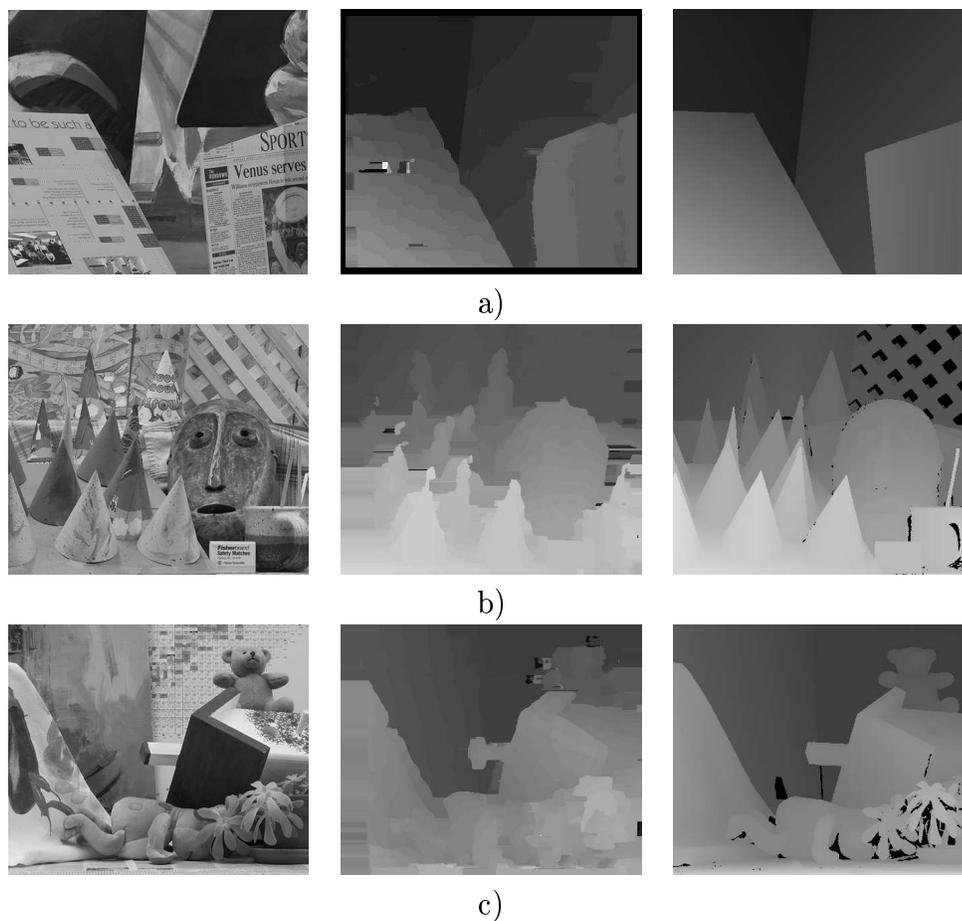


Figura 23: Imágenes de prueba de la universidad de Middlebury, sus mapas de disparidad obtenidos aplicando el algoritmo 3.1 y el mapa de disparidad con los valores reales. a) Venus. b) Cones. c) Teddy.

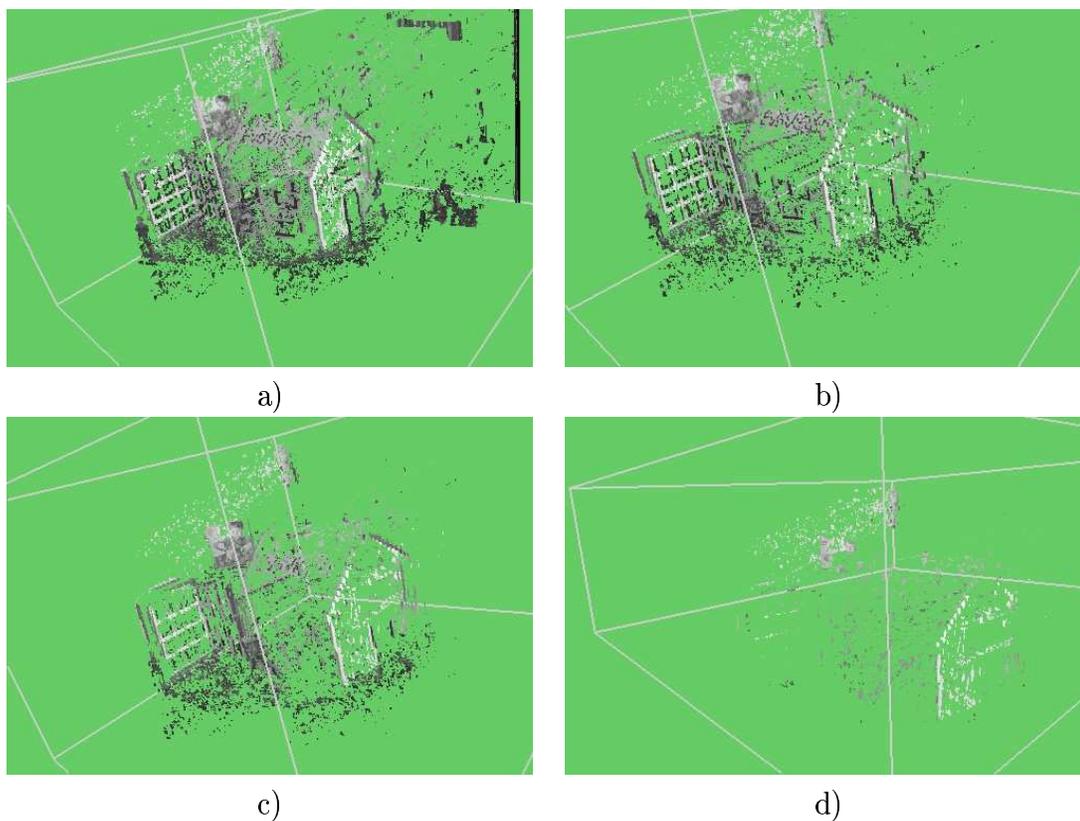


Figura 24: Triangulación al aplicar diferentes valores en los criterios de aceptación en el algoritmo de correspondencia. a) $R = 100$ pix. $U = 0.80$ $V = \pm 5$ pix. b) $R = 100$ pix. $U = 0.85$ $V = \pm 3$ pix. c) $R = 100$ pix. $U = 0.90$ $V = \pm 3$ pix. d) $R = 100$ pix. $U = 0.95$ $V = \pm 0$ pix.

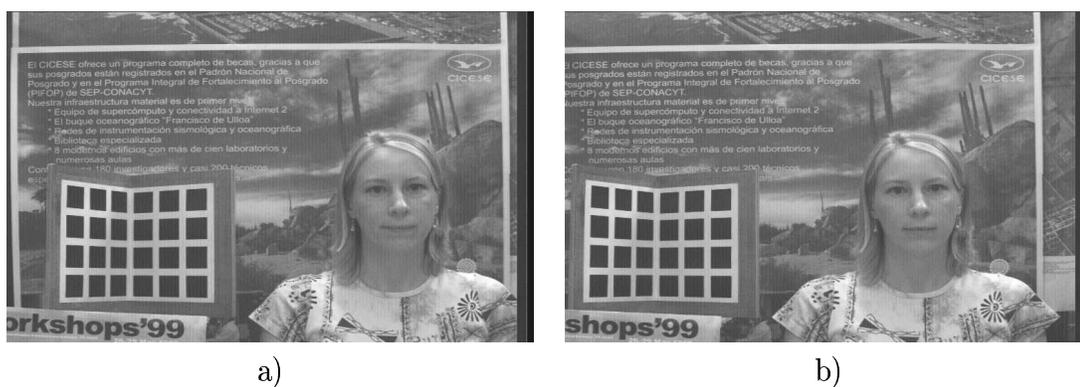


Figura 25: Segundo par de imágenes utilizadas para realizar pruebas.

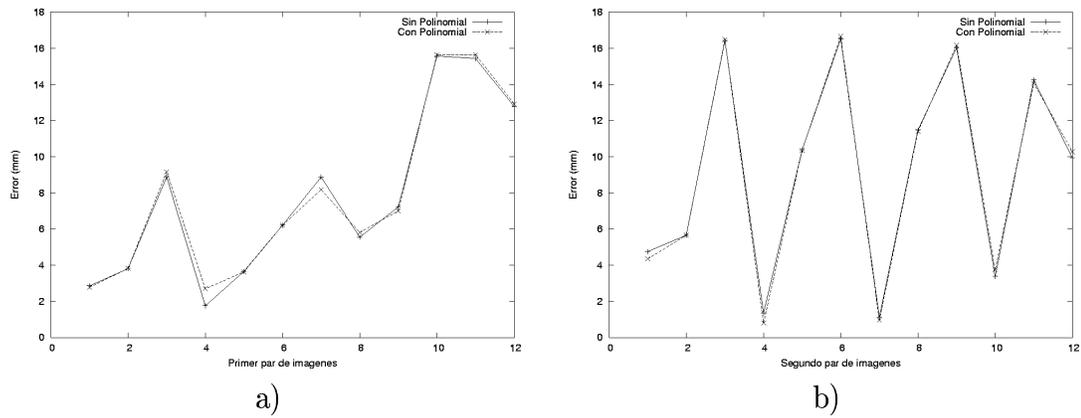


Figura 26: Comparación del modelo tridimensional aplicando y no aplicando la corrección del método polinomial para (a) el par estéreo de la figura 16 y (b) el par estéreo de la figura 25.

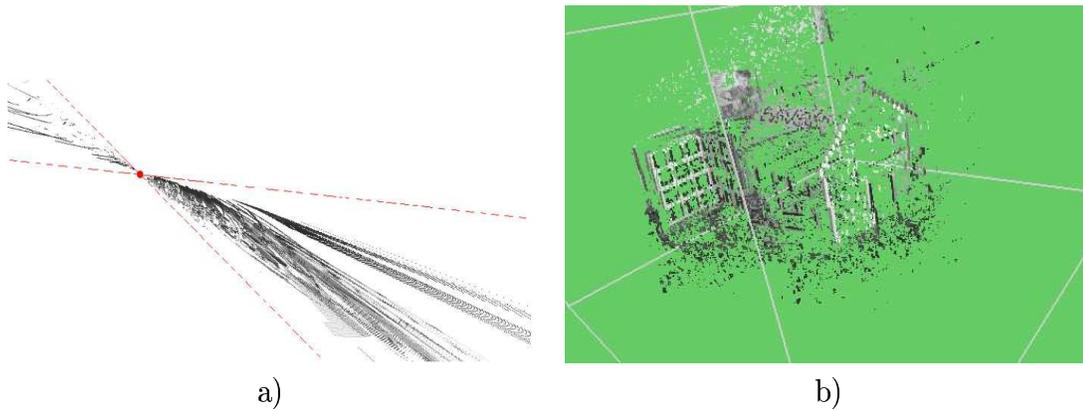


Figura 27: Triangulación con errores. Importancia de los criterios de aceptación en el algoritmo de correspondencia: a) $R = 100$ pix. $U = 0.50$ $V = \pm 10$. b) $R = 100$ pix. $U = 0.85$ $V = \pm 3$ pix.

Parte II
Enfoque Bio-inspirado.

Capítulo V

Introducción a los algoritmos evolutivos.

En la primera parte de este trabajo se ha presentado cómo se aborda geoméricamente el problema de la reconstrucción tridimensional. A partir de ahora, en la segunda parte, se muestra un enfoque diferente, basado en procesos naturales, en donde el problema se plantea como un problema de búsqueda. Es por esto que en este capítulo se aborda uno de los métodos más utilizados para resolver los problemas de búsqueda: los algoritmos evolutivos.

V.1 Métodos clásicos de búsqueda y optimización.

Existen muchos métodos que resuelven problemas de búsqueda y de optimización (Reklaitis (1983)). La mayoría usan un procedimiento determinístico *punto-por-punto* para acercarse a la solución óptima. Arrancan con una solución aleatoria, y mediante reglas de transición que se basan en información local, sugieren la dirección que ha de tomar la búsqueda. Si sobre esta dirección se encuentra una mejor solución que la anterior; ésta es la nueva solución y el procedimiento se repite por un determinado número de veces. La figura 28 ilustra el procedimiento anterior.

La diferencia entre los diversos métodos casi siempre radica en la forma en que se elige la dirección de búsqueda. De acuerdo a la forma en que esta dirección se elige,

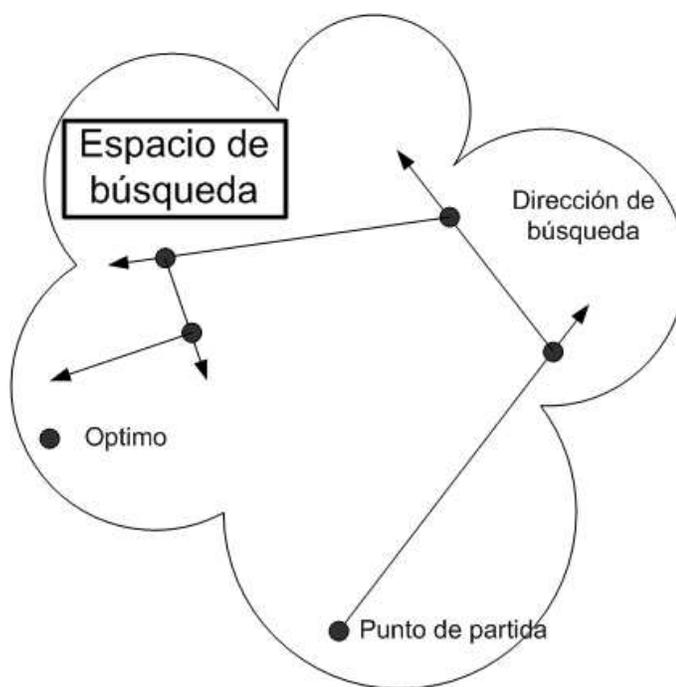


Figura 28: Métodos clásicos de búsqueda y optimización. La mayoría de los métodos clásicos utilizan un procedimiento determinístico punto-por-punto

los métodos de búsqueda y optimización se pueden clasificar en dos grupos: *métodos directos* y *métodos basados en gradiente* (Reklaitis (1983)). En los primeros, la búsqueda se dirige sólo por la función objetivo y las restricciones; mientras que los segundos usan derivadas de primer y segundo orden de la función objetivo y/o las restricciones para guiar la búsqueda. Los métodos directos usualmente son lentos, y requieren de muchas iteraciones para llegar a la mejor solución debido a que no usan derivadas; pero por la misma razón tienden a ser genéricos y se les puede aplicar a muchos problemas sin grandes cambios. Lo contrario sucede con los métodos basados en gradiente, que llegan rápidamente a la mejor solución pero resultan ineficientes en problemas no diferenciables

o discontinuos. Además, ambos tipos de métodos comparten ciertas dificultades:

- La convergencia a una solución óptima depende mucho de la solución inicial.
- La mayoría de los métodos quedan *atascados* en un óptimo local.
- Un método eficiente para cierto tipo de problema, puede resultar ineficiente para otro.
- Generalmente son ineficientes en problemas con un espacio de búsqueda discreto.

Todo lo anterior sugiere que los métodos clásicos de búsqueda y optimización pueden enfrentar dificultades en la solución de problemas prácticos. En lo que resta de este capítulo, se describirán los algoritmos evolutivos, que pueden aliviar las dificultades mencionadas arriba y que están sustituyendo rápidamente a los métodos clásicos.

V.2 Principios Básicos.

Los algoritmos evolutivos (en inglés Evolutionary Algorithm: EA) han demostrado ser un enfoque general y flexible que se adapta a cualquier problema de búsqueda y optimización. Una de las principales diferencias con respecto a los métodos clásicos, es que los EA's utilizan una *población* de soluciones por iteración, en lugar de una sola solución. Si el problema en el que se está usando tiene un sólo óptimo, se espera que el EA converja hacia la solución óptima; y si tiene varios óptimos el EA tiene la capacidad de encontrar el mejor, o bien, permite conservar varias soluciones en su población final con una pequeña modificación al algoritmo.

V.2.1 La evolución natural.

Para mover la población de soluciones sobre el espacio de búsqueda, los EA's utilizan el principio de la evolución natural, donde los más aptos sobreviven y prosperan. Desde hace más de 100 años Charles Darwin, sin conocer nada sobre genética, realizó una extensa investigación sobre la evolución natural e identificó tres principios (Darwin (1859)):

- El ciclo reproductivo.
- La selección natural.
- La diversidad por variación.

El ciclo reproductivo explica la naturaleza iterativa de la evolución mediante el proceso de nacimiento y muerte. Todos los individuos son mortales, pero a través de la reproducción crean hijos que los reemplazan, los descendientes forman la siguiente generación de la especie, y a veces, forman una nueva variedad de la especie mediante aislamiento geográfico y especialización. La selección natural es el proceso mediante el cual los individuos se adaptan al entorno que los rodea. En este proceso, los individuos que llevan a cabo sus funciones vitales de mejor manera tienen mayores probabilidades de sobrevivir, y son ellos los que consiguen reproducirse. Así, en cada generación de individuos, aparecen nuevas alteraciones que convierten a los descendientes en mejores o peores individuos que sus padres, en la capacidad de adaptarse a su entorno, consiguiendo así la diversidad por variación. El entorno se encarga de decidir si las variantes incorporadas a la especie permanecen o no. Si los descendientes llegan a reproducirse

con dichos cambios tienen la posibilidad de transmitirse de nuevo a sus propios hijos y el nuevo rasgo puede llegar a conservarse, hasta que ya no sea de utilidad a la especie.

Darwin no pudo haber la explicación del porqué los individuos tenían esa capacidad de variación, fue Gregor Mendel quien poco después ofreció la respuesta al introducir las bases de la genética. Cada gen posee información codificada sobre un individuo y durante el proceso reproductivo se combinan las características de ambos progenitores para crear a sus descendientes, a esto se le conoce como *cruzamiento*. De esta manera los hijos resultan ser diferentes de ambos padres y no una réplica de alguno de ellos. Además, durante la combinación de genes, periódicamente se cometen errores, en consecuencia se producen *mutaciones* que provocan una nueva diferenciación de los hijos respecto de los padres.

Estas ideas inspiraron a numerosos científicos a aplicar conceptos similares para la resolución de problemas en sus respectivos campos, que eran susceptibles de ser expresados como problemas de búsqueda y optimización. En el área computacional comenzó con John H. Holland en la Universidad de Michigan, Ingo Rechenberg y Hans-Paul Schwefel en la Universidad Técnica de Berlín, y Lawrence J. Fogel en el Instituto Tecnológico de California (CalTech) (Schwefel (1995)). Las primeras aplicaciones de estas técnicas evolutivas aparecen en los años 70's. En la década de los 80's hubo un período de indiferencia, pero posteriormente se dió un resurgimiento en los años 90's. Actualmente se han desarrollado un gran número de técnicas consideradas como algoritmos evolutivos, entre las que destacan los algoritmos genéticos (GA), estrategias de evolución (ES), y una gran cantidad de variantes.

V.3 Estructura General.

A pesar de que existen diferentes técnicas de algoritmos evolutivos, todas ellas comparten los siguientes rasgos fundamentales:

- Una población de individuos, cada uno de los cuales representa la solución potencial al problema en cuestión.
- Un mecanismo de reproducción con en el que se producen cambios en las características de los diferentes individuos.
- Un criterio de evaluación, mediante el cual, diferentes soluciones se comparan de un modo objetivo, o subjetivo; según convenga.
- Un mecanismo de selección por el cual ciertas soluciones son destinadas a ser las generadoras de la siguiente población de individuos y a iniciar un nuevo ciclo evolutivo.

Al igual que en la teoría de la evolución natural, en los algoritmos evolutivos los individuos deben comprobar que son aptos para sobrevivir y/o reproducirse mediante un criterio de evaluación al que se le llama *función de aptitud*. Este criterio califica el lugar que ocupan en el espacio de búsqueda, y aquellos que obtengan los valores más altos tienen mayor probabilidad de ser seleccionados para el proceso de reproducción. La reproducción se lleva a cabo mediante los llamados *operadores evolutivos*. En ella se combina la información de uno o varios individuos (no necesariamente dos) para dar lugar a nuevos individuos. Este proceso se repite iterativamente, es decir, generación tras generación. De esta manera el espacio de búsqueda se va explorando, moviéndose

en la dirección que indiquen los mejores individuos. En la figura 29 se muestra la estructura general de los algoritmos evolutivos.



Figura 29: Diagrama de flujo general de un algoritmo evolutivo.

A continuación se explican cada uno de estos pasos.

V.3.1 Generación de la población.

La elección de una buena representación de la solución es muy importante para cualquier problema de búsqueda. En el caso de los algoritmos evolutivos es importante que los individuos sean diseñados de tal manera que resuelvan el problema satisfactoriamente.

Cada individuo que conforma la población representa una solución dentro del espacio de búsqueda del problema. La forma general de un problema de búsqueda y/o optimización es la siguiente:

Minimizar/Maximizar :

$$f(x_1, x_2, \dots, x_n)$$

sujeto a :

$$g(x_1, x_2, \dots, x_n)$$

\vdots

$$h(x_1, x_2, \dots, x_n)$$

Para :

$$x_1^{min} \leq x_1 \leq x_n^{max}$$

\vdots

$$x_n^{min} \leq x_n \leq x_n^{max}$$

Donde las diferentes x_i representan las variables de decisión, las cuales definen el espacio de búsqueda. El espacio a su vez se encuentra acotado por las restricciones $g(\cdot)$, $h(\cdot)$, etc. Así, cada individuo de la población del EA es en realidad un vector de valores por cada variable de decisión: $ind_i = (x_{i1}, x_{i2}, \dots, x_{in})$.

V.3.2 Evaluación.

La evaluación de un individuo significa calcular su valor con la función objetivo, penalizándolo si incurre en la violación de una o más restricciones, cuantificando así su posición en el espacio de búsqueda. A esta métrica que considera la función objetivo y las restricciones para asignarle un mérito relativo a cada individuo se le conoce como *función de aptitud*. La función de aptitud se diseña específicamente para cada

problema y tiene la capacidad de asignar a un individuo un valor de aptitud que indica qué tan bueno o malo es como solución del problema.

V.3.3 Selección.

El principal objetivo del operador de selección es identificar buenas soluciones y determinar qué individuos dejarán descendencia y en qué cantidad. De esta manera la búsqueda se orienta hacia aquellas soluciones más promisorias. La selección permite que los individuos más aptos se reproduzcan con mayor frecuencia y con menor frecuencia los menos aptos, negando en ocasiones el derecho de reproducción de algunos individuos. Existen varias maneras de llevar a cabo la selección. Las más comunes son: *selección por torneo, selección proporcional, selección por rango.*

Selección por Torneo.

La selección por torneo es de las más populares y sencillas. En ella se eligen dos individuos que compiten entre si comparando su valor de aptitud. El derecho a reproducirse lo gana aquel con un valor mayor o menor, según se esté minimizando u maximizando el problema (ver figura 30, inciso (a)). Este proceso se repite hasta que se obtengan el número de ganadores necesarios para generar la nueva población. La elección de los competidores puede ser sistemática o aleatoriamente. Independientemente del método de elección, es muy probable que el mismo individuo participe en varias competencias, pudiendo ganar el derecho de reproducirse varias veces. Aquel individuo con el mayor valor de aptitud ganará todos los torneos en los que participe permitiendo la conservación de sus características en la nueva población, mientras que el que tenga menor valor perderá todos los torneos, condenándose a la extinción.

Selección Proporcional.

En este tipo de selección, el número de veces en que cada individuo obtiene el derecho de reproducción, es proporcional a su valor de aptitud. Entre mayor valor de aptitud, mayor número de veces lo obtendrá. El mecanismo de este operador funciona como una ruleta (ver figura 30, inciso (b)) donde la rueda tiene N divisiones que representan el tamaño de la población, el ancho de cada división está en proporción a la función de aptitud de cada miembro. Sea f_{total} la suma de los valores de aptitud de toda la población, y sea f_i el valor de aptitud del individuo i , entonces la superficie de la ruleta que le corresponde a cada individuo es f_i/f_{total} . La ruleta girará M veces, según el número de padres que necesite la nueva población, eligiendo cada vez al individuo que señale el indicador de la ruleta.

Selección por Rango.

La selección proporcional puede llegar a tener un problema de *escalamiento*. Por ejemplo, si en una población existe un individuo con un valor de aptitud mucho mas grande que el resto de la población, ocupará la mayor parte de la superficie de la ruleta provocando su selección con demasiada frecuencia, lo cual puede provocar la convergencia hacia un óptimo local. Una manera de evitar este problema es mediante la selección por rango. A cada individuo se le asigna un rango de acuerdo a la magnitud de su valor de aptitud, asignándole al de menor valor el número 1, y al de mayor valor N . Entonces a cada individuo se le asigna un nuevo valor de aptitud proporcional al rango. Después se aplica el operador de selección proporcional con los nuevos valores, de manera normal.

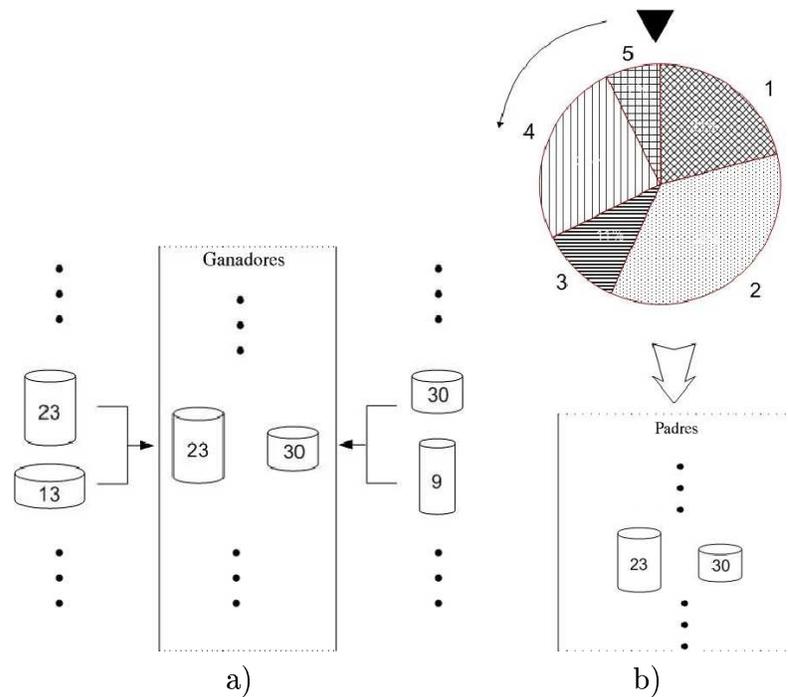


Figura 30: Algunos operadores de selección. (a) Selección por torneo. (b) Selección proporcional.

V.3.4 Reproducción.

La reproducción consiste en aplicar uno o mas operadores \mathcal{O} sobre los individuos seleccionados en la etapa de selección (llamados *padres*: x_i) y obtener la creación de los individuos que conformarán la siguiente generación (llamados *hijos*: y_i). La aplicación del operador se hace por cada componente del vector de valores del individuo

$$y_i = \mathcal{O}(x_i).$$

A estos operadores se les conoce como *operadores evolutivos*. Los operadores evolutivos, junto con una buena representación de la solución del problema son esenciales para el buen rendimiento de un EA. Al igual que la representación del problema, los operadores evolutivos que se utilicen dependen mucho del tipo de problema que se esté resolviendo. Generalmente son dos los operadores que se aplican: mutación y cruzamiento. Existen operadores específicos para problemas en los que se desea obtener más de una solución óptima. Uno de ellos es el de repartición. Por lo regular, la aplicación de estos operadores es estocástica, es decir, su aplicación depende de una función de probabilidad. Aunque en algunos casos es determinística, es decir, se aplica siempre, ya sea a toda la población o a sólo una parte de ella. A continuación se describen los operadores de mutación, cruzamiento y repartición.

Mutación.

La mutación (*mutation* en inglés), altera al azar la información de un solo individuo para crear un nuevo hijo. Evita la pérdida de diversidad en la población y por lo mismo, a la mutación se le considera un operador de exploración del espacio del búsqueda. Al igual que su equivalente en la naturaleza, la mutación incrementa el potencial de variación entre las generaciones de la población. A pesar de que la utilización de los operadores evolutivos depende del problema, un operador de mutación debería considerar los siguientes aspectos para obtener un buen rendimiento:

- **Alcanzabilidad.** Ya que el operador realiza movimientos a través del espacio de búsqueda, se espera que pueda trasladar a un individuo del punto p al punto q en

cierto número de generaciones g , $g < \infty$. La alcanzabilidad se realiza paulatinamente mediante lo que se conoce como *saltos* en el espacio de búsqueda.

- **Escalabilidad.** La longitud de los saltos (fuerza de mutación) debe acoplarse a la cercanía del individuo a una solución óptima. Entre más cercano se encuentre, el salto debe ser más pequeño.

A continuación se presentan algunos operadores de mutación:

Mutación binaria.

Este operador es para representaciones binarias (0, 1). Consiste en cambiar cada bit de 1 a 0, o viceversa, de acuerdo a una probabilidad de mutación p_m . Esto implica generar un número aleatorio por cada bit, lo cual puede resultar computacionalmente costoso. Para evitarlo, Goldberg (Goldberg (1989)) desarrolló la *mutación de reloj*, la cual, después de que un bit ha sido mutado, determina la localización del siguiente bit a mutar con una distribución exponencial. El mecanismo es simple, primero se genera un número aleatorio $r \in [0, 1]$, después se calcula el siguiente bit a mutar saltándose los siguientes $\eta = -p_m \ln(1 - r)$ bits.

Mutación aleatoria.

Este operador fue propuesto por Michalewicz (Michalewicz (1982)). La forma más simple de alterar la información de un individuo es aplicarle una función aleatoria considerando todo el espacio de búsqueda

$$y_i = r_i(x_i^{max} - x_i^{min}), \quad (34)$$

donde r_i es un número aleatorio en $[0, 1]$ que sigue una distribución uniforme, y x_i^{max} , x_i^{min} son los límites superior e inferior, respectivamente, de la variable de decisión x_i (ver figura 31, inciso (a)).

Mutación no uniforme.

Michalewicz también propone el operador de mutación no uniforme. En este caso, la probabilidad de crear un nuevo individuo cerca del padre es mayor que crearlo lejos del mismo. Sin embargo, conforme avanzan las generaciones, la probabilidad de crear al individuo más cerca, va creciendo

$$y_i = x_i + \tau(x_i^{max} - x_i^{min})(1 - a_i^{(1-g/g_{max})^b}) \quad (35)$$

τ sólo puede tomar dos valores: 1 o -1, con una probabilidad de 0.5. El parámetro g_{max} es el número máximo de generaciones, mientras que b es un parámetro definido por el programador. El comportamiento de éste operador es similar al de la mutación aleatoria durante las primeras generaciones, mientras que en las últimas se comporta como la función de Dirac (ver figura 31, inciso (b)).

Mutación Normal.

Un operador simple y popular es la mutación normal, la cual utiliza una distribución gaussiana para provocar la alteración en un individuo. La mutación se aplica por cada componente del vector de valores que compone el individuo

$$y_i = x_i + N(0, \sigma_i) \quad (36)$$

Aquí, el parámetro σ_i determina la fuerza de mutación, y es un valor fijo definido por el programador, es muy importante que su valor se establezca correctamente de acuerdo al problema para que el operador funcione correctamente. Nótese que habrá tantos parámetros como componentes en el vector de valores. En el caso de los problemas cuyas variables de decisión están acotadas se debe cuidar que el nuevo individuo no quede fuera del espacio de búsqueda.

Mutación Polinomial.

La mutación polinomial aplica una distribución polinomial para provocar la alteración en el individuo, en lugar de la distribución gaussiana.

$$y_i = x_i + (x_i^{max} - x_i^{min})\delta_i,$$

donde el parámetro δ_i se calcula a partir de la distribución de probabilidad polinomial $\mathcal{P}(\delta) = 0.5(\eta_m + 1)(1 - |\delta|)^{\eta_m}$

$$\delta_i = \begin{cases} (2r_i)^{1/(\eta_m+1)} - 1, & \text{si } r_i < 0.5, \\ 1 - [2(1 - r_i)]^{1/(\eta_m+1)}, & \text{si } r_i \geq 0.5 \end{cases} \quad (37)$$

La distribución polinomial es similar a la distribución de la mutación no uniforme, a excepción del parámetro η_m definido por el programador, que regula el grado de apertura del área bajo la curva a partir del padre (ver figura 31, inciso (b)). Otra diferencia es que con la mutación polinomial la distribución no cambia dinámicamente con las generaciones, permanece fija.

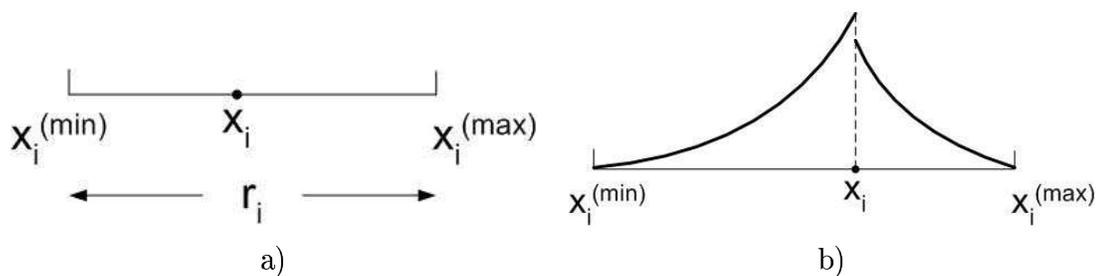


Figura 31: Algunos operadores de mutación. (a) Mutación aleatoria. (b) Mutación no uniforme.

Cruzamiento.

El operador de cruzamiento (*crossover* en inglés, también llamado recombinación) realiza un intercambio de información entre dos o más individuos; produciendo uno o dos hijos, de acuerdo al tipo de cruzamiento que se realice. La idea central es que diferentes segmentos de información, provenientes de padres con un buen grado de adaptación deberían producir hijos que conserven las similitudes entre los padres y aprovecharlas para realizar una búsqueda minuciosa en los espacios entre ellos; por esta razón se dice que el cruzamiento es un operador de explotación, ya que procura agotar la búsqueda en regiones con mayor valor de aptitud, en generaciones sucesivas de selección y cruzamiento. Un buen operador de cruzamiento debe considerar los siguientes aspectos para obtener un buen rendimiento:

- **Imparcialidad.** Cada individuo que tome parte en el cruzamiento debe contribuir en la misma proporción.
- **Especialización.** Un operador de cruzamiento debe diseñarse específicamente de acuerdo al problema que se esté resolviendo.

A continuación se presentan algunos operadores de cruzamiento:

Cruzamiento binario.

Se utiliza cuando un individuo se representa por una cadena binaria $(1, 0)$. Es bastante simple, necesita de dos padres y consiste en elegir un punto en la cadena, y a partir de él intercambiar esa porción entre las dos cadenas y así crear dos nuevos individuos. Véase la figura 32, inciso (a).

Cruzamiento Lineal.

También se le conoce como baricéntrico, una de sus primeras implementaciones fue reportada por Wright (Wright (1991)). Consiste en generar una combinación lineal de dos padres X^1 y X^2 para generar un hijo:

$$y_i = (1 - \alpha)x_i^1 + \alpha x_i^2 \quad (38)$$

donde α es una variable aleatoria en $[0, 1]$ que sigue una distribución uniforme.

Cruzamiento binario simulado.

Simulated Binary Crossover (SBX por sus siglas en inglés) lo propone Deb (Deb (2001)). Trabaja con dos padres y genera dos hijos. Simula el principio del cruzamiento binario de la siguiente manera:

β_i es un factor de dispersión que se define como la razón de la diferencia absoluta de los hijos entre los padres

$$\beta_i = \left| \frac{y_i^2 - y_i^1}{x_i^2 - x_i^1} \right|$$

Primero se crea un número aleatorio u_i entre 0 y 1. Después, a partir de una función de distribución de probabilidad específica, la ordenada de β_{qi} se encuentra de tal manera que el área bajo la curva de probabilidad de 0 a β_{qi} es igual al número aleatorio u_i . La función de distribución de probabilidad de la que se habla debe tener el mismo *poder de búsqueda* que el cruzamiento binario, y es la siguiente

$$\mathcal{P}(\beta_i) = \begin{cases} 0.5(\eta_c + 1)\beta_i^{\eta_c}, & \text{si } \beta_i \leq 1; \\ 0.5(\eta_c + 1)\frac{1}{\beta_i^{\eta_c+2}}, & \text{de otra manera} \end{cases}$$

Donde η_c es cualquier número no negativo. Un valor grande resulta en una probabilidad alta de que los hijos queden cerca de los padres, mientras que un valor pequeño permite obtener individuos lejanos a los padres (véase la figura 32, inciso (b)). Usando esta expresión el valor de β_{qi} puede calcularse, dependiente de el área bajo la curva igual a u_i

$$\beta_{qi} = \begin{cases} 2u_i^{1/(\eta_c+1)} & \text{si } u_i < 0.5 \\ \frac{1}{2}(1 - u_i)^{1/(\eta_c+1)} & \text{si } u_i \geq 0.5 \end{cases}$$

Después de obtener β_{qi} de la distribución de probabilidad de arriba, se procede a calcular los hijos

$$y_i^1 = 0.5[(1 + \beta_{qi})x_i^1 + (1 - \beta_{qi})x_i^2] \quad (39)$$

$$y_i^2 = 0.5[(1 - \beta_{qi})x_i^1 + (1 + \beta_{qi})x_i^2] \quad (40)$$

En la la figura 32, inciso b) sed parecia que entre mayor es η_c el área bajo la curva es más estrecha, lo que provoca que el hijo quede cerca del padre. El cruzamiento SBX

tiene dos propiedades esenciales:

1. La diferencia entre los hijos está en proporción a los padres.
2. Los hijos cercanos a los padres son monotónicamente más probables de escoger que aquellos que son lejanos.

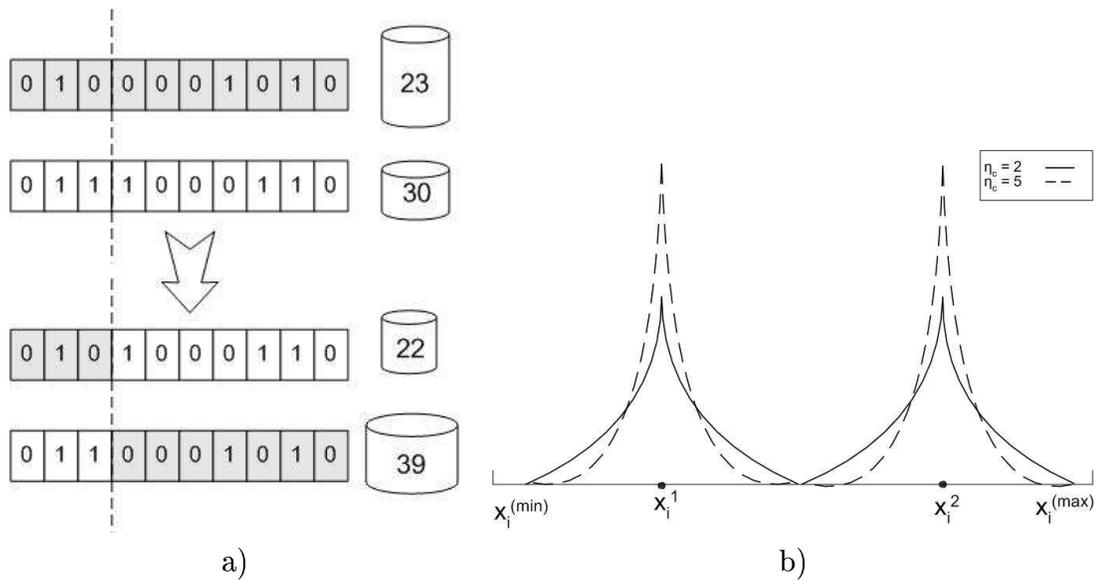


Figura 32: Algunos operadores de cruzamiento. (a) Cruzamiento binario. (b) Cruzamiento SBX.

Operadores de nicho: La repartición.

Por sus características, un algoritmo evolutivo explora el espacio de búsqueda hasta converger hacia una región del espacio. El objetivo de este trabajo no es encontrar una solución global total, sino varias soluciones locales, a esto se le llama objetivo

multimodal. Una solución multimodal implica encontrar varias soluciones óptimas. Para evitar la convergencia es necesario introducir un mecanismo para mantener la diversidad de la población durante todo el proceso.

En la naturaleza, los recursos como el agua y la comida presentes en un ecosistema son variados pero limitados y deben compartirse por todas las especies vivas en este ecosistema. Si hay abundancia de recursos las especies prosperan y se multiplican; si hay escasez de estos recursos el número de individuos de las especies disminuye y sólo los más aptos sobreviven. El ecosistema busca, por si mismo, el equilibrio y no una lucha competitiva, inútil y desgastadora; y lo hace mediante *nichos ecológicos*. El nicho es el rol que cumple cada organismo dentro del ecosistema y depende de sus características, su grado de adaptabilidad y su conducta. El nicho ecológico se compone del conjunto de los factores físicos, químicos y fisiológicos que necesita un organismo para vivir.

Inspirado en el principio de nichos ecológicos y en el equilibrio que intenta mantener la naturaleza, John Holland (Holland (1975)) propuso el principio de nichos en los algoritmos evolutivos. Aquí los nichos representan una región donde existe una solución óptima, alrededor de la cual se reúnen un conjunto de individuos. Véase la figura 33, inciso b). Estos individuos, al compartir el mismo nicho, comparten un recurso común: el valor de aptitud. Entre más individuos conformen un nicho, la cantidad del recurso asignado a cada uno es más pequeña. Este mecanismo se aplica generación tras generación, hasta crear un equilibrio donde el tamaño de la población en un nicho y el número de nichos son proporcionales a la disponibilidad del recurso. El objetivo de los operadores de nicho es que la población se distribuya sobre los mejores lugares dentro del espacio de búsqueda. El número de individuos en un nicho debe ser proporcional

a los recursos (aptitud) de ésta. El principio de nichos ecológicos se aplica de maneras diferentes en los algoritmos evolutivos, siendo uno de ellos el operador de repartición (*sharing* en inglés) propuesto por Goldberg (Goldberg (1989)).

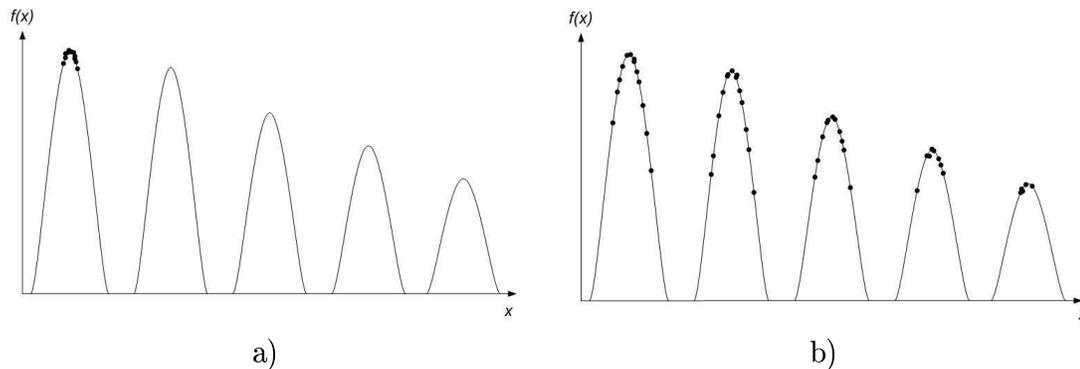


Figura 33: Una función objetivo multimodal. A cada pico se le considera como nicho. a) Resultado de un EA sin aplicar operador de repartición. b) Resultado de un EA aplicando operador de repartición.

En el operador de repartición (*sharing*, en inglés) el valor de aptitud de un individuo baja de acuerdo al número de individuos dentro de su *zona de influencia* (nicho), con los que debe compartir recursos. Así el valor de aptitud compartido de un individuo apt'_i es igual a su valor de aptitud original apt_i dividido entre su contador de nicho m_i :

$$apt'_i = \frac{apt_i}{m_i} \quad (41)$$

El contador de nicho representa el número de individuos que comparten el nicho con el individuo i y se calcula mediante la sumatoria de una función de repartición Sh del individuo i con respecto a los demás individuos:

$$m_i = \sum_{j=1}^N Sh(d_{i,j})$$

La función de repartición es una métrica de la distancia d entre dos individuos i y j . La función de repartición más utilizada tiene la siguiente forma

$$Sh(d_{i,j}) = \begin{cases} 1 - \frac{d_{i,j}}{\sigma_{rep}} & \text{si } d_{i,j} \leq \sigma_{rep} \\ 0 & \text{de otra manera} \end{cases}$$

Donde σ_{rep} representa el radio de la zona de influencia de cada individuo. Por lo general se utiliza la distancia euclidiana para medir la distancia entre individuos

$$d_{i,j} = \sqrt{\sum_{k=1}^q (x_{ik} - x_{jk})^2},$$

donde q representa el número de componentes del vector de variables que conforma cada individuo. Así, cuando muchos individuos se encuentran muy cercanos unos con otros, todos contribuyen a la disminución de sus respectivos valores de aptitud. Como resultado, queda limitada la proliferación de individuos en esa zona del espacio de búsqueda, dando oportunidad a la población emergente de poblar otros nichos, después de un determinado número de generaciones se habrán poblado todos los nichos y cada uno de éstos poseerá una cantidad estable de individuos.

V.3.5 Criterio de paro.

El algoritmo evolutivo avanzará generación tras generación evolucionando la población hacia la o las mejores soluciones en el espacio de búsqueda, hasta que se le diga que se detenga. En ocasiones, cuando no se sabe con certeza la región donde se encuentra

la solución óptima (o en su caso las soluciones), o bien cuando el espacio de búsqueda resulta difícil de modelar, es complicado definir un criterio de paro para el algoritmo evolutivo. Además, la naturaleza estocástica de la población inicial y de la generación de la nueva población agrega un nivel de complejidad más. Generalmente, cuando cierto porcentaje de la población converge en un punto del espacio de búsqueda el algoritmo se detiene, aunque este criterio no es aplicable para problemas de objetivo multimodal; también se puede aplicar el criterio de parar cuando el mejor individuo de la población no cambia después de cierto número de generaciones. Otro criterio es definir *a priori* el número de generaciones que el algoritmo llevará a cabo. Este último es el más utilizado, ya que se puede establecer un número grande de generaciones y llevar a cabo pruebas estadísticas analizando la convergencia de los valores de aptitud a través de las generaciones; y así estimar el número de generaciones apropiado y utilizarlo en el futuro.

V.4 Técnicas Evolutivas.

Dos de las versiones más utilizadas de algoritmos evolutivos son los algoritmos genéticos y las estrategias de evolución. Aunque hoy en día la diferencia entre ellas se ha acortado significativamente y han surgido muchas variantes. Ambas técnicas utilizan indistintamente representación binaria y real, y ambas utilizan los operadores reproductivos sin distinción, junto con un buen número de operadores híbridos y operadores nuevos, que han surgido de acuerdo a las necesidades de diversos problemas.

V.4.1 Algoritmos Genéticos.

Propuestos por John Holland y complementados por Goldberg. Esta técnica se basa en la genética de los organismos vivos. La cual, dice que a la estructura fundamental que contiene los rasgos de un individuo se le conoce como *cromosoma*. Uno o más cromosomas especifican un organismo como un todo. Al conjunto de cromosomas se le conoce como *genotipo*, y a su traducción en rasgos físicos; es decir, al organismo resultante se le llama *fenotipo*. Un cromosoma consta de cierto número de subestructuras individuales llamadas *genes*. Cada gen codifica una característica particular del organismo; y la localización o *locus* del gen dentro del cromosoma determina cuál de varios valores diferentes de esta característica el gen representa. A los diferentes valores que un gen puede representar se les conoce como *alelos*.

En un algoritmo genético (en inglés *Genetic Algorithm: GA*), cada solución representa un *cromosoma*, y un conjunto de cromosomas forman la población de soluciones. Las variables de decisión que conforman cada cromosoma son codificadas en cadenas binarias, y se les llama *genes*. La posición de cada variable de decisión es el *locus* de la variable, similarmente, los *alelos* representan los diferentes valores que una variable de decisión puede tomar. El *genotipo* es la concatenación de unos y ceros que forman el cromosoma, mientras que el *fenotipo* es la decodificación del cromosoma en números reales para su evaluación en la función de aptitud.

Los GA's basan su esfuerzo para resolver el problema en alfabetos de baja cardinalidad para codificar la información (generalmente, un alfabeto binario 1, 0) y en la capacidad del operador de cruzamiento para construir bloques, esto es, cadenas cortas que pueden ser optimizadas independientemente una de otra para después juntarlas

y conformar la solución óptima. La mutación se aplica escasamente mediante eventos aleatorios que consisten en intercambiar el caracter por otro del pequeño alfabeto, además sólo se aplica a un pequeño número del total de la población. En la figura 34 se muestra la estructura de un algoritmo genético.

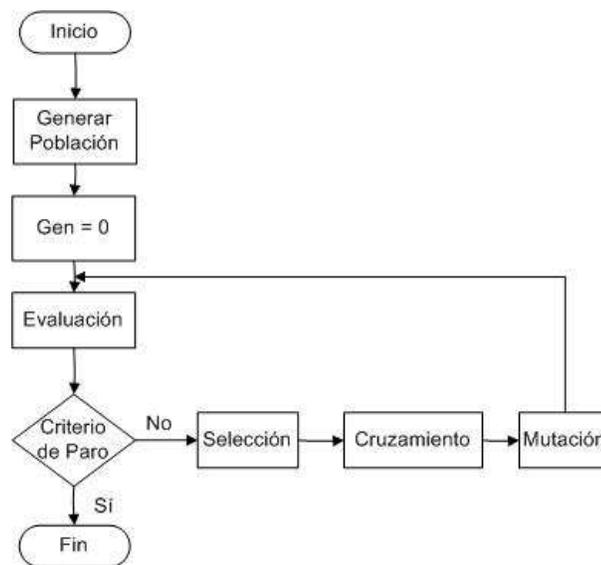


Figura 34: Estructura general de un algoritmo genético.

V.4.2 Estrategias de Evolución.

Esta técnica fue desarrollada de manera independiente por Ingo Rechenberg y Hans-Paul Schwefel (Schwefel (1995)) y se basa en el mecanismo de selección natural de Darwin. En este caso la población se conforma de individuos que compiten con los demás por su derecho a reproducirse y/o sobrevivir a través de las generaciones. Cada individuo se compone de un vector de valores que pertenecen al conjunto de los números

reales que representan el conjunto de variables de decisión. Las estrategias de evolución (en inglés *Evolution Strategies*: ES) son diferentes a los GA's en dos aspectos principalmente, en primer lugar, basan su esfuerzo evolutivo en la *fuerza de mutación* y aplican poco los operadores de cruzamiento, de hecho en un principio las ES's no contaban con un operador de cruzamiento. En segundo lugar, las ES's no codifican la representación del problema, trabajan directamente en el espacio de búsqueda del problema.

Existen diferentes tipos de ES's, Bi-miembros, Multimiembros y los Autoadaptables:

- **Estrategia de evolución bi-miembro: (1+1)-ES.** Esta es la forma más simple. En cada generación se toma un sólo padre para crear un hijo aplicando el operador de mutación. Y si el valor de aptitud del hijo resulta mejor que el del padre, se conserva el primero y se deshecha el segundo.
- **Estrategia de evolución Multi-miembro: $(\mu+\lambda)$ -ES y (μ,λ) -ES.** En este tipo de ES, un tamaño determinado de población μ se utiliza para generar λ hijos. Existen dos formas de aplicar el concepto de multimiembro, en la primera, llamada estrategia *mas*, la población para la nueva generación se elige de entre los mejores padres e hijos; mientras que la estrategia *coma*, sólo toma en cuenta a los hijos, los padres son desechados después de cada generación. En $(\mu+\lambda)$ -ES los padres pueden sobrevivir a través de las generaciones, lo cual puede controlarse mediante el parámetro κ . Así, $\kappa = n$ permite que un individuo sobreviva por n generaciones. Nótese que una $(\mu+\lambda)$ -ES con $\kappa = 1$ es en realidad una (μ,λ) -ES.
- **Estrategia de evolución autoadaptable.** La autoadaptabilidad permite a las ES's ser flexibles y más cercanas a la evolución natural y fue desarrollada por

Bäck (Deb (2001)) y se le puede aplicar para los dos tipos de ES anteriores. La autoadaptación es recomendable en el caso de que se necesite conocer la solución óptima con una precisión arbitraria, o bien, cuando la función objetivo y/o la solución óptima cambien con el tiempo. Existen tres maneras diferentes de aplicar la autoadaptabilidad en una estrategia de evolución:

1. *Aplicando un meta-ES.* Consiste en utilizar dos niveles de evolución; en el primero se optimizan los parámetros de la segunda estrategia de evolución y en la segunda se trabaja con el problema de manera normal, pero con los parámetros que el primer nivel arrojó.
2. *Aplicando una matriz de covarianza.* Consiste en registrar el comportamiento de la población para cierto número de generaciones y calcular su varianza y covarianza, modificando los parámetros de los operadores de acuerdo al resultado que se desea, en las generaciones subsecuentes.
3. *Uso explícito de parámetros de control autoadaptables.* La idea es similar al punto anterior, sólo que aquí las varianzas y covarianzas forman parte del vector solución, evolucionando junto con las variables de decisión. En consecuencia se debe modificar la función de aptitud para que considere las nuevas variables.

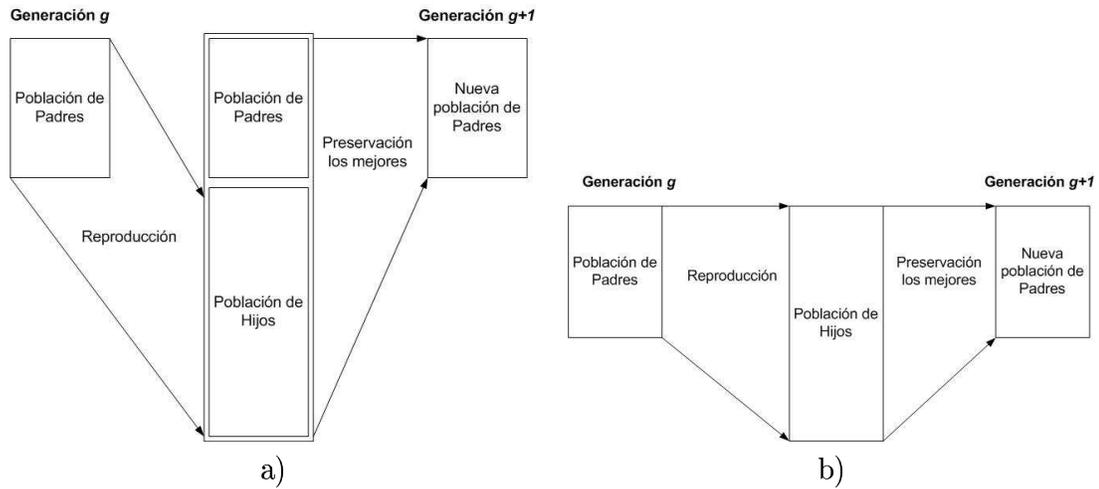


Figura 35: Estructura general de una estrategia de evolución. (a) $(\mu+\lambda)$ -ES. (b) (μ,λ) -ES

V.5 Conclusiones.

Los algoritmos evolutivos (EA's) han tenido gran aceptación como algoritmos de búsqueda y optimización robustos en los últimos diez años. La idea de encontrar soluciones imitando a la naturaleza y aplicar los *operadores evolutivos* como selección, mutación y cruzamiento han provocado un interés que le va ganando terreno a los métodos clásicos. Esto se debe a las ventajas que ofrece el aplicar un algoritmo evolutivo; entre las cuales, las más importantes son las siguientes:

- Son simples y fáciles de implementar. Además son de propósito general, esto es, que pueden aplicarse a diferentes problemas sin cambios significativos a la estructura del algoritmo.
- No necesita gran conocimiento *a priori* sobre el problema, el conocimiento *emerge*

a través de las generaciones. No es necesario para el programador conocer a detalle las restricciones del problema, sean lineales o no lineales.

- El principio de evolución natural permite una búsqueda global efectiva, basada en la selección de los mejores individuos de una población que cambia generación tras generación.

Existen diferentes técnicas o *corrientes* de algoritmos evolutivos, como la *Programación Evolutiva* propuesta por Fogel y la *Programación Genética* que propone John Koza (Koza (1992)). Sin embargo, entre los más utilizados destacan los algoritmos genéticos (GA's) y las estrategias de evolución (ES's), los cuales han sido descritos en este capítulo. La existencia de diferentes técnicas que ponen distinto énfasis en el rol de los distintos operadores evolutivos se debe más que nada a la falta de una teoría unificadora, la cual debería especificar el objetivo de cada operador y permitir una comparación objetiva entre las posibles combinaciones de operadores para medir su *poder de combinación*. La deficiencia es mayor en los GA's, quienes tienen mayor base empírica. En cambio, en las ES's se han hecho más esfuerzos para entender los algoritmos teóricamente (Hans-Georg (2001)).

La diferencia entre estas dos técnicas ha ido disminuyendo con el tiempo, pues debido a su flexibilidad, hay aplicaciones que utilizan GA's con valores reales en lugar de binarios y ES's con valores binarios en lugar de reales. Las configuraciones de población: (μ, λ) y $(\mu + \lambda)$ aplican en los algoritmos genéticos bajo el nombre de *selección truncada* y *selección elitista* respectivamente. La autoadaptabilidad de los ES's actualmente se encuentra en fase experimental en los GA's.

En el siguiente capítulo se propone un enfoque novedoso para afrontar el problema de la reconstrucción tridimensional, el cual es tratado como un problema de búsqueda. Se ha decidido para tal propósito utilizar algoritmos evolutivos por las razones expuestas en este capítulo, y además, porque el problema de reconstrucción representa un objetivo multimodal, ya que exige varias puntos para conformar un modelo tridimensional (varias soluciones óptimas). Dentro de los algoritmos evolutivos, se ha decidido utilizar estrategias de evolución, pues el espacio de búsqueda maneja número reales (espacio euclidiano tridimensional), que aunque podría ser abordado de igual manera por un GA, se prefiere respaldar este trabajo con los resultados teóricos que han alcanzado las ES's y además se apega a las filosóficas iniciales de ambas técnicas.

Capítulo VI

Reconstrucción 3D: Algoritmo de búsqueda basada en el comportamiento de las abejas.

Dentro de las ciencias computacionales, existen diversas metodologías que han sido inspiradas por observaciones de los diferentes procesos naturales, como las redes neuronales, cómputo evolutivo, etc. Muchas de estas metodologías simulan un fenómeno natural para aplicaciones cuyo objetivo no necesariamente es el mismo que el proceso natural que simulan. A estos métodos se les conoce como bio-inspirados. Su objetivo es diseñar sistemas artificiales adaptativos, flexibles, y robustos, capaces de resolver cierto problema por sí mismos. Dentro de la visión por computadora se han implementado con éxito técnicas bio-inspiradas para abordar problemas desde un punto de vista diferente (Boumaza y Louchet (2001); Kass *et al.* (1987); McInerney y Terzopoulos (1995); Olague *et al.* (2004)).

Un ejemplo es el algoritmo de las moscas. El cual es una técnica basada en algoritmos evolutivos para la exploración del espacio tridimensional en un sistema de visión para un robot móvil (Boumaza y Louchet (2001)). En este algoritmo la solución se representa por la mayor parte de la población, no por un solo individuo. Cada uno de ellos representa un punto en el espacio 3D y evoluciona evaluando una función de aptitud que, basada en las imágenes, intenta concentrar la población en los bordes de

objetos que conforman la escena. Sin embargo, para cierto tipo de imágenes, donde predominan las zonas de baja textura el criterio de buscar los bordes puede no ser suficiente; de hecho la consideración de la textura debería ser implementada, ya que por lo general, los objetos relevantes en una escena poseen gran cantidad de textura. Además, cada individuo representa una parte de la solución total; y como tal es necesario que exista una interacción explícita entre los individuos para acarrear mejores resultados. Sería deseable que un algoritmo con características del algoritmo de las moscas fuera complementado con la información que se acaba de mencionar. Por esta razón se ha realizado una investigación para aplicar las características de las que se está hablando, inspirándose en el comportamiento de un insecto inteligente y organizado: la abeja.

En este capítulo se desarrolla una nueva técnica bio-inspirada que no realiza el trabajo masivo de triangular cada pixel de las imágenes de la escena. En su lugar, se plantea el problema de la reconstrucción de forma inversa; en vez de construir el modelo 3D a partir de la imagen, se propone una búsqueda en el espacio tridimensional para encontrar el modelo tridimensional que sea coherente a la escena que está en las imágenes. De esta manera, el problema a resolver puede verse como un problema de búsqueda, donde el espacio de búsqueda está definido en el espacio tridimensional, y se comprueba su validez en las imágenes. En la primera sección de este capítulo se plantea la reconstrucción como problema de búsqueda, posteriormente se da una explicación de la estructura social de una colmena de abejas con el fin de dar a entender el origen de la técnica bio-inspirada. Una vez definido esto, se procede a plantear el algoritmo, su estructura y las partes que lo componen. Finalmente se muestran las pruebas que se realizaron sobre el algoritmo para demostrar su rendimiento y convergencia a una

solución óptima, con diferentes operadores evolutivos.

VI.1 Planteamiento de la reconstrucción 3D como problema de búsqueda.

Generalmente en la visión por computadora, se lleva a cabo la reconstrucción tridimensional como se ha planteado en la primera parte de esta tesis: usando un enfoque que se basa en las relaciones geométricas entre la escena y al menos dos imágenes de esta escena (ver capítulos II, III, IV). La matriz de proyección permite la transformación de las coordenadas de la escena a la imagen; a lo cual se le conoce como *enfoque directo*. Por otro lado, la transformación de las coordenadas de la imagen a la escena se le conoce como *enfoque inverso*, el cual se lleva a cabo mediante triangulación de los píxeles correspondientes en dos imágenes de la escena. Al proceso de correspondencia se le considera entre los más complejos en el cual, es frecuente cometer errores debido a la ambigüedad que se presenta al elegir el píxel adecuado y a los errores inherentes en el cálculo de la geometría epipolar.

Para evitar el problema de correspondencia se propone abordar el problema de reconstrucción aplicando el enfoque directo, en el cual, se buscan los puntos con coordenadas (x, y, z) que sean coherentes con la escena; basándose en la información proporcionada por las proyecciones en las imágenes izquierda (u_i, v_i) y derecha (u_d, v_d) . De esta manera, el problema se convierte en un problema de búsqueda en el espacio tridimensional, donde el objetivo es encontrar un modelo tridimensional que se acerque más a la escena real. Formalmente, se define como sigue

$$\begin{aligned}
\text{Maximizar :} & \quad f(\mathbf{P}) \\
\text{sujeto a :} & \quad x^{\min} \leq x_i \leq x^{\max} \\
& \quad y^{\min} \leq y_i \leq y^{\max} \\
& \quad z^{\min} \leq z_i \leq z^{\max}
\end{aligned} \tag{42}$$

Donde $f : \mathbb{R}^3 \rightarrow \mathbb{R}$ es una función multimodal que recibe como entrada un conjunto de n puntos en el espacio tridimensional $\mathbf{P} = \{x_i, y_i, z_i \mid i = 1..n\}$. Estos puntos no deben salir del espacio de búsqueda, el cual es en realidad el espacio euclidiano acotado por los valores de restricción $\mathfrak{R} = \{x^{\min}, x^{\max}, y^{\min}, y^{\max}, z^{\min}, z^{\max}\}$. f se encarga de proyectar los puntos del conjunto \mathbf{P} sobre las imágenes con ayuda de la matriz de proyección para evaluar qué tan coherente es cada uno de estos puntos con respecto a la escena. La salida de esta función es un número real en $[0, 1]$ que representa la evaluación de coherencia. El objetivo entonces es encontrar varios puntos (soluciones óptimas) que conformen un modelo tridimensional lo más cercano posible a la escena real.

Como se dijo en la sección V.5, se ha decidido utilizar las estrategias de evolución como herramienta de búsqueda. Dentro de la estrategia de evolución se modela el comportamiento de las abejas para alcanzar el objetivo de una manera inteligente y coordinada. Antes de proponer el algoritmo basado en el proceso de búsqueda de las abejas se presenta una breve reseña del comportamiento social de las abejas para comprender de una mejor manera su importancia dentro de este trabajo.

VI.2 Comportamiento de las abejas.

El comportamiento de las abejas desde que nacen hasta el final de sus días ha sido estudiado por muchos investigadores, entre los que destacan Karl von Frisch (von Frisch (1967)) y recientemente Elaine Crist (Crist (2004)). Una colonia de abejas consiste de unos cuantos miles de abejas que han desarrollado mecanismos de comunicación que les permite una organización funcional y compleja para cooperar en tareas de supervivencia; como la construcción y mantenimiento de la colmena y recolección de agua y comida; lo cual sobrepasa por mucho las capacidades de una sola abeja, lo que las obliga a trabajar en conjunto. Este comportamiento colectivo de la colonia carece de un control central. Se organizan en un estricto sistema de castas donde cada individuo tiene un rol; el cual va cambiando de acuerdo a la edad de cada abeja. La información e instrucciones que recibe una abeja para realizar alguna actividad, viene de alguna otra que entra en contacto con ella. La estructura de la colonia es tal, que la decisión de los individuos para realizar alguna tarea se basa en información local sin una referencia a la situación global de la colonia. Al trabajo local de muchos individuos para lograr un objetivo global se le llama *auto-organizado*. La auto-organización de tareas en una colonia requiere de dos mecanismos de comunicación. El primero, *retroalimentación positiva* donde aquellas abejas que realizan alguna actividad útil para la colonia piden la cooperación de aquellas que están inactivas o que hacen alguna tarea innecesaria. El segundo, *retroalimentación negativa* donde una abeja debe enterarse de que continuar con la tarea que lleva a cabo es irrelevante y la deje de hacer. La comunicación entre los individuos permite a la colonia responder apropiadamente al entorno actual, y adaptarse rápidamente a cualquier cambio en el entorno.

VI.2.1 Organización social: castas.

En la sociedad de abejas existe una estricta distribución de funciones estando la actividad reproductora relegada a la casta reproductora, constituida por la reina y los machos, mientras que el resto de funciones de intendencia, como fabricación y limpieza del nido, aprovisionamiento, mantenimiento de la temperatura, alimentación de las larvas, etc. corre a cargo de la casta obrera, formada por las hembras estériles.

Casta Reproductora.

La reina. La reina es la única hembra fértil de la colmena, y por lo tanto la madre del resto del enjambre. Por lo general es única aunque, excepcionalmente, pueden existir colmenas con dos reinas. Puede vivir hasta 7 años por lo que puede sobrevivir a muchas generaciones de hijas, que colaboran con ella en la cría y cuidado de sus hermanas. La reina desempeña una doble función: coordinadora y madre.

La reina es una pieza fundamental en la organización de la colonia, pues regula y coordina la actividad del resto de los individuos químicamente, mediante la llamada *feromona real*. La feromona real ejerce una poderosa atracción sobre el resto de los miembros de la colonia, siendo muy apreciada por todas las obreras, ávidamente la recogen lamiendo el cuerpo de la reina, se distribuye por todos los elementos de la colmena por trofalaxia ¹. También actúa sobre las obreras inhibiendo el desarrollo de los ovarios, por lo que permanecen estériles. La presencia de la feromona real regula la construcción de celdas de cría, dependiendo de su nivel, obliga a las obreras a construir celdas grandes o pequeñas. Cuando al inicio de la primavera existen pocas abejas

¹Transferencia de alimento boca a boca

en la colonia, la cantidad de feromona real es alta. Las obreras se ocupan de elaborar nuevos panales cuyas celdillas son de tamaño pequeño, en donde nacerán futuras obreras que fortalecerán la colonia. En verano, cuando la colmena tiene muchos miembros, la cantidad de feromona real es menor, provocando la construcción de celdillas de cría grandes para machos, y a construir celdas reales, preparando a la colmena para el enjambrazón ². Lo mismo pasa cuando la reina envejece y fabrica poca jalea real, en este caso la colmena se prepara para la sustitución de la reina.

Los machos. Constituyen el otro elemento fundamental en la casta reproductora, siendo los individuos de mayor tamaño de la colmena. También se les llama zánganos, por su falta de participación en la vida laboral de la colmena, al carecer de cestillas para el polen y tener lengua pequeña. De recién nacidos y jóvenes permanecen en la zona de cría donde los alimentan las obreras, posteriormente se alimentan solos de la miel almacenada en la colmena y salen de la zona de cría. Se dice que no pertenecen a ninguna colmena, pues son admitidos en todas. Con sus enormes y coalescentes ojos, localizan a la reina en vuelo y desempeñan el importantísimo papel de fecundarla, siendo muy atraídos por la feromona real que la reina segrega. Los zánganos sólo están presentes en la colmena mientras existan reinas que fecundar, al final del otoño se produce la expulsión de los zánganos de la colonia, en cuyo exterior mueren de frío y hambre. En ocasiones, las obreras originan verdaderas matanza de zánganos, cuando éstos no quieren marcharse de la colmena.

²Proceso por el cual las abejas crean la división de la colmena tendiente a perpetuar la especie, una reproducción de la misma.

Casta obrera.

La casta más abundante de la colmena está formada por los miembros más pequeños de la colonia, las obreras. Éstas son hembras que se desarrollan en celdillas de pequeño diámetro. Por su escasa alimentación en la vida larvaria, estas hembras resultan estériles. Su nombre *obreras* es debido a que se encargan de todas las funciones de intendencia de la colonia, siendo éstas muy variadas dependiendo de su edad y estado fisiológico. Las obreras tienen una morfología especial, con órganos adaptados al desempeño de funciones, como el desarrollado aparato bucal con el que succionan el néctar, que almacenan en el buche, para transportarlo a la colmena, y en sus patas posteriores muestran unas estructuras rodeadas de rígidas sedas, llamadas *cestillas*, con las que sujetan las bolitas de polen en su viaje de retorno a la colmena. Su ovipositor ³ ha perdido las funciones propias, transformándose en un aguijón, órgano vulnerante conectado con una glándula de veneno, con el que defienden la colmena y sus provisiones de los extraños, incluso a costa de su propia vida.

Existe una secuencia de actividades que realizan las obreras de acuerdo a su edad. La vida de una obrera desde que nace hasta que muere se puede resumir en las etapas que a continuación se presentan, si bien no siempre todas las abejas siguen este patrón de conducta, se puede tomar como base para describir sus diferentes funciones.

1ra. Etapa: Limpiadoras. Acabadas de nacer y salir de su celdilla, comienzan a limpiarla de las fundas que han dejado en el interior, por las mudas de su desarrollo larvario. Y durante dos o tres días más siguen limpiando las celdillas de sus hermanas

³Conducto que desarrollan las hembras para la puesta de huevos

recién nacidas. De esta manera esas celdillas quedan preparadas para una siguiente postura de huevos por parte de la reina. También extraen de la colmena los cadáveres de las larvas vecinas que hayan muerto.

2da. Etapa: Nodrizas. Al estar limpiando celdillas, las abejas limpiadoras se encuentran con celdas ocupadas por larvas que demandan su alimento mediante feromonas. Lo cual, provoca que sus glándulas productoras de jalea real comiencen a desarrollarse, lo cual se alcanza entre los seis y doce días de edad; entonces comienzan a alimentar con jalea real a larvas de menos de 3 días. Parte de estas nodrizas forman la corte real, que atraídas por la feromona real, alimentarán a la reina. Para la fabricación de la jalea real las abejas nodrizas necesitan alimentarse intensamente de polen, de donde obtienen las proteínas necesarias, ya que la jalea real no es otra cosa que una solución con proteínas de alto valor nutritivo.

3ra. Etapa: Variable. A partir de la tercera semana de vida, las abejas realizan diferentes actividades que están menos relacionadas con la edad y que suceden según su estado fisiológico y las necesidades que tenga la colmena, estas actividades son flexibles y repetitivas. Se alternan las tareas de limpieza de cadáveres, suciedad, maduración de néctar (fabricación de miel), construcción de celdas, recolección y almacenamiento de agua y alimento, y a continuación se describe cada una.

- **Cereras.** Las obreras poseen cuatro glándulas situadas en el abdomen que producen una fina película de cera. Esta secreción la recogen con sus patas y la amasan con las mandíbulas. Con la cera estirada construyen las celdillas en forma de prisma hexagonal, cuyo eje está un poco inclinado, lo que evita que se caiga su

contenido. Las cereras también se encargan de sellar las celdas, tanto las que tienen miel madura como las que contienen larvas, además reparan los panales que están rotos.

- **Almaceneras.** Son aquellas abejas que recogen el néctar que traen las abejas recolectoras y lo llevan a las celdillas de reservas. Junto con el néctar regurgitado cada abeja aporta enzimas digestivos que intervendrán en la fermentación del néctar y la fabricación de la miel. Hasta esta etapa las obreras no estaban lo suficientemente desarrolladas para volar, sin embargo, siendo almaceneras comienzan vuelos de orientación en la entrada de la colmena.
- **Guardianas.** Las guardianas se encargan de la vigilancia y defensa de la colonia. Sólo a finales de la tercera semana, la bolsa del veneno está repleta del mismo y algunas abejas se hacen guardianas. Vigilan la entrada para evitar que abejas de otras colonias o insectos extraños entren en la colmena y pretendan robar la miel. Las abejas son reconocidas, fundamentalmente, por su olor propio, segregado por una glándula llamada *glándula de Nassanof*, por eso, los machos, que carecen de ella, pueden entrar libremente en todas las colmenas.
- **Ventiladoras.** Una abeja por sí sola no es capaz de regular su temperatura, pero considerando el conjunto de la colmena como un superorganismo, éste puede mantener una cierta independencia térmica del ambiente. La colmena como superorganismo es capaz de mantener la temperatura de su centro entre unos 20 a 35 °C. Para el desarrollo de las larvas es preciso que en la zona de cría mantenga

una temperatura casi constante comprendida entre los 32° y 36 °C. Durante el verano, las abejas se colocan de forma que, con el movimiento de las alas, producen corrientes de aire dirigidas (aire acondicionado) y ventilan la colmena rebajando su temperatura. Para ello, algunas abejas situadas en la entrada, se alinean una tras otra y baten sus alas creando corrientes de aire fresco del exterior que ventilan la colmena. Estas abejas son las ventiladoras. Por otro lado, para producir calor durante el invierno las abejas lo obtienen del calor metabólico, haciendo que sus músculos torácicos se contraigan y relajen, manteniendo la misma actividad que durante el vuelo, aunque las alas permanezcan inmóviles. Además, para evitar la pérdida de calor por disipación del mismo por la superficie, las abejas se apiñan formando bolas y reduciendo la superficie de enfriamiento.

- **Rastreadoras.** Esta es la última etapa en la vida de una obrera, las rastreadoras son las que salen de la colmena en busca de alimento principalmente, aunque también se encargan de buscar agua, o un nuevo lugar donde vivir. La abeja rastreadora representa un excelente ejemplo de la organización y distribución de tareas, característico de las abejas. Se divide en dos roles diferentes, con el fin de evitar la búsqueda exhaustiva y masiva. El primero de estos roles es el de **abeja exploradora**; el cual lleva a cabo la búsqueda de alimento en las inmediaciones de la colmena. Cuando el alimento es escaso, se han visto abejas a 80 kilómetros de distancia de la colmena (Beekman *et al.* (2004)). Por lo general, son pocas las abejas que se dedican a explorar, esto es con el fin de minimizar el esfuerzo. La fuente de alimento de la colmena es el néctar de flores. Las exploradoras buscan zonas pobladas con flores. Cuando el alimento por fin se encuentra, entra en acción

el segundo rol: el de **abeja recolectora**. El cual consiste en la explotación masiva del recurso alimenticio.

Las abejas pueden ser rastreadoras a partir de las tres semanas de vida; y asumen uno u otro rol alternadamente según las necesidades de la colmena. Es la tarea más agotadora, y tras un período de actividad incesante suelen morir de agotamiento. Para darse cuenta del esfuerzo que implica, basta decir que para que una abeja llene su buche de néctar, es necesario que visite de 1000 a 1500 flores. La actividad cooperativa de la colmena es tal que una colmena puede almacenar hasta 1 kilogramo de miel al día, lo que significa que las abejas recolectoras deberán de hacer un promedio de 5 millones de viajes diarios.

VI.2.2 ¿Porqué abejas?

Como se dijo al principio de este capítulo, se han buscado nuevas formas de abordar problemas abstrayendo la esencia de procesos que se ven en la naturaleza y aplicándolos a diferentes problemas. La idea es trasladar la complejidad del problema hacia un nuevo punto de vista, consiguiendo una mejor comprensión y como consecuencia una forma novedosa y comparable en sus resultados con las anteriores. Con la presente tesis, se ha llevado a cabo una investigación sobre el comportamiento de las abejas para desarrollar un algoritmo inteligente que enfrente el problema de reconstrucción tridimensional de forma satisfactoria. Después de dar una introducción sobre el comportamiento de las abejas, a continuación se enlistan las principales características que motivan la utilización de su comportamiento como paradigma de reconstrucción.

- **Inteligencia.** Por mucho tiempo se ha pensado que las abejas reaccionan instintivamente de acuerdo al entorno que las rodea (von Frisch (1967)), pero en los últimos años los especialistas en la conducta animal se convencen de que actúan con cierto grado de autonomía, pero aún gobernadas por las necesidades colectivas, las cuales las dicta la reina mediante la segregación de feromonas ⁴ (Crist (2004)). Prueba de ello es su forma de comunicación basada en movimientos corporales, el cual ha sido bautizado como *el lenguaje de la danza*. Este lenguaje es la herramienta principal con la que las abejas exteriorizan información sobre localización de alimento, agua, de nueva casa e incluso, de clima. La intensidad de esta danza representa el grado de urgencia de la necesidad que están comunicando. Además son capaces de discriminar las opciones que se les presentan de acuerdo a diversos criterios, como la distancia, riqueza de la fuente de alimento, el clima y hasta el riesgo que implicaría el aventurarse a conseguir determinado recurso. Son también capaces de tomar decisiones de acuerdo a las necesidades de la colmena. En el caso de decisiones importantes como optar por una fuente de alimento o la mejor localización para construir una nueva colmena, tienen la capacidad de llegar a un consenso. Este tipo de inteligencia colectiva y *auto-organizada* sería de gran utilidad para enfrentar el problema de reconstrucción.
- **División de trabajo.** Una colmena se compone de miles de individuos, de entre los cuales, sólo unos pocos son enviados en busca de lo que es necesario para la colmena (abejas exploradoras). Una vez que lo encuentran, éstas regresan y

⁴Las feromonas son sustancias químicas, producidas por un animal que afecta a la conducta de otros de su misma especie.

reclutan compañeras mediante su lenguaje de danza, sólo entonces salen grandes enjambres en pos de lo que la compañera informó. De esta manera las abejas siguen un enfoque *divide y vencerás*, con lo que se ahorra esfuerzo laboral. Mientras algunos miembros de la colmena recolectan alimentos, otros se dedican a cuidar de las crías, de la reina, ventilar la colmena, protegerla de invasores, etc. gracias a un estricto régimen de castas. El enfoque divide y vencerás puede ser utilizado para llevar a cabo la reconstrucción en pequeñas partes de la escena.

- **Comunicación.** Para coordinar su trabajo y para que el esfuerzo individual forme parte de un todo organizado, las abejas deben mantenerse conscientes de la presencia de sus compañeras para el caso de que algunas de ellas se percate de la necesidad de la colmena antes que otras y se los transmita si es el caso. Con todo, este grado de libre albedrío, es limitado para una simple obrera, pues es la reina quien expresa las necesidades principales de la colmena a través de sus feromonas que viajan en el aire; pero son las abejas obreras las que deciden cómo y dónde realizar las actividades que se les pide, en otras palabras; se les dice que hacer, pero ellas deciden cómo llevarlo a cabo. Este grado de abstracción en el desempeño de una tarea resulta atractivo para aplicarlo en el problema de reconstrucción.
- **Sentido de orientación.** Como ya se ha comentado, las abejas poseen cierto grado de inteligencia, pero éste es tal, que son capaces de abstraer la información del exterior y transformarlo a su lenguaje, el cual realizan en el interior de su colmena. En la danza son capaces de transmitir con gran precisión la distancia

que las compañeras deben recorrer y la dirección con respecto a un punto de referencia a la que un recurso se encuentra. Junto con el sentido de orientación la colmena de abejas aplica las características que se acaban de numerar para un proceso que es esencial para su supervivencia: el proceso de búsqueda. De este proceso se habla con mayor detalle en la siguiente sección, ya que es el punto medular del algoritmo que se propone en este trabajo.

VI.2.3 El proceso de búsqueda.

Una de las tareas más importantes que lleva a cabo la colonia es la búsqueda, ya sea de alimento (el néctar de las flores), de agua o de un nuevo lugar para vivir. En el caso del alimento, la cooperación para buscar y recolectar nectar les permite aprovechar el de mejor calidad aún cuando esté muy disperso. La recolección de néctar se lleva a cabo en dos etapas: la exploración y la recolección. Las abejas rastreadoras en sus dos roles (exploradoras y recolectoras) son las que se encargan de llevar a cabo esta tarea.

El proceso comienza cuando una abeja que está inactiva recibe el estímulo de buscar alimento. Entonces toma el rol de *exploradora* y sale de la colmena en busca de fuentes de alimento valiéndose de su sentido del olfato y de la vista. Si después de cierto tiempo no encuentra lo que busca regresa a la colmena y vuelve a un estado inactivo o bien, a realizar otras tareas dentro de la colmena. En cambio, si tuvo éxito, toma una muestra de néctar del lugar y regresa. Al llegar comienzan a *reclutar* compañeras comunicándoles la dirección y distancia de la fuente de alimento mediante una danza. Si la distancia de la colmena a la fuente de alimento es corta (100 metros) la exploradora realiza la *danza en círculo* (round dance) con la que esparce el olor del tipo de

flor en el que ha estado para que las compañeras busquen este tipo de néctar hacia cualquier dirección. Para distancias largas la búsqueda se restringe estrictamente hacia determinada dirección y distancia; mediante la *danza con meneo* (waggle dance), donde la exploradora agita rápidamente el abdomen mientras camina mirando hacia una dirección particular; después da media vuelta regresando al punto de inicio y vuelve a menear el abdomen, vuelve a dar la media vuelta, pero esta vez por el lado contrario, y así sigue, describiendo un ocho al caminar, agitando el abdomen en la parte de en medio de este ocho (ver figura 36). La danza representa la codificación del lugar exacto donde fue encontrado el alimento. La duración del meneo es proporcional a la distancia que recorrió la exploradora para encontrar el alimento, y el ángulo de dirección con respecto al acimut solar (la componente horizontal de la dirección hacia el sol) es representado durante la danza en el ángulo entre la vertical dentro de la colmena y la dirección a la que apunta la cabeza de la abeja durante el meneo. La danza dura algunos minutos o hasta horas; dependiendo del entusiasmo con el que la exploradora la ejecute. El entusiasmo es proporcional a la riqueza del recurso y a las necesidades de la colmena. La abeja que está bailando tiene la capacidad de dejar de hacerlo si juzga que su información resulta irrelevante por sus compañeras. Si esto ocurre la abeja vuelve al estado de inactividad. En caso contrario las abejas que estén cerca de la exploradora recibirán la información y algunas acudirán al llamado convirtiéndose en *recolectoras*. Irán a la fuente de alimento y recolectarán todo el néctar posible. Al regresar lo pasan a otras abejas que se encargan de almacenarlo y repiten la danza para reclutar más compañeras, o bien, si el recurso se ha agotado, regresan al estado de inactividad. La recolección de néctar se encuentra resumido en el diagrama de flujo de la figura 37. Lo

que se acaba de decir aplica por cada abeja; tómesese en cuenta que el proceso se lleva a cabo en forma paralela y asíncrona; es algo extremadamente complejo si se contempla que son miles de abejas las que están involucradas en la tarea.

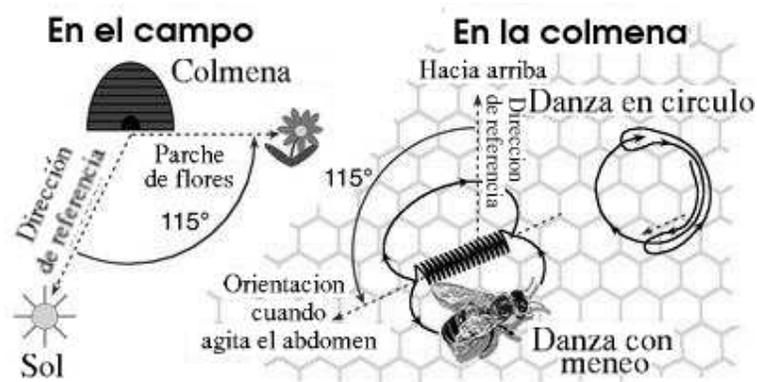


Figura 36: Significado de la danza con meneo.

VI.3 El proceso de búsqueda como paradigma de reconstrucción.

Ahora se procede a abstraer la esencia del proceso de búsqueda para desarrollar el algoritmo de reconstrucción. En la figura 37 se muestra un diagrama de flujo que representa las ideas expuestas en la sección anterior. Básicamente, el proceso de búsqueda se compone de tres tareas principales: *exploración*, *reclutamiento* y *recolección*, las cuales son asíncronas y paralelas, es decir, las llevan a cabo diferentes abejas, y mientras una está realizando una tarea, otra puede estar en la siguiente. Una abeja pasa de un estado inactivo a la etapa de exploración en la cual juega el papel de exploradora y viaja

distancias considerables en busca de recursos potenciales. En caso de no encontrarlos regresa al nido y regresa al estado inactivo, o bien, pasa a la etapa de reclutamiento en donde comparte información sobre la localización de recursos tales como néctar, polen, agua y material para construcción ⁵ y dependiendo de la calidad de los recursos anima a otras compañeras a recolectar dichos recursos. Este proceso se lleva a cabo indefinidamente, gobernado por las necesidades de la colmena.

VI.3.1 Estructura del algoritmo.

A continuación se presenta el algoritmo basado en el proceso de búsqueda de las abejas para obtener un modelo tridimensional sobre la escena. Este algoritmo se basa en el paradigma de estrategias de evolución $(\mu + \lambda)$ -ES. Sin embargo, la aplicación de los operadores evolutivos se hace de manera diferente. En una estrategia de evolución tradicional los operadores de mutación y cruzamiento se aplican o no se aplican sobre los padres de acuerdo a una probabilidad p_m y p_c . En este trabajo se sigue el paradigma propuesto por Boumaza (Boumaza y Louchet (2001)), donde una parte de la nueva población se genera exclusivamente de la mutación, otra parte del cruzamiento, y una parte más se genera aleatoriamente (véase figura 38). A continuación se presenta el pseudo-código:

Algoritmo 6.1

1. Inicializar población con μ_E individuos.
2. Evaluar población de acuerdo a una función de aptitud.

⁵Los materiales de construcción consisten en resinas de árboles que las abejas utilizan como *cemento*.

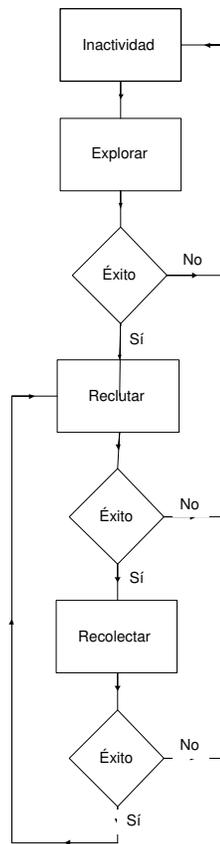


Figura 37: El proceso de búsqueda de las abejas se compone de tres tareas principales: exploración, reclutamiento y recolección.

3. Hacer: generación = 0.
4. Mientras no se llegue al número de generaciones preestablecido hacer lo siguiente:
 - (inicia ciclo)
 - (a) Aplicar selección por torneo.
 - (b) Crear α_E hijos con el operador de mutación.

- (c) Aplicar selección por torneo.
- (d) Crear β_E hijos con el operador de cruce.
- (e) Crear γ_E individuos generados aleatoriamente.
- (f) Evaluar α_E, β_E y γ_E , donde $\lambda_E = \alpha_E + \beta_E + \gamma_E$ de acuerdo a la función de aptitud.
- (g) Juntar poblaciones de padres e hijos en una sola: $\lambda_E + \mu_E$.
- (h) Aplicar operador de repartición a la población fusionada.
- (i) Escoger los mejores μ_E individuos.
- (j) Hacer: generación = generación+1.

(fin de ciclo)

5. Asignación de recolectoras a cada exploradora.

6. Asignación de espacio de búsqueda.

7. Por cada exploradora que reclutó recolectoras, hacer lo siguiente:

(inicia ciclo)

- (a) Inicializar población con μ_R individuos.
- (b) Evaluar población de acuerdo a una función de aptitud.
- (c) Hacer: generación = 0.
- (d) Mientras no se llegue al número de generaciones preestablecido hacer lo siguiente:

(inicia ciclo)

- i. Aplicar selección por torneo.
 - ii. Crear α_R hijos con el operador de mutación.
 - iii. Aplicar selección por torneo.
 - iv. Crear β_R hijos con el operador de cruce.
 - v. Crear γ_R individuos generados aleatoriamente.
 - vi. Evaluar α_R, β_R y γ_R , donde $\lambda_R = \alpha_R + \beta_R + \gamma_R$ de acuerdo a la función de aptitud.
 - vii. Juntar poblaciones de padres e hijos en una sola: $\lambda_R + \mu_R$.
 - viii. Aplicar *sharing* a la población fusionada.
 - ix. Escoger los mejores μ_R individuos.
 - x. Hacer: generación = generación+1.
- (fin de ciclo)

(fin de ciclo)

(fin del algoritmo)

En la primera línea del pseudocódigo, se inicializa la población de μ_E abejas exploradoras aleatoriamente mediante un generador de números aleatorios tomado de `VXL`⁶. Este generador arroja números de 32 bits con alto grado de aleatoriedad, uniformemente distribuidos. Para conocer un poco más sobre el generador de números aleatorios ver (Marsaglia y Zaman (1991)). En la línea dos se evalúa la población generada de acuerdo a la función de aptitud para las abejas exploradoras (ver VI.3.2). A partir de la línea

⁶ *Vision X Libraries*. Conjunto de librerías en C++ *open source* de gran utilidad para la comunidad de visión por computadora.

cuatro comienza la etapa de exploración con un ciclo de evolución para la población de abejas exploradoras; donde a partir de μ_E padres se crea una nueva población de λ_E hijos de los cuales:

- α se crean por mutación.
- β se crean por cruzamiento.
- γ se crean aleatoriamente.

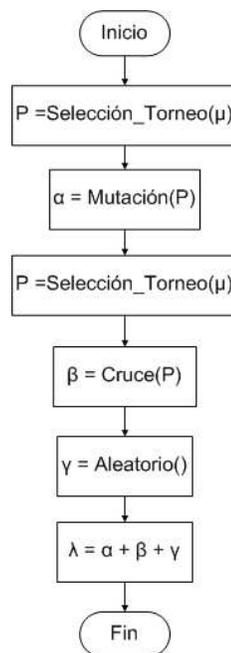


Figura 38: Diagrama de flujo que detalla el proceso de generación de la nueva población.

La población γ se utiliza para ayudar a mantener la diversidad en la población y para explorar nuevas regiones dentro del espacio de búsqueda. Con esto se pretende

simular lo que realmente ocurre en la naturaleza, pues el proceso de búsqueda de las abejas es asíncrono, es decir, las exploradoras van saliendo poco a poco de la colmena buscando nuevas fuentes de alimento. La aparición de una población aleatoria en cada generación simula esta acción. La selección de los mejores individuos para la aplicación de operadores se hace por torneo (ver sección V.3.3). Posteriormente, esta nueva población se evalúa. A continuación se juntan ambas poblaciones ($\mu_E + \lambda_E$) y se aplica el operador sharing, y en consecuencia la aptitud de algunos individuos disminuye (ver VI.3.3). De la conjunción de poblaciones se escogen los mejores μ_E individuos para conformar la nueva población, como se lleva a cabo en una estrategia de evolución ES- $(\mu + \lambda)$. Esta población se usa como entrada para la siguiente generación del ciclo evolutivo, y este ciclo sigue hasta que se satisfaga la condición de paro, que en este caso, es cumplir con determinado número de generaciones.

Posteriormente viene la fase de reclutamiento de recolectoras, compuesta por las líneas cinco y seis del algoritmo. A cada exploradora se le asigna determinado número de recolectoras de acuerdo a la evaluación que obtuvo por la función de aptitud, estas recolectoras buscarán alimento alrededor del punto donde la exploradora se encuentra, dentro de una pequeña región que representa su espacio de búsqueda. El espacio de búsqueda está en función inversamente proporcional a la distancia que existe desde la ubicación de las cámaras hasta el punto donde se encuentra la abeja exploradora.

En la línea siete se aprecia un ciclo que se lleva a cabo por cada exploradora que logró reclutar recolectoras, este ciclo marca el inicio de la *fase de explotación*. A partir de aquí la estructura del algoritmo es similar a la de la línea cuatro; existe un ciclo de evolución donde se crean λ_R hijos (aplicando mutación, cruzamiento y aleatoriamente)

a partir de μ_R padres (las abejas reclutadas), se juntan las poblaciones $\mu_R + \lambda_R$ y se conservan los μ_R mejores. El algoritmo termina cuando se ha alcanzado la condición de paro, que de nueva cuenta, es cumplir con determinado número de generaciones. El resultado es un conjunto de puntos en el espacio tridimensional que se ajustan a las coordenadas de los puntos de mayor relevancia en una escena (aquellos con mayor textura). En la figura 39 se presenta el diagrama de flujo del algoritmo que se acaba de describir. En lo que resta de este capítulo se explican a detalle las distintas partes del algoritmo y se presentan algunas pruebas para validar la contribución de cada una de ellas.

VI.3.2 Funciones de aptitud.

Como se dijo anteriormente, el algoritmo de las abejas formula el problema de la reconstrucción 3D como el problema de buscar una representación (un modelo) de la escena. Así, la población de individuos es un conjunto de puntos en el espacio tridimensional (x, y, z) que evolucionan a través de las generaciones gracias a los operadores evolutivos; cuya función objetivo debe contemplar el mejor ajuste posible a la escena. En un algoritmo evolutivo la función objetivo es equivalente a la función de aptitud con la que se evalúa a cada individuo. En el algoritmo que se presenta en este trabajo se utilizan dos funciones de aptitud, una para las abejas exploradoras y otra para las abejas recolectoras.

Abejas exploradoras.

Una abeja exploradora tiene como objetivo explorar el espacio de búsqueda y encontrar las mejores fuentes del recurso que se busca. Para llevar a cabo su trabajo, la

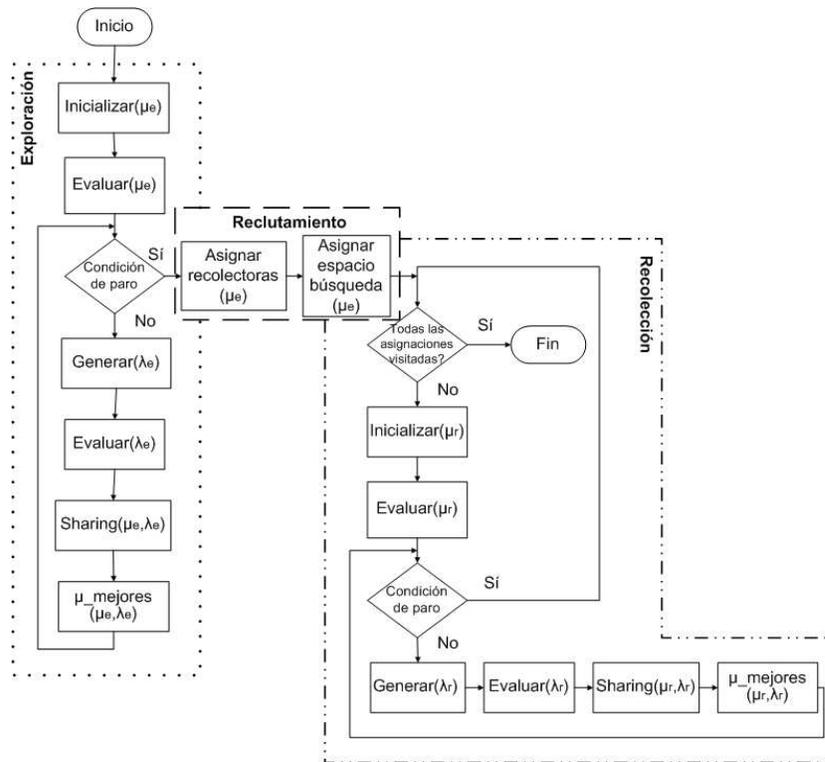


Figura 39: Diagrama de flujo que describe el algoritmo basado en el proceso de búsqueda de las abejas.

abeja exploradora debe aplicar dos criterios: el primero es saber si efectivamente se encuentra *parada* sobre un objeto y el otro es obtener una muestra del recurso para reclutar compañeras. Para corroborar que la abeja está próxima a la superficie visible de un objeto, se propone verificar en sus proyecciones en las dos imágenes. El problema se convierte entonces en un problema de búsqueda tridimensional cuyos criterios de búsqueda utilizan la información de las imágenes izquierda y derecha p_{Izq} y p_{Der} respectivamente. Por lo tanto será necesario proyectar sus coordenadas en cada imagen cada vez que se evalúe un individuo. Esto se lleva a cabo aplicando la matriz de

proyección calculada previamente con el método Faugeras-Toscani que se explica en la sección II.6.2. La función de aptitud para las abejas exploradoras se define como

$$\mathcal{F}_E = g(p_{Izq}, p_{Der}) \times f(p_{Izq}, p_{Der}), \quad (43)$$

donde $g(p_{Izq}, p_{Der})$ satisface el primer criterio, saber si la proyección de una abeja sobre la imagen está *parada* sobre un objeto. Esto se logra aplicando la función gradiente (∇). El gradiente es un operador diferencial que se utiliza en el análisis de imágenes para la detección de bordes (Gonzalez y Woods (2002)). Las derivadas pueden definirse como la razón de cambio de una función. La razón de cambio en una imagen es mayor cerca de los bordes porque existe un cambio en la intensidad del nivel de gris. El gradiente se aplica localmente píxel a píxel para ambas coordenadas (u, v) mediante una aproximación por diferencias, en este trabajo se utiliza el gradiente de sobel

$$\begin{aligned} \nabla_u &= (I(u+1, v-1) + 2[I(u+1, v)] + I(u+1, v+1)) - \\ &\quad (I(u-1, v-1) + 2[I(u-1, v)] + I(u-1, v+1)) \\ \nabla_v &= (I(u-1, v+1) + 2[I(u, v+1)] + I(u+1, v+1)) - \\ &\quad (I(u-1, v-1) + 2[I(u, v-1)] + I(u+1, v-1)) \end{aligned}$$

Donde $I(u,v)$ representa el valor de intensidad de nivel de gris en la coordenada (u,v) de una imagen. La detección de bordes se define entonces como el módulo del gradiente

$$\|\nabla\| = \sqrt{\nabla_u^2 + \nabla_v^2}$$

Considerando el gradiente en ambas imágenes, $g(p_{Izq}, p_{Der})$ queda definida de la siguiente manera

$$g(p_{Izq}, p_{Der}) = \|\nabla(p_{Izq})\| \times \|\nabla(p_{Der})\| \quad (44)$$

La detección de bordes no garantiza que una abeja se encuentre *parada* sobre un objeto, se puede presentar la situación de que una abeja que no está parada sobre un objeto, al momento de proyectarla sobre la imagen pareciera que sí lo está. Para evitar este problema se puede utilizar la información alrededor de la proyección. Para una abeja que está sobre un objeto las vecindades de sus proyecciones en ambas imágenes deberán ser semejantes. Por el contrario, si el punto 3D no está sobre una superficie, la proyección en una imagen será diferente en la otra. Este criterio se ilustra en la figura 40, inciso b). Para medir la semejanza entre las dos proyecciones se utiliza la correlación cruzada normalizada (ZNCC), definida en el capítulo IV.

$$f(p_I, p_D) = \frac{\sum_{i=-n}^n \sum_{j=-n}^n (I(x_I + i, y_I + j) - \overline{I(x_I, y_I)}) \cdot (I(x_D + i, y_D + j) - \overline{I(x_D, y_D)})}{\sqrt{\sum \sum (I(x_I + i, y_I + j) - \overline{I(x_I, y_I)})^2 \cdot \sum \sum (I(x_D + i, y_D + j) - \overline{I(x_D, y_D)})^2}}$$

El segundo criterio consiste en evaluar la calidad del recurso que ha encontrado la abeja exploradora. Se podría utilizar ZNCC como medida de calidad, sin embargo por sí sola no es suficiente, ya que mide el grado de semejanza, sin importar que en la región exista poca textura. Para ello se introduce un operador específico para medir textura: la homogeneidad. El descriptor de homogeneidad mide la semejanza que existe entre determinado píxel con respecto a sus vecinos cercanos, que en este caso, equivale a la medida de riqueza de determinado alimento. La homogeneidad está dada por

$$homog = \sum_i^n \sum_j^n \frac{M(i, j)}{1 + |i - j|}$$

Donde $M(i, j)$ es la *Matriz de Co-ocurrencia*, cada elemento de esta matriz representa la probabilidad relativa de que dos píxeles separados por cierta distancia en la imagen, tengan como niveles de gris i, j respectivamente.

Para que la función de aptitud favorezca a las exploradoras que hayan encontrado un recurso de mejor calidad, se ha agregado la siguiente cláusula de penalización

Sea la homogeneidad de una abeja i : $homog_I$ y $homog_D$ en la imagen izquierda y derecha respectivamente. Si $homog_I + homog_D > 0.3$ y $f(p_I, p_D) < 0.7$ entonces $\mathcal{F}_E = 0$

De esta manera la función de aptitud favorece a las abejas exploradoras colocadas en lugares con mucha textura.

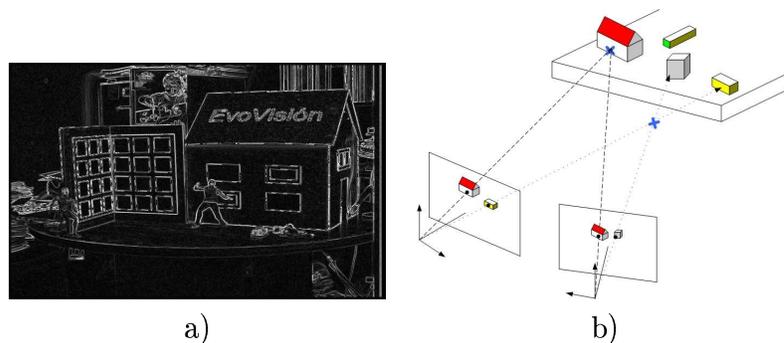


Figura 40: La función de aptitud de las exploradoras contempla dos cosas principalmente: Los contornos de los objetos con el gradiente de sobel (a) y la verificación de que la abeja se pose sobre el objeto (b).

Abejas recolectoras.

La función de aptitud para las abejas recolectoras está diseñada para explotar las intermediaciones de la fuente de alimento encontrada por una exploradora. Asume que la exploradora se encuentra parada sobre un objeto (el recurso que se busca) y que éste es de buena calidad. No se utiliza la función gradiente ya que en esta etapa del algoritmo se desea reconstruir hacia el interior del objeto previamente identificado, no en sus bordes. Sin embargo se necesita medir el grado de correlación con ZNCC para que la abeja no quede volando sobre el objeto de la misma manera que para las exploradoras. Además, el descriptor de homogeneidad ya no se maneja como una penalización sino que se incluye en la función de aptitud, esto es para favorecer a las recolectoras que se encuentren en regiones con alta textura sobre el objeto. En consecuencia el modelo tridimensional resultante conservará las regiones más significativas de la escena.

La función de aptitud para las abejas recolectoras queda de la siguiente manera

$$\mathcal{F}_{\mathcal{R}}(p_I, p_D) = f(p_I, p_D) \times h(p_I, p_D) \quad (45)$$

Donde $h(p_I, p_D)$ representa la medida de homogeneidad de ambas imágenes

$$h(p_I, p_D) = (1 - \text{homog}_I) \times (1 - \text{homog}_D)$$

La razón de aplicar la homogeneidad de esta manera (puede verse como *el complemento de la homogeneidad*) es porque se desea evitar las zonas semejantes (baja textura). Y la homogeneidad lo que mide es precisamente el grado de semejanza. Por eso se aplica el complemento.

VI.3.3 Los operadores evolutivos

Los operadores evolutivos aplicados de una generación a otra, representan el desplazamiento de un lugar a otro dentro del espacio de búsqueda por parte de las abejas. En esta sección se describe la funcionalidad de los operadores de selección, mutación, cruzamiento y repartición.

Selección.

La selección elige a las abejas con mejor valor de aptitud para que funjan como padres de la nueva población. De las diferentes formas de llevar a cabo la selección en la sección V.3.3, se ha elegido la selección por torneo, ya que es de las más populares y sencillas. Como se aprecia en la figura 38, antes de aplicar la mutación y el cruzamiento se eligen los individuos más aptos aplicando la selección por torneo a la población de padres. Primero se define el número de ganadores que se desean, después se eligen dos soluciones que compiten entre sí, aquella que obtenga un valor más alto en la función de aptitud gana su derecho a reproducirse. El número de torneos que se llevan a cabo depende de la población de mutación (α) y cruzamiento (β) que se deban generar.

Mutación.

El objetivo del operador de mutación es simular el desplazamiento que lleva a cabo una abeja dentro del espacio de búsqueda de una generación a otra. Como el espacio de búsqueda pertenece al conjunto de los reales, no se puede aplicar la mutación binaria. Se ha descartado la mutación aleatoria ya que, no encaja con el proceso de búsqueda de las abejas, ya que la generación de un nuevo hijo en cualquier lugar del espacio de búsqueda no se puede considerar como el movimiento coherente de una abeja de una generación a

otra. La mutación no uniforme, la polinomial y la normal son similares en el sentido de que las tres otorgan una mayor probabilidad a que hijo sea generado cerca del padre; lo cual es lo que se busca para el algoritmo. La mutación uniforme y la polinomial son parecidas, sin embargo, la primera no permite controlar la distancia a la que se genera el hijo con respecto del padre, mientras que la segunda si lo permite. Así, de los operadores de mutación que se muestran en la sección V.3.4, para el algoritmo desarrollado aquí son viables la mutación normal y la polinomial. En la sección VI.4.1 se presenta un estudio para elegir cuál de los dos se ha de implementar en la experimentación.

Cruzamiento.

Al igual que la mutación, el cruzamiento simula el desplazamiento de una abeja en el espacio, pero de una forma diferente; ya que extrae las similitudes de dos individuos de buen valor de aptitud para generar uno o dos nuevos individuos. Por la misma razón que para la mutación binaria, no se puede implementar el cruzamiento binario. Los dos operadores para espacios reales son el cruzamiento lineal y el binario simulado. Ambos son útiles para este problema. El primero introduce el concepto de colinearidad. Pues al cruzar dos individuos, el hijo resultante se encuentra en la línea recta que une a los padres. Así, el operador lineal favorece a la explotación de superficies planas, llenando paulatinamente los huecos entre individuos de buen valor de aptitud. Sin embargo si los individuos que se cruzan están sobre dos superficies distintas se produce un hijo que queda volando en el espacio. Esto no ocurre con SBX, el cual también puede llenar los huecos alrededor de los padres, permitiendo que se controle la distancia entre padres e hijos, además produce dos hijos en lugar de uno. En la sección VI.4.2 se presenta un estudio para elegir cuál de los dos se ha de implementar en la experimentación.

Repartición.

La repartición es un mecanismo de mantenimiento de la diversidad de la población, sirve para evitar la convergencia de la población hacia un único punto óptimo. La repartición permite que un algoritmo evolutivo alcance un objetivo considerado como más difícil que encontrar el óptimo global, el cual es encontrar múltiples soluciones que no necesariamente tienen el mismo valor de aptitud. En este caso, los nichos en los que se agrupan las soluciones representan regiones en donde el recurso que las abejas estén buscando son mejores (los objetos de la escena con la información más significativa). Conformando entre todas las regiones la solución global: un modelo tridimensional de la escena. El operador de repartición que se aplica en el algoritmo es el que se explica en la sección V.3.4.

VI.3.4 Reclutamiento.

En el proceso natural de búsqueda de algún recurso, las exploradoras que hayan encontrado una fuente abundante realizan la danza frenéticamente, con el fin de reclutar más compañeras. Entre mejor sea la calidad del recurso más compañeras pondrán atención a la danza e irán hacia la fuente. Esto es lo que se conoce como la etapa de reclutamiento.

La simulación de la etapa de reclutamiento se encuentra en las líneas cinco y seis del algoritmo 6.1: la asignación de recolectoras a cada exploradora y asignación de un espacio de búsqueda alrededor de cada exploradora.

Asignación de recolectoras.

La asignación de recolectoras es proporcional al valor en la función de aptitud que cada exploradora haya obtenido. Aquella con un valor de aptitud mayor será a la que se le

asignen mas recolectoras. La proporción p_i que le corresponde a cada exploradora se obtiene dividiendo su valor de aptitud entre la suma del de todas las exploradoras

$$p_i = \frac{apt_i}{\sum_{j=1}^N apt_j} \quad (46)$$

Y así se obtiene el número de recolectoras r_i que cada exploradora es capaz de reclutar

$$r_i = p_i \times \lambda_R, \quad (47)$$

donde λ_R es la población total de recolectoras. Nótese que es posible que no todas las abejas exploradoras

Asignación del espacio de búsqueda.

Las abejas reclutadas por cada exploradora irán a la localidad que ésta última les indicó, explotando el recurso en las cercanías de donde está la exploradora. Con este propósito se define un espacio de búsqueda por cada grupo de recolectoras. El tamaño de este espacio de búsqueda es proporcional a la distancia que existe desde la colmena (que es el lugar donde se encuentran las cámaras) hasta el punto donde se encuentra la exploradora. De esta manera, aquellas que están mas cercanas a la colmena obtendrán un espacio de búsqueda más grande con respecto a las que se encuentren a gran distancia, ya que las zonas más cercanas se les considera de mayor interés. Esta premisa es congruente con lo que ocurre en la realidad, pues las abejas prefieren concentrar sus esfuerzos en las zonas más cercanas a la colmena; a menos que la urgencia o la calidad de un recurso sea muy alta en un lugar lejano, con respecto a otros más cercanos.

Como punto de partida, la exploradora más cercana recibe un espacio de búsqueda de tamaño fijo τ_i . Sus compañeras reciben una versión reducida de éste, que empuerñece cada vez más conforme se van alejando. El volumen del espacio de búsqueda que se le asigna a la exploradora mas cercana está definido como el espacio de búsqueda que le correspondería a cada abeja exploradora si el volumen del espacio de búsqueda inicial V_O se repartiera equitativamente, en forma de cubo; así, τ_i representa la arista de este cubo

$$\tau_i = \sqrt[3]{V_O/\mu_E},$$

donde μ_E es la población de exploradoras. La reducción del espacio de búsqueda se basa en la información de distancia en píxeles d_i entre las proyecciones de la exploradora i en las imágenes.

$$d_i = \sqrt{(u_I - u_D)^2 + (v_I - v_D)^2}$$

Entre mayor sea la distancia, mayor es la profundidad a la que se encuentra la exploradora i . Con esta información se calcula la reducción red_i a la que cada espacio de búsqueda se somete

$$red_i = 0.5(1 - u_i) + u_i ,$$

donde u_i representa el grado de preferencia que cada espacio posee de acuerdo a su distancia de la colmena: d_i/d_{maximo} . La reducción multiplica el tamaño de la arista del espacio fijo para definir el tamaño de la arista que le corresponde a cada espacio

$$\tau'_i = \tau_i \times red_i \quad (48)$$

Nótese que el valor de red_i está en el rango $[0.5, 1]$; donde 0.5 se aplica al espacio de mayor distancia y 1 se aplica al más cercano.

VI.4 Experimentación.

En esta sección se hace una comparación entre los operadores evolutivos de los que se habló en la sección anterior. Para las pruebas se utilizó el par estéreo de la figura 16. A continuación se muestran pruebas para medir la efectividad de los diferentes operadores de mutación, cruzamiento y repartición; así como diferentes valores para sus parámetros.

VI.4.1 Pruebas con el operador de Mutación.

Las gráficas que se muestran aquí se refieren a la comparación de la mutación normal y la polinomial descritos antes. Los resultados presentados se obtuvieron a partir de una serie de 10 ejecuciones independientes, sobre el par estéreo de la figura 16, cada una de 150 generaciones para la etapa de exploración y 75 para la etapa de recolección. En la tabla VIII se muestran los parámetros que se usaron para esta prueba.

Una manera de comparar la efectividad de cada operador es midiendo el número de individuos exitosos que cada operador produce durante la generación en que fueron creados. Es decir, aquellos que son lo suficientemente aptos como para permanecer en la población que pasa a la siguiente generación. En la figura 41 se muestra el porcentaje de individuos exitosos creados por mutación. Se incluyó en la gráfica el promedio de

Tabla VIII: Parámetros del algoritmo para la comparación entre operadores de mutación.

Mutación Polinomial		Mutación Normal	
Población:			
μ_E	100	μ_E	100
λ_E	200	λ_E	200
μ_R	1000	μ_R	1000
λ_R	2000	λ_R	2000
Mutación:			
η_m	25	σ_{α_X}	2
		σ_{α_Y}	2
		σ_{α_Z}	2
Cruzamiento (SBX):			
η_c	2	η_c	2
Repartición:			
σ_{rep}	25	σ_{rep}	25
Hijos por operador: exploradoras			
Mutación α_E :	60	Mutación:	60
Cruzamiento:	10	Cruzamiento:	10
Aleatorio:	30	Aleatorio:	30
Hijos por operador: recolectoras			
Mutación:	600	Mutación:	600
Cruzamiento:	300	Cruzamiento:	300
Aleatorio:	100	Aleatorio:	100

éxitos en las 10 ejecuciones, así como el máximo y mínimo por cada generación. En el inciso (a) está la comparación para la etapa de exploración y en el inciso (b) para la etapa de recolección.

En la etapa de exploración la mutación normal claramente obtiene más éxito que la polinomial, sin embargo en la etapa de recolección ambos obtienen resultados muy similares. Un buen desempeño en esta prueba significa que el operador está generando

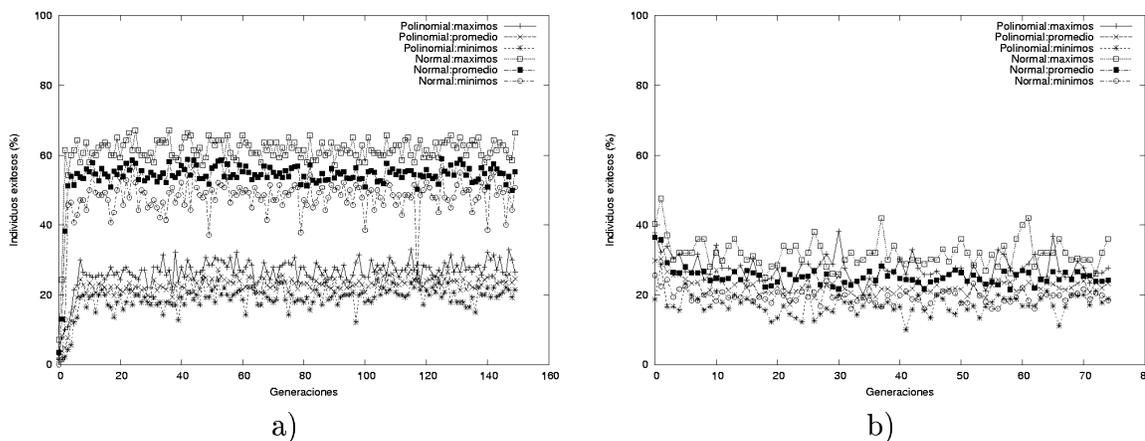


Figura 41: Comparación entre mutación normal y polinomial: porcentaje de individuos exitosos. (a) Etapa de exploración. (b) Etapa de recolección. (b)

individuos que por sí solos son lo suficientemente competitivos como para mantenerse en la población. Pero esta prueba no es decisiva, hace falta evaluar la aportación colectiva del total de individuos; para esto se lleva a cabo la prueba sobre el promedio del valor de aptitud y la diversidad de población. La figura 42 muestra la influencia de cada tipo de mutación sobre el valor de aptitud promedio y sobre el diversidad de la población, tanto en la etapa de exploración como en la de recolección. Para juzgar la calidad de los resultados del algoritmo, se deben revisar las dos gráficas simultáneamente. Como la solución se compone por el conjunto de individuos, el valor de aptitud promedio de éstos es un criterio de calidad. Sin embargo, la diversidad de estos individuos es un criterio de calidad complementario que también es necesario tomar en cuenta. Esta diversidad representa la diferencia entre los individuos de la población que para este problema se traduce como la distribución sobre el espacio de búsqueda. Para medir la diversidad de una población P de n individuos ($div(P)$) se ha decidido calcular el grado

de dispersión de la población total sobre las imágenes:

$$div(P) = \sqrt{\sum_{i=0}^n [(u_i - \bar{u}) + (v_i - \bar{v})]}$$

Donde u_i y v_i son las coordenadas del píxel donde se proyecta cada abeja. \bar{u} y \bar{v} representan la media de toda la población.

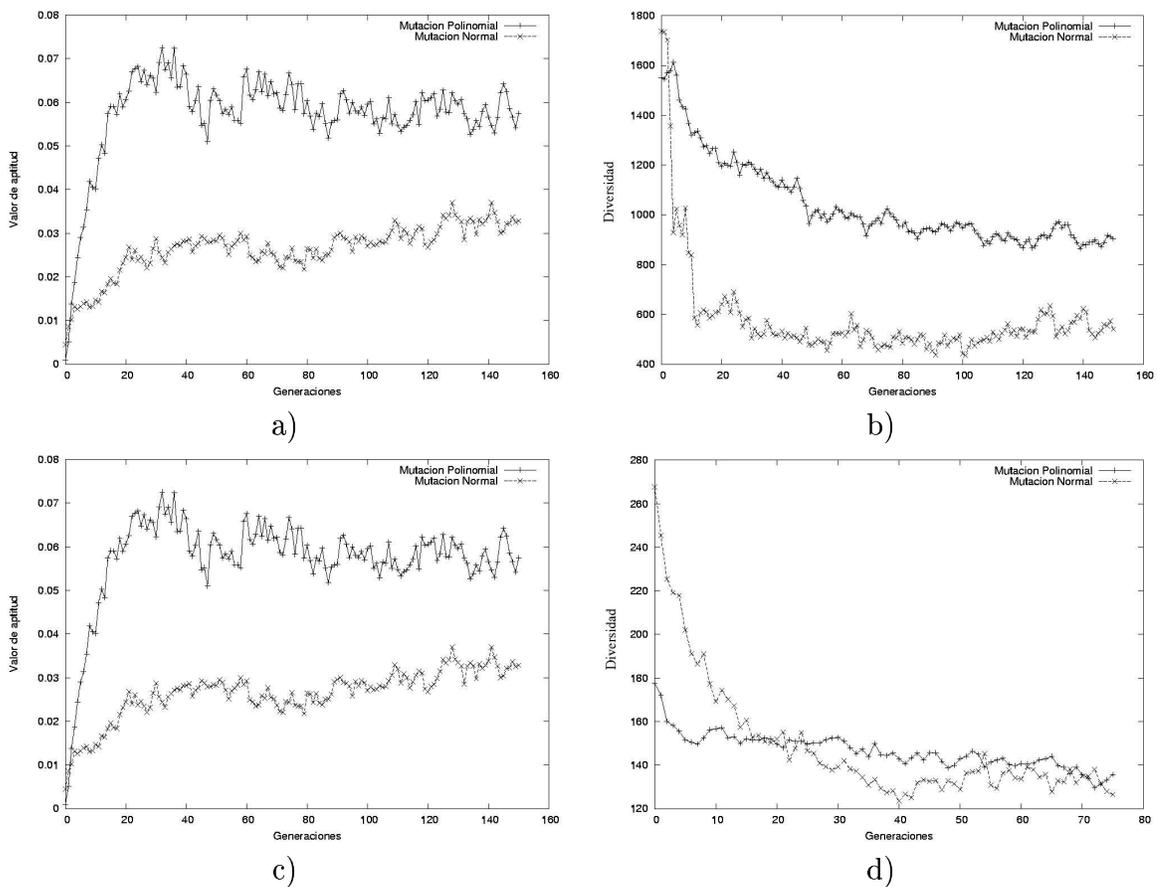


Figura 42: Comparación entre mutación normal y polinomial: Valor de aptitud promedio y diversidad de la población. (a y b) Etapa de exploración. (c y d) Etapa de recolección.

La figura 42, inciso (a), pone de manifiesto la supremacía del operador polinomial considerando el conjunto de individuos como la solución total. Lo mismo ocurre para el inciso (b), la mutación polinomial posee una mayor diversidad, lo que significa que la solución abarcará un espacio tridimensional mas amplio que la mutación normal. Nótese que entre más grande es el valor de aptitud, la diversidad es menor. De esto se hablará en la sección VI.4.3.

VI.4.2 Pruebas con el operador de Cruce.

Con el fin de comparar los resultados de los dos operadores de cruzamiento se han realizado las mismas pruebas que a los operadores de mutación sobre el mismo par de imágenes. Los parámetros para el algoritmo son similares, aquellos que cambian se muestran en la tabla IX. La decisión sobre qué operador usar para el capítulo de experimentación se obtiene a partir de la comparación del número de cruzamientos exitosos y de la evolución a través de las generaciones del valor de aptitud promedio simultáneamente con la diversidad. Nótese que el operador lineal no tiene parámetros; esto se debe a que el que utiliza es generado automáticamente, como variable aleatoria (ver sección V.3.4).

La figura 43 ilustra el porcentaje de individuos que, creados por cruzamiento, han resultado exitosos. En el inciso (a) está la comparación para la etapa de exploración y en el inciso (b) para la etapa de recolección. En la figura se puede apreciar que ambos operadores obtienen resultados muy parecidos en ambas etapas. Por claridad sólo se han graficado el promedio de los individuos exitosos en las 10 ejecuciones. En la etapa de exploración el porcentaje de exitosos se mantiene alrededor del 20% a lo largo de las

Tabla IX: Parámetros del algoritmo para la comparación entre operadores de cruzamiento.

Cruzamiento SBX		Cruzamiento lineal	
Mutación Polinomial:			
η_m	25	η_m	25
Cruzamiento:			
η_c	2	-	-

generaciones, en cambio, en la etapa de recolección se presenta con cierta inestabilidad entre el 15 y 20%.

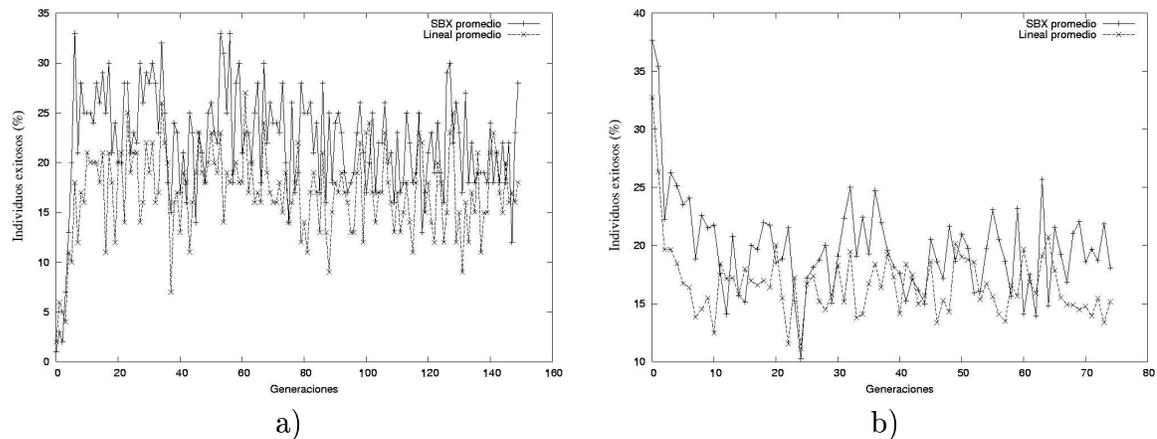


Figura 43: Comparación entre cruzamiento SBX y lineal: porcentaje de individuos exitosos. (a) Etapa de exploración. (b) Etapa de recolección. (b)

La figura 44 muestra la influencia de cada operador de cruzamiento sobre el valor de aptitud promedio y sobre la diversidad de la población. Al igual que se dijo para los

operadores de mutación, tanto el valor de aptitud como la diversidad son importantes para obtener una buena reconstrucción. Pues entre mayor sea el valor de aptitud, menor es la diversidad. Sobre esta cuestión se hablará con mayor detalle en las pruebas con el operador de repartición.

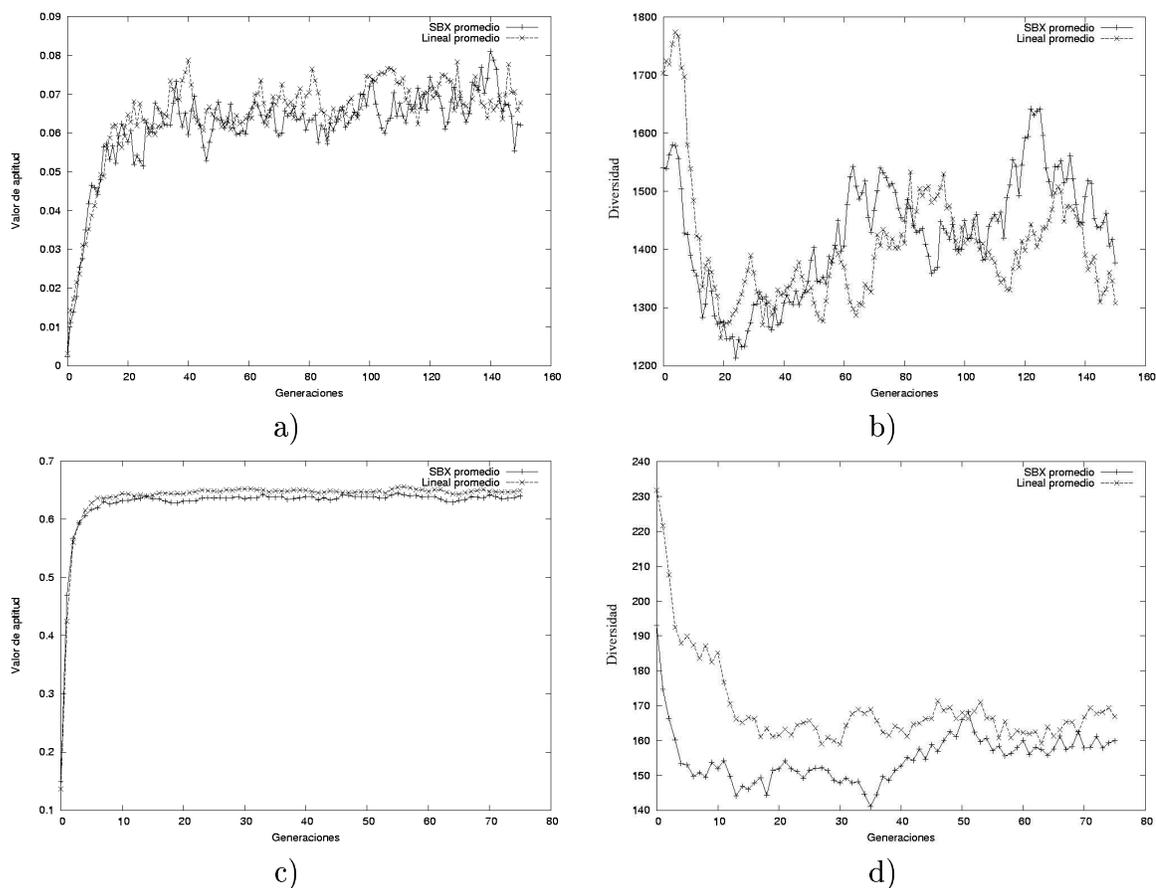


Figura 44: Comparación entre cruzamiento SBX y lineal: Valor de aptitud promedio y diversidad de la población. (a y b) Etapa de exploración. (c y d) Etapa de recolección.

Al igual que en la figura 43, ambos operadores muestran rendimientos muy similares en ambas etapas. Esto permite predecir que la elección de uno u otro operador proporcionará resultados muy similares.

VI.4.3 Pruebas con el operador de repartición.

En la prueba que sigue a continuación, se compara el resultado del algoritmo con coeficientes de repartición diferentes (σ_{rep}), los demás parámetros son los de la tabla VIII y permanecen fijos.

En el inciso (a) de la figura 45 se muestra la evolución del valor de aptitud a través de las generaciones con distintos valores para σ_{rep} . Nótese que entre más disminuye σ_{rep} , el valor de aptitud es más grande. Esto se debe a que un coeficiente de repartición pequeño reduce la diversidad de la población (figura 45, inciso (b)). El valor de aptitud para σ_{rep} pequeños es grande, no obstante los resultados de reconstrucción serán pobres, pues una gran parte del espacio tridimensional no será representado. Por otro lado, con un coeficiente de repartición mayor, la diversidad de la población es más importante como se aprecia en la figura 45, inciso (b), en este caso, los individuos visitan un gran número de óptimos locales. Estos óptimos tendrán un valor de aptitud más bajo, lo que provoca la reducción del valor de aptitud promedio de la población. Estas gráficas refuerzan la idea mencionada antes, que la calidad de los resultados debe juzgarse no sólo sobre el valor de aptitud, sino sobre la diversidad de la población.

Viendo ambas gráficas simultáneamente se aprecia fácilmente los dos objetivos que deben alcanzarse. Desde este punto de vista es posible formular el problema como bi-objetivo: obtener el mejor valor de aptitud con la mayor diversidad.

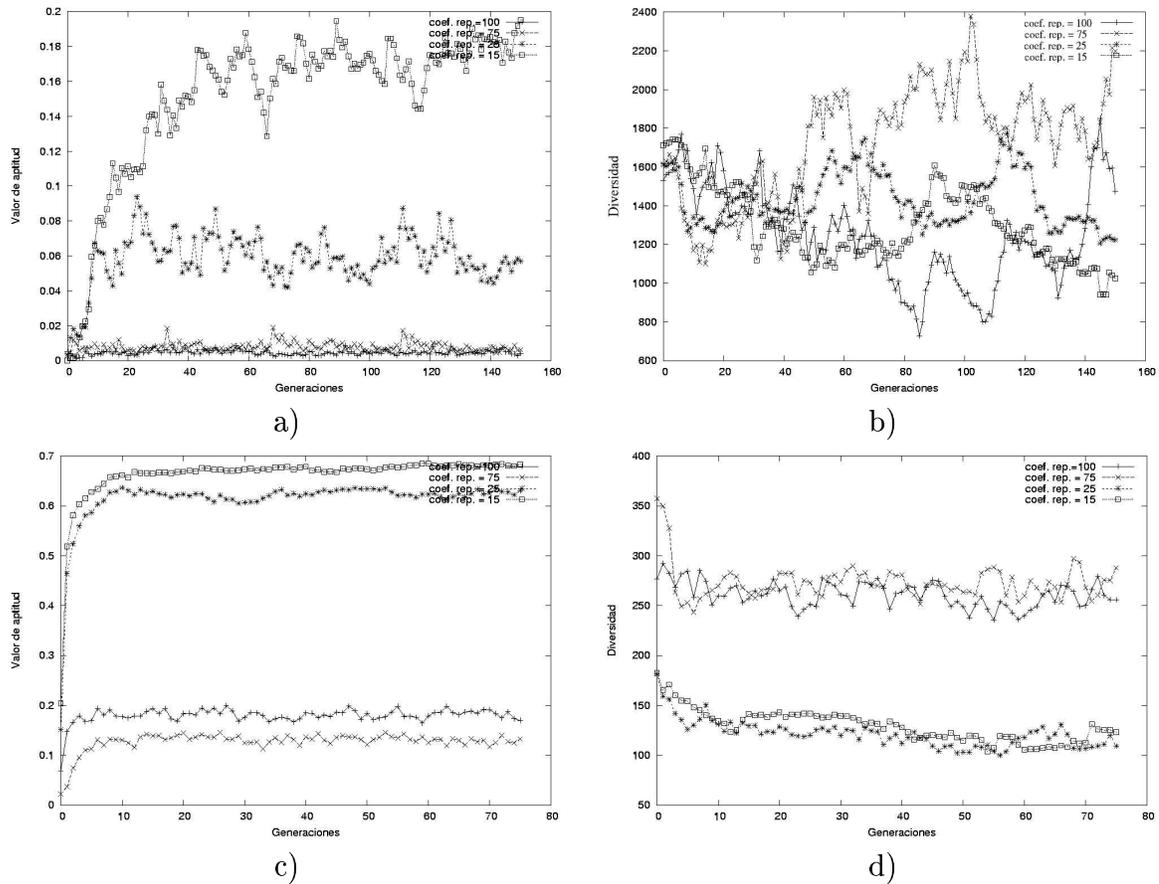


Figura 45: Comparación entre diferentes coeficientes de repartición: Valor de aptitud promedio y diversidad de la población. (a y b) Etapa de exploración. (c y d) Etapa de recolección.

VI.5 Conclusiones.

En este capítulo se desarrolló el algoritmo de búsqueda basado en las abejas. Se ha abstraído la esencia de un proceso natural y con ella se ha desarrollado un nuevo enfoque para abordar el problema de reconstrucción tridimensional. El nuevo enfoque consiste en desarrollar una técnica inteligente que favorezca las regiones que por su textura y

cercanía, son las más significativas de la escena, resultando en una reconstrucción quasisensa, pero dirigida y mejorada. Se ha decidido aplicar el comportamiento de las abejas, ya que se ha demostrado que son capaces de interpretar, memorizar y discriminar la información según criterios flexibles de acuerdo a las circunstancias (Crist (2004); von Frisch (1967)). Conceptos tan complejos como reconocimiento, búsqueda, comprensión, toma de decisiones y consenso, los aplican de una manera cooperativa y organizada en la lucha diaria que sostienen para sobrevivir. A todos estos conceptos se les considera del mayor interés dentro de la visión por computadora, lo que permite afirmar que un algoritmo inspirado en el comportamiento de las abejas puede aportar nuevos puntos de vista sobre problemas que aún no se han resuelto dentro de esta disciplina.

Los operadores evolutivos son fundamentales para el proceso de búsqueda, ya que simulan el movimiento de las abejas en el espacio. Se analizaron diversos operadores, sobre los de mutación se puede decir que el polinomial es mejor para este problema que la mutación normal; ya que su capacidad para alcanzar diversidad es fundamental para un problema multimodal como éste. Entre los operadores de cruzamiento, se ha decidido por SBX, pues aunque ambos han demostrado ser iguales en cuanto a efectividad, SBX cuenta con una variable determinística para controlar la distancia a la que los hijos son generados η_c ; no así el cruzamiento lineal, cuya variable estocástica α provoca una tendencia al comportamiento impreciso del algoritmo. Del operador de repartición se evaluaron diferentes valores para el coeficiente de repartición σ_{rep} para elegir el que mejor se acomode al tipo de imágenes que se utilizan. observando las gráficas del valor de aptitud y la diversidad se puede intuir que este problema podría abordarse desde el punto de vista de la búsqueda multi-objetivo.

Para complementar los comentarios que se acaban de mencionar, es necesario presentar resultados con imágenes de escenas reales y revisar el comportamiento del algoritmo. Esto se lleva a cabo en el siguiente capítulo, y por último, en el capítulo VIII se presentan las reflexiones finales y en perspectiva, así como el trabajo futuro sobre el desarrollo, adaptación y aplicación de este algoritmo a otras tareas del área de la visión por computadora.

Capítulo VII

Experimentación.

En este capítulo se lleva a cabo la reconstrucción tridimensional aplicando el algoritmo de búsqueda basado en abejas. Al igual que para el método de triangulación del capítulo IV, se utilizan los pares estéreo de la figura 16 y 25. En el capítulo anterior se realizaron experimentos con varios operadores. Aquí se aplican los mejores operadores para obtener resultados sobre reconstrucción tridimensional.

El algoritmo de búsqueda basado en abejas se ha implementado como complemento en una aplicación llamada *xcv*, que utiliza el paquete de funciones de programación en lenguaje C++, portables y eficientes *VXL*. *VXL* (en inglés, *Vision x Libraries*) diseñadas para facilitar la investigación en el campo de la visión por computadora. *VXL*, posee un conjunto de algoritmos numéricos programados en C++ (*vnl*), administradores de imágenes (*vil*) y de video (*vidl*), interfaz gráfica de usuario y especializadas en visión por computadora: procesamiento de imágenes (*vipl*), geometría epipolar (*mvl*) entre otras.

El objetivo de este experimento es obtener el modelo tridimensional del par estéreo 16 y 25. Habiendo comparado los diferentes operadores en el capítulo anterior, se ha decidido usar mutación polinomial y cruzamiento binario simulado (SBX). Se presentarán distintas ejecuciones del experimento con una población diferente cada una, los demás parámetros permanecerán fijos. El número de generaciones será de 60 para las

exploradoras, y de 30 para las recolectoras. Esto debido a que en las pruebas anteriores la valor de aptitud y el de diversidad se estabilizan alrededor de la generación 45 para las exploradoras y de 15 para las recolectoras. Exceptuando las poblaciones, todas las pruebas tienen los mismos parámetros de la tabla X. El total de los parámetros que se usaron para generar el modelo son los siguientes:

Tabla X: Parámetros del algoritmo para generar un modelo 3D a partir de los pares estéreos de las figuras 16 y 25.

Prueba 1		Prueba 2		Prueba 3	
Población:					
μ_E	200	μ_E	800	μ_E	1600
λ_E	400	λ_E	1600	λ_E	3200
μ_R	2000	μ_R	8000	μ_R	16000
λ_R	4000	λ_R	16000	λ_R	32000
Mutación Polinomial					
η_m	25	η_m	25	η_m	25
Cruzamiento (SBX):					
η_c	2	η_c	2	η_c	2
Repartición:					
σ_{rep}	25	σ_{rep}	25	σ_{rep}	25
Hijos por operador: exploradoras					
Mutación:	240	Mutación:	960	Mutación:	1920
Cruzamiento:	40	Cruzamiento:	160	Cruzamiento:	320
Aleatorio:	120	Aleatorio:	480	Aleatorio:	960
Hijos por operador: recolectoras					
Mutación:	2400	Mutación:	9600	Mutación:	19200
Cruzamiento:	1200	Cruzamiento:	4800	Cruzamiento:	9600
Aleatorio:	400	Aleatorio:	1600	Aleatorio:	3200

VII.1 Experimento I.

El experimento I consiste en aplicar el algoritmo al par estéreo de la figura 16 y obtener un modelo tridimensional disperso o quasi-denso. En la figura 46, se muestran tres diferentes modelos con las poblaciones de la tabla X. La primera columna de la figura contiene las proyecciones en la imagen izquierda de las abejas que conforman el modelo tridimensional de la segunda columna. En los incisos (a) y (b) se muestra la primera prueba, con una población inicial de $\mu_E = 200$ padres y $\lambda_E = 400$ hijos para las exploradoras, y $\mu_R = 2000$ padres y $\lambda_E = 4000$ hijos para las recolectoras. Se debe recordar que el algoritmo se basa en una estrategia de evolución $(\mu+\lambda)$ -ES, por lo que al final de cada generación, cuando se preservan los μ mejores individuos, padres e hijos tienen la misma oportunidad de ser elegidos, siempre y cuando su valor de aptitud sea competitivo. Con esta población se genera un modelo tridimensional con 2200 abejas (incluyendo exploradoras y recolectoras). Son pocas las zonas que se pueden reconstruir con tan pocos puntos, tomando en cuenta que cada una de las imágenes tiene una dimensión de 768×484 , es decir, de 371,712 puntos. Sin embargo, se aprecian varios cúmulos de abejas en donde consideraron que era una zona de gran riqueza, principalmente en la mira de calibración. Donde las abejas se organizaron y trabajaron en conjunto para reconstruir una buena parte de uno de los planos de la mira de calibración.

En los incisos (c) y (d), la población final que lleva a cabo la reconstrucción es de 8,800 individuos, incluyendo exploradoras y recolectoras. Se aprecia un gran incremento en los detalles de la reconstrucción. La mira de calibración está más poblada de abejas, así como la caja encima de la mira, con el grabado de la cara de una persona. Obsérvese

que ya se aprecian los bordes geométricos de la casa en el modelo tridimensional. El muñeco enfrente de la casa también se ve poblado de abejas que se han posado sobre él. En los incisos (e) y (f) la población es de 17,600 individuos en total. Los bordes del plano izquierdo de la mira están completamente reconstruídos, lo mismo ocurre con el muñeco enfrente de la casa. En la caja encima de la mira, el grabado con la cara de una persona ha sido completamente reconstruído, así como los bordes del tejado de la casa, los bordes verticales y la palabra Evovisión, que está escrita en el tejado. Obsérvese cómo la reconstrucción se centra en los bordes, como la cuadrícula de la mira, y las orillas de la casa, también enfoca sus esfuerzos en las zonas de alta textura como las letras del tejado y el grabado de la caja encima de la mira. La reconstrucción es significativa a pesar de estar usando menos del 5 % de los 371,712 puntos que se podrían utilizar.

VII.1.1 Comparación con la triangulación.

Es posible comparar la triangulación y el algoritmo de búsqueda basado en abejas cualitativa y cuantitativamente. La primera es comparando ambos modelos tridimensionales y la segunda es comparando las mediciones de ambos métodos sobre el único objeto del que se conocen sus dimensiones en la escena: la mira de calibración. Para hacer una comparación justa, véase la figura 24, inciso (d) y la figura 46, inciso (f). El primer modelo 3D se compone por 12,661 puntos, mientras que al segundo lo conforman 17,600 individuos. Hay casi cinco mil puntos de diferencia, pero esta diferencia es mucho menor que la diferencia entre 46, (f) y 24, (c), teniendo este último 62,251 puntos triangulados.

El modelo obtenido con el algoritmo de búsqueda tiene una calidad mejor que el modelo obtenido por triangulación. De hecho, pareciera que el modelo obtenido por triangulación tuviera muchos menos puntos de los que en realidad tiene. Esta ilusión se debe a que una parte considerable de ellos está fuera del lugar que le corresponde, como se puede ver con los puntos que están en la parte de atrás de la casa; también se debe a que los puntos se encuentran muy dispersos. Aún el modelo de la figura 46, inciso (d), que con sus 8,800 individuos, podría considerarse como un modelo mejor que el modelo triangulado, ya que los objetos se ven más definidos en el resultado del algoritmo bio-inspirado que en el del geométrico. De la misma manera se puede comparar 46, (f) con 24, (c). A pesar de la diferencia de puntos, los modelos son comparables gracias a la filosofía de agruparse y trabajar en equipo tomada del comportamiento de las abejas; con lo cual se ganan dos cosas, en primer lugar una reconstrucción compacta de los lugares más significativos y en segundo lugar una propagación de la precisión, esto es, una exploradora que está posada sobre un objeto atraerá a varias recolectoras sobre el mismo, evitando así que queden individuos *volando*, y no parados sobre los objetos.

La manera cuantitativa de comparar ambos métodos es medir las dimensiones de la mira de calibración. En la tabla VI se presentan las medidas del método de triangulación para doce esquinas de la cuadrícula de la mira de calibración; y en la tabla XI se presentan las medidas de los mismos puntos usando el método bio-inspirado, en particular, del modelo de la figura 46, inciso (d). Compárense ambas tablas, claramente se ve un error más pequeño en XI. El error más grande en esta tabla es de 9.97 milímetros, mientras que en la tabla VI es de 15.57 milímetros, el error promedio también es más pequeño; un 3.90 mm contra 7.72mm de la triangulación. Sin embargo, el enfoque

Tabla XI: Medida de error en el modelo tridimensional del par estéreo de la figura 16

X Escena	Y Escena	Z Escena	X Modelo	Y Modelo	Z Modelo	Dif. X	Dif. Y	Dif. Z	Dist. Euclidiana
175	0	0	173.69	-1.15	-4.97	1.30	1.15	4.97	2.72
175	0	105	174.23	-8.43	134.68	0.77	8.43	29.69	9.97
175	0	225	172.95	-1.42	226.97	2.05	1.42	1.97	2.33
70	0	0	64.40	13.15	-0.39	5.60	13.15	0.39	4.37
70	0	105	69.94	1.06	101.34	0.06	1.06	3.66	2.18
70	0	225	59.68	8.07	229.56	10.32	8.07	4.56	4.79
0	70	0	-0.16	69.76	-5.94	0.16	0.24	5.94	2.51
0	70	105	1.78	68.59	105.74	1.78	1.41	0.71	1.97
0	70	225	-5.24	64.31	223.99	5.24	5.69	1	3.45
0	175	0	4.99	179.22	15.86	4.98	4.22	15.86	5
0	175	105	1.48	168.58	104.30	1.48	6.42	0.69	2.93
0	175	225	-0.16	191.11	229.38	0.15	16.11	4.39	4.54

geométrico tiene el error más pequeño, 1.75 mm contra 2.18 mm del bio-inspirado; aunque la diferencia es muy pequeña.

VII.2 Experimento II.

El objetivo de este experimento es obtener el modelo tridimensional, disperso o quasi-denso, del par estéreo de la figura 25. Utilizando los parámetros de la tabla X, se obtuvieron los resultados que se aprecian en la figura 47

Esta imagen es muy diferente a la primera. Posee más textura, pero le hacen falta bordes rectilíneos. Nótese que los individuos se siguen agrupando, sin embargo, en la ausencia de bordes, no existe una precisión tan buena como la del par de imágenes

Tabla XII: Medida de error en el modelo tridimensional del par estéreo de la figura 25

X Escena	Y Escena	Z Escena	X Modelo	Y Modelo	Z Modelo	Dif. X	Dif. Y	Dif. Z	Dist. Euclidiana
175	0	0	168.57	-7.00	-8.70	6.43	7.00	8.70	4.70
175	0	105	163.98	3.24	121.43	11.02	3.24	16.44	10.85
175	0	225	189.08	13.05	200.57	14.08	13.05	24.43	7.18
70	0	0	90.44	-2.23	-1.39	20.44	2.23	1.39	4.9
70	0	105	116.99	7.60	105.17	46.99	7.59	0.17	7.4
70	0	225	96.36	48.49	218.55	26.36	48.49	6.45	9
0	70	0	0.59	40.06	2.78	0.59	29.94	2.78	5.77
0	70	105	24.27	74.76	102.34	24.27	4.76	2.65	5.62
0	70	225	0.60	91.37	223.51	0.6	21.36	1.48	4.84
0	175	0	1.76	162.87	2.12	1.75	12.12	2.12	4
0	175	105	-6.69	172.86	105.69	6.69	2.14	0.69	3
0	175	225	193.19	275.13	207.89	193.19	100.13	17.10	17.6

anterior. Aunque en esta imagen está la mira de calibración, la blusa de la persona que está junto a la mira representa un recurso más rico, ya que posee más textura, y por lo tanto una gran cantidad de abejas convergen en ella.

VII.2.1 Comparación con la triangulación.

Una vez más, se toma como medida de comparación la mira de calibración aunque para esta imagen la mira de calibración no posee demasiados puntos.

Comparando cuantitativamente la tabla VII con la tabla XII ambas poseen un error similar. En este caso el error más grande lo obtuvo el método bio-inspirado con 17.6 mm, contra el apenas inferior 16.52 del método geométrico. Lo mismo ocurre con el error más pequeño, el cual pertenece al método geométrico. Sin embargo el mejor promedio

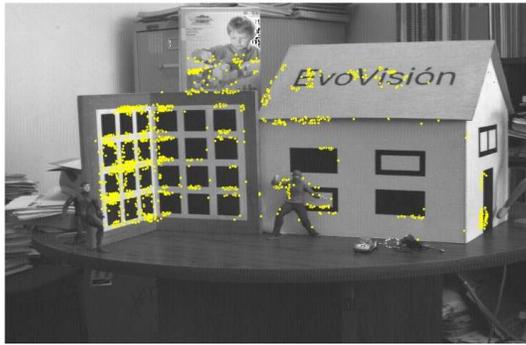
lo tiene el bio-inspirado, con un 7.07mm, mientras que la triangulación alcanza los 9.27 mm en su promedio. Esto se debe a que una buena parte de los errores en la tabla VII están por encima de 10mm.

VII.3 Conclusiones.

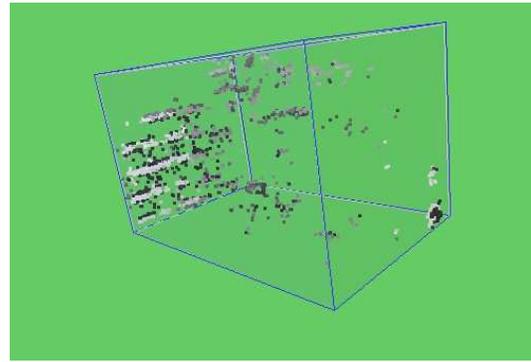
Hasta antes de este capítulo se explicó el proceso de búsqueda de recursos de una colmena de abejas, se abstraigo la esencia del concepto y se adaptó a un algoritmo evolutivo: las estrategias de evolución. Se propusieron diferentes operadores genéticos para explorar y explotar el espacio de búsqueda de la mejor manera posible. En este capítulo se aplica experimentalmente el algoritmo a dos pares estéreo reales, con características diferentes uno del otro.

La ambigüedad sobre la localización exacta de un punto en el espacio la generan problemas como la oclusión y la falta de textura. esto ocasiona modelos tridimensionales con demasiados puntos incoherentes con el modelo original. Una manera de evitar los resultados erróneos es aplicando rigurosos criterios que los descarten. Sin embargo, esto acarrea una reconstrucción con huecos debido a la dispersión de los puntos que prevalecen. En este capítulo se ha mostrado que el algoritmo es comparable cualitativa y cuantitativamente con el método popular de triangulación. La ventaja del método basado en el proceso de búsqueda de las abejas es la robustez contra puntos dispersos y fuera de lugar, pues genera agrupaciones compactas de puntos coherentes con la escena que se está modelando. Esto se debe al comportamiento cooperativo, organizado e inteligente basado en el comportamiento de las abejas que se le proporcionó al algoritmo, con el cual se obtiene una reconstrucción dispersa mejorada o quasi-densa. Este

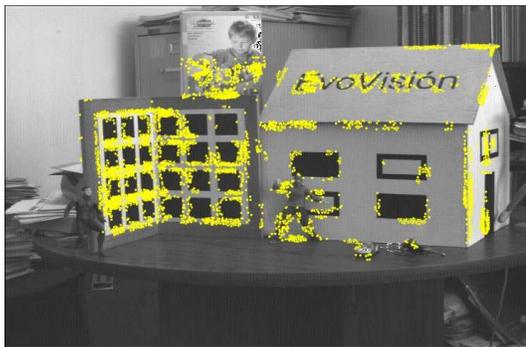
comportamiento puede verse como una estrategia *divide y vencerás*, en la cual unos pocos individuos exploran el espacio de búsqueda y una vez que encuentran un recurso factible guían a otros individuos, usando información de textura y de correlación, para que exploten el recurso.



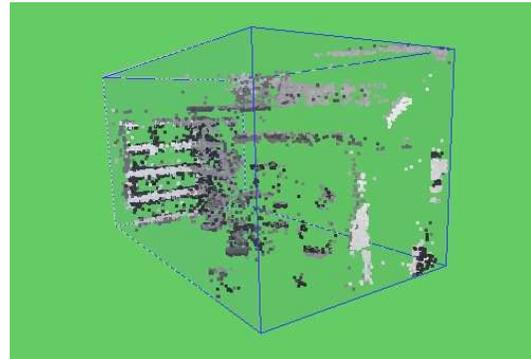
a)



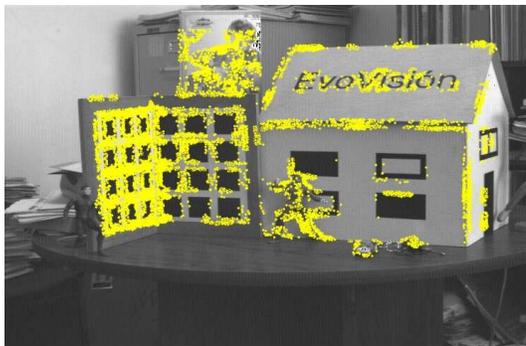
b)



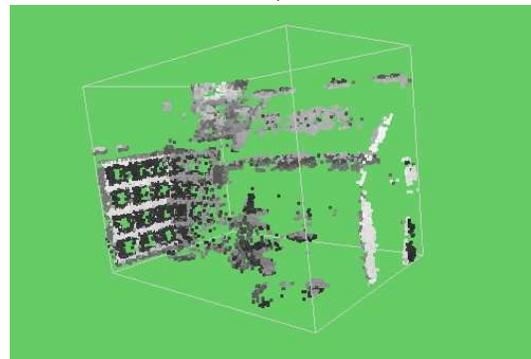
c)



d)



e)



f)

Figura 46: Resultados del experimento I con el par estéreo de la figura 16

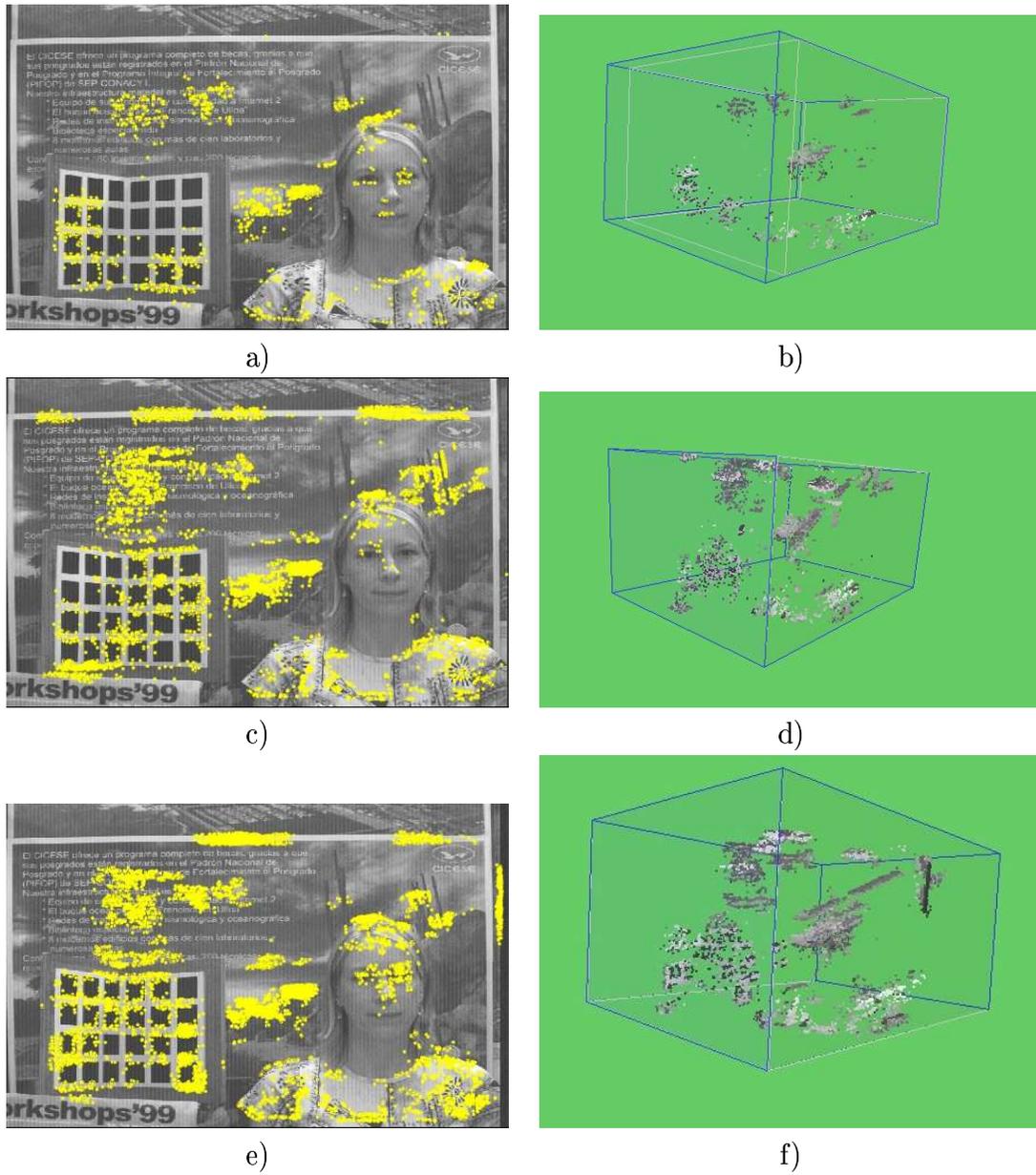


Figura 47: Resultados del experimento II con el par estéreo de la figura 25

Capítulo VIII

Conclusiones y trabajo futuro.

En este trabajo se ha propuesto un novedoso enfoque biológicamente inspirado para abordar, desde otro punto de vista, el problema de la reconstrucción tridimensional: un algoritmo basado en el proceso de búsqueda de las abejas.

La reconstrucción tridimensional es un problema complejo que aún no ha sido resuelto satisfactoriamente, existen diversas técnicas y métodos propuestos por los especialistas en visión por computadora y en fotogrametría, pero ninguno de ellos resuelve el problema de manera general que pueda afrontar cualquier circunstancia. La reconstrucción puede llevarse a cabo mediante la geometría de múltiples vistas. La más simple de ellas es la que está compuesta por dos imágenes. Una buena reconstrucción a partir de dos imágenes contiene muchas zonas ambiguas de las que no se puede determinar fácilmente su posición en el espacio. Estas zonas están sujetas a problemas de oclusión y de baja textura principalmente. Esto ocasiona modelos tridimensionales con demasiados puntos incoherentes con el modelo original. Uno de los objetivos de este trabajo fue conocer y analizar a fondo el enfoque geométrico. De esto trata la primera parte del trabajo. Sin esta parte del trabajo hubiera sido difícil dar con el proceso natural adecuado para encarar el mismo problema, pero desde otra perspectiva.

El algoritmo que se desarrolló en el capítulo VI y los resultados del capítulo VII

demuestran que el proceso de búsqueda de recursos de una colmena de abejas es apropiado para encarar el problema de reconstrucción tridimensional, ya que en esencia, el objetivo de este proceso natural es encontrar objetos en el espacio tridimensional. La forma en que la colmena de abejas encara este problema es una estrategia *divide y venceras*, ya que los individuos poseen un rol, que, aunque pueda cambiar con el tiempo, les permite cumplir con un objetivo tan grande como la alimentación de la colonia entera mediante esfuerzos individuales pero coordinados y organizados, como la exploración y recolección de néctar.

Trasladando estas ideas al problema de reconstrucción, el algoritmo que se propone en este trabajo pretende evitar resultados erróneos, enfocándose en la búsqueda de información significativa en las imágenes, en este caso la textura y el grado de correlación para utilizarla en el espacio tridimensional; una vez encontradas las zonas ricas en textura, el algoritmo se encarga de *propagar la precisión*, es decir, mediante el proceso de reclutamiento, una forma de comunicación compleja que utilizan las abejas para transmitir información y necesidades, van agrupándose alrededor de las zonas más significativas de la escena tridimensional conformando una reconstrucción dispersa mejorada o hasta quasi-densa. Esta comunicación y cooperación entre los individuos es la que permite la reconstrucción de zonas compactas de la escena, lo que sin duda es más significativa que una reconstrucción densa pero con errores, y que una reconstrucción dispersa no inteligente.

El proceso de búsqueda de la colmena de abejas es un proceso flexible y multi-propósito. Así debe ser también el algoritmo que se basa en este proceso, por eso, se plantea el siguiente trabajo futuro:

- *Optimización orientada al procesamiento en tiempo real.* Buscar una simplificación del algoritmo de manera que pueda obtener resultados aceptables en períodos de tiempo muy pequeños. Esto con el fin de utilizar el algoritmo para aplicaciones en tiempo real, como la navegación de un robot o la detección de obstáculos; tareas que no requieren de una reconstrucción tridimensional muy precisa.
- *Estudiar su aplicación en otros problemas.* Actualmente se ha considerado la posible aplicación en problemas de reconocimiento y clasificación de objetos, tanto en el espacio tridimensional, como en el plano de la imagen. Esto podría verse como el paso natural, después de llevar a cabo la búsqueda y descubrimiento de un objeto, lo que sigue es identificarlo. Otra tarea en que podría ser aplicado el enfoque es en el seguimiento de objetos o personas (en inglés tracking) mediante video.
- *Paralelización.* El algoritmo actual es una simplificación del proceso de búsqueda natural, el cual es paralelo y asíncrono. Sería interesante paralelizar el algoritmo, lo que lo haría más fiel al proceso natural y más rápido. Esto podría ser una solución para el primer punto .
- *Comparación con su antecesor.* El algoritmo basado en el proceso de búsqueda de las abejas recoge algunas características del algoritmo de las moscas (Boumaza y Louchet (2001)). Sería interesante la comparación para hacer evidente las diferencias y las semejanzas entre estos dos métodos.

- *Múltiples imágenes.* Estudiar la factibilidad de aplicar el enfoque a una reconstrucción con múltiples imágenes, o bien su aplicación en el problema de la siguiente mejor vista (en inglés *Next Best View*).
- *Aplicar Estrategias de evolución autoadaptable.* Como se vió en el capítulo V, existen estrategias de evolución cuyos operadores de reproducción evolucionan junto con la población. Se espera que al aplicar esta característica llegue más rapido a una solución óptima.

Bibliografía

- Beekman, M., D. J. T. Sumpter, N. Seraphides, y F. Ratnieks 2004. “Comparing foraging behavior of small and large honey-bee colonies by decoding waggle dances made by foragers”. *Functional Ecology*, 18:829-835 p.
- Boumaza, A. y J. Louchet 2001. “Dynamic flies: Using real-time parisian evolution in robotics”. *Applications of Evolutionary Computing. Lecture Notes in Computer Science*, Springer-Verlag, 2037:288-297 p.
- Criminisi, A., I. Reid, y A. Zisserman 2000. “Single view metrology”. *International Journal of Computer Vision*, 40(2):123-148 p.
- Crist, E. 2004. “Can an insect speak? the case of the honeybee dance language”. *Social Studies of Science.*, 34(1):7-43 p.
- Darwin, C. 1859. “On the origin of species by means of natural selection”. Murray, London, primera edition. 459 pp.
- Davis, J., R. Ramamoorthi, y S. Rusinkiewicz 2003. “Spacetime stereo: a unifying framework for depth from triangulation”. En: “Computer Vision and Pattern Recognition, CVPR’2003”, volume 2. IEEE. Madison, Wisconsin, 359-366 p.
- Deb, K. 2001. “Multi-objective optimization using evolutionary algorithms”. John Wiley & sons, Baffins Lane, Chichester, West Sussex, PO19 1UD, England, first edition. 497 pp.
- Fielding, G. y M. Kam 2000. “Wiegthed matchings for dense stereo correspondence”. *Journal of the Pattern Recognition Society*, 33:1511-1524 p.

- Forsyth, D. y J. Ponce 2002. "Computer vision: A modern approach". Prentice Hall, first edition. 693 pp.
- Goldberg, D. E. 1989. "Genetic algorithms for search, optimization, and machine learning.". Addison-Wesley, Reading, MA. USA. 372 pp.
- Gonzalez, R. y R. Woods 2002. "Digital image processing". Prentice Hall, Inc., Saddle River, New Jersey, second edition. 793 pp.
- Gruen, A. 1985. "Adaptive least squares correlation: A powerful image matching technique". South African Journal of Photogrammetry, Remote Sensing and Cartography, 14(3):175-187 p.
- Hans-Georg 2001. "The theory of evolution strategies". Springer-Verlag, Berlin. 380 pp.
- Harris, C. J. y M. Stephens 1988. "A combined corner and edge detector". En: "4th Alvey Vision Conference". 147-151 p.
- Hartley, R. y P. Sturm 1997. "Triangulation". Computer Vision and Image Understanding, 68(2):146-157 p.
- Hartley, R. I. y A. Zisserman 2004. "Multiple view geometry in computer vision". Cambridge University Press, second edition. 655 pp.
- H.C., L.-H. 1981. "A computer algorithm for reconstructing a scene from two projections". Nature, 293(September):133-135 p.
- Hernández, B. 2002. "Un nuevo detector sub-píxel paramétrico de esquinas múltiples, empleando técnicas de optimización tradicionales y cómputo evolutivo". Tesis de Maestría, Centro de Investigación Científica y de Educación Superior de Ensenada, Ensenada, Baja California. 254 pp.

- Holland, J. 1975. "Adaptation in natural and artificial systems". Universidad of michigan press, Michigan, first edition. 211 pp.
- Kass, M., A. Witkin, y D. Terzopoulos 1987. "Snakes: active contour models". En: "First International Conference on Computer Vision", volume 1. 259-268 p.
- Koch, R. y L. V. Good 1998. "Cumuli, panorama and vanguard project overview". Lecture Notes in Computer Science, Springer-Verlag, 1506:1-13 p.
- Koza, J. R. 1992. "Genetic programing: On the programming of computers by means of natural selection". MIT press, Cambridge, MA. USA. 840 pp.
- Marchand, E. y F. Chaumette 1999. "An autonomous active vision system for complete and accurate 3d scene reconstruction". International Journal of Computer Vision, 32(3):171-194 p.
- Marr, D. 1982. "Vision". W. H. Freeman and Company, New York. 360 pp.
- Marsaglia, G. y A. Zaman 1991. "A new class of random number generators". Annals of Applied Probability, 1(3):462 pp.
- McInerney y Terzopoulos 1995. "Topologically adaptable snakes". En: "International Conference of Computer Vision, ICCV'95". Cambridge, MA. USA. 840-845 p.
- Michalewicz, Z. 1982. "Genetic algorithms + data structures = evolution programs". W. H. Freeman and Company, New York. 385 pp.
- Olague, G., F. Fernández, C. Pérez, y E. Luton 2004. "The infection algorithm: An artificial epidemic approach to dense stereo matching". Parallel Problem Solving from Nature VIII. Lecture Notes in Computer Science, 3242:622-632 p.

- Pérez, C. B. 2004. "Apareamiento denso de un par estereocópico usando un enfoque de vida artificial". Tesis de Maestría, Centro de Investigación Científica y de Educación Superior de Ensenada, Ensenada, Baja California. 163 pp.
- Pollefeys, M., R. Koch, V. Maarten, y L. V. Gool 1998. "Metric 3d surface reconstruction from uncalibrated image sequences". Lecture Notes in Computer Science, Springer-Verlag, 1506:139-154 p.
- Press, W., S. Teukolsky, W. Vetterling, y B. Flannery 1994. "Numerical recipes in c". Cambridge University Press, 40 West 20th Street, New York, second edition. 994 pp.
- Ramirez, R. 2000. "Síntesis de imágenes a partir de fotografías". Tesis de Maestría, Centro de Investigación Científica y de Educación Superior de Ensenada, Ensenada, Baja California. 218 pp.
- Reklaitis, G. V. 1983. "Engineering optimization, methods and applications". Wiley, New York. 704 pp.
- Scharstein, D. y R. Szelinski 2002. "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms". International Journal of Computer Vision, 47(1/2/3):7-42 p.
- Schwefel, H. P. 1995. "Evolution and optimum seeking". Wiley, New York, first edition. 444 pp.
- Sun, C. 2002. "Fast stereo matching using rectangular subregioning and 3d, maximum-surface techniques". International Journal of Computer Vision, 47(1/2/3):99-117 p.
- Sun, J., N.-N. Zheng, y H.-Y. Shum 2003. "Stereo matching using belief propagation". IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI), 25(7):787-800 p.

- Thakoor, S., J. Morookian, J. Chahl, B. Hine, y S. Zornetzer 2004. “Bees: Exploring mars with bioinspired technologies”. IEEE Computer, 37(9):38-47 p.
- von Frisch, K. 1967. “The dance language and orientation of bees”. Harvard University Press, Cambridge, MA. USA. 566 pp.
- Wright, A. 1991. “Genetic algorithms for real parameter optimization”. En: Rawlins, G. J. E., editor, “Foundations of Genetic Algorithms (FOGA)”, volume 1. Morgan Kaufmann. 205-218 p.