

ADECUACIÓN DE TÉCNICAS DE DESCRIPCIÓN VISUAL UTILIZANDO INFORMACIÓN 3D Y SU APLICACIÓN EN ROBÓTICA

Abel Pacheco Ortega

Director de tesis: **DR. JESUS SAVAGE CARMONA**
DR. WALTERIO W. MAYOL CUEVAS

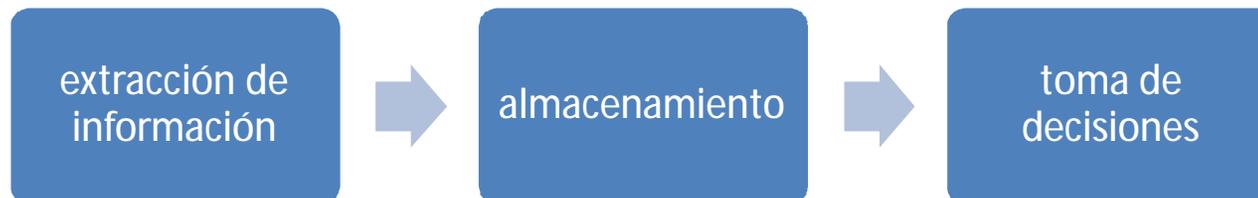
Visión

- **Visión humana**

Es una tarea de procesamiento de información, que describe con las imágenes lo que está presente en el mundo real y dónde se encuentra

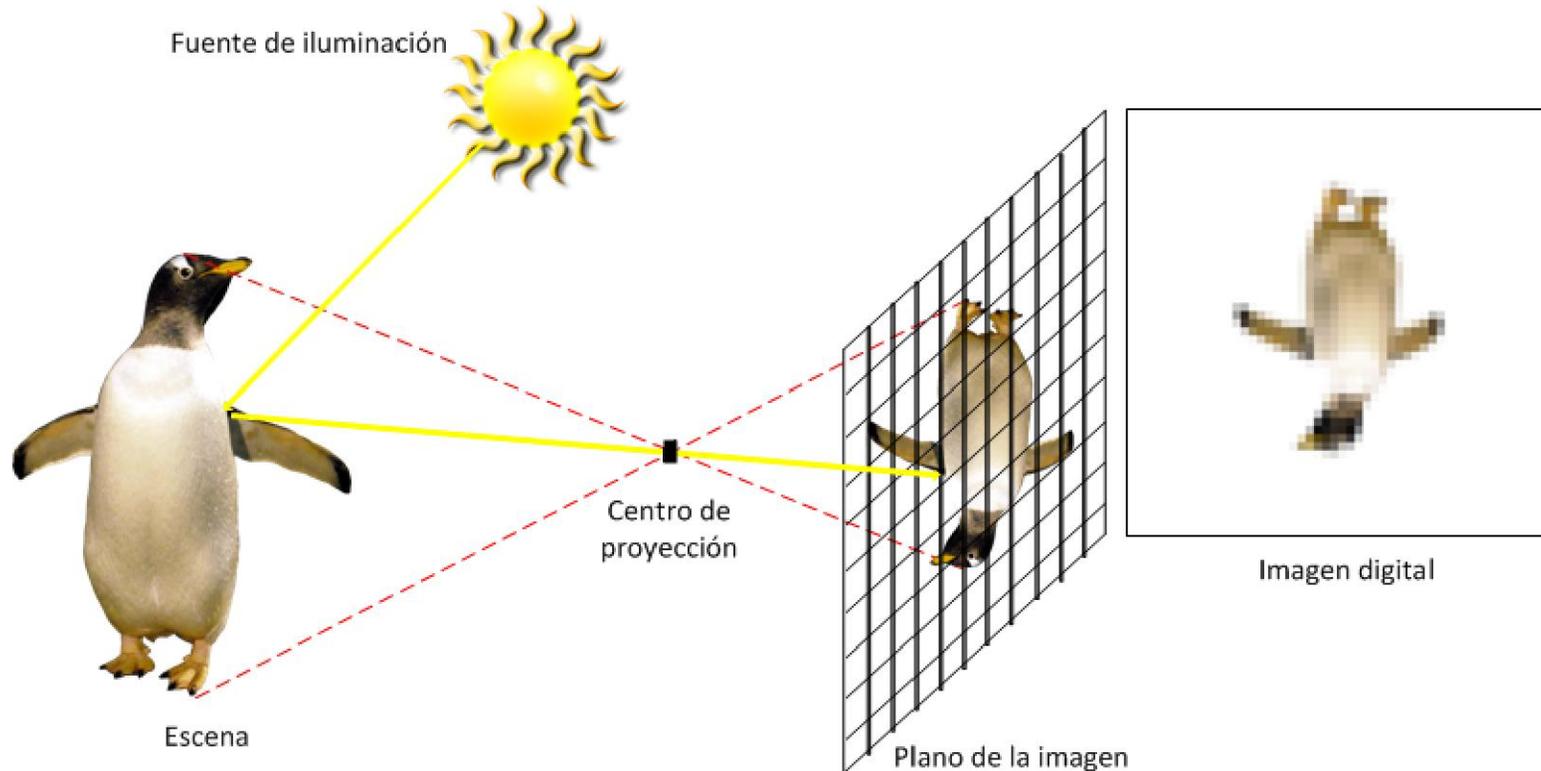
- **Visión por computadora**

Disciplina cuyo objetivo es realizar decisiones a partir de características del mundo real extraídas a través de imágenes digitales del mismo.



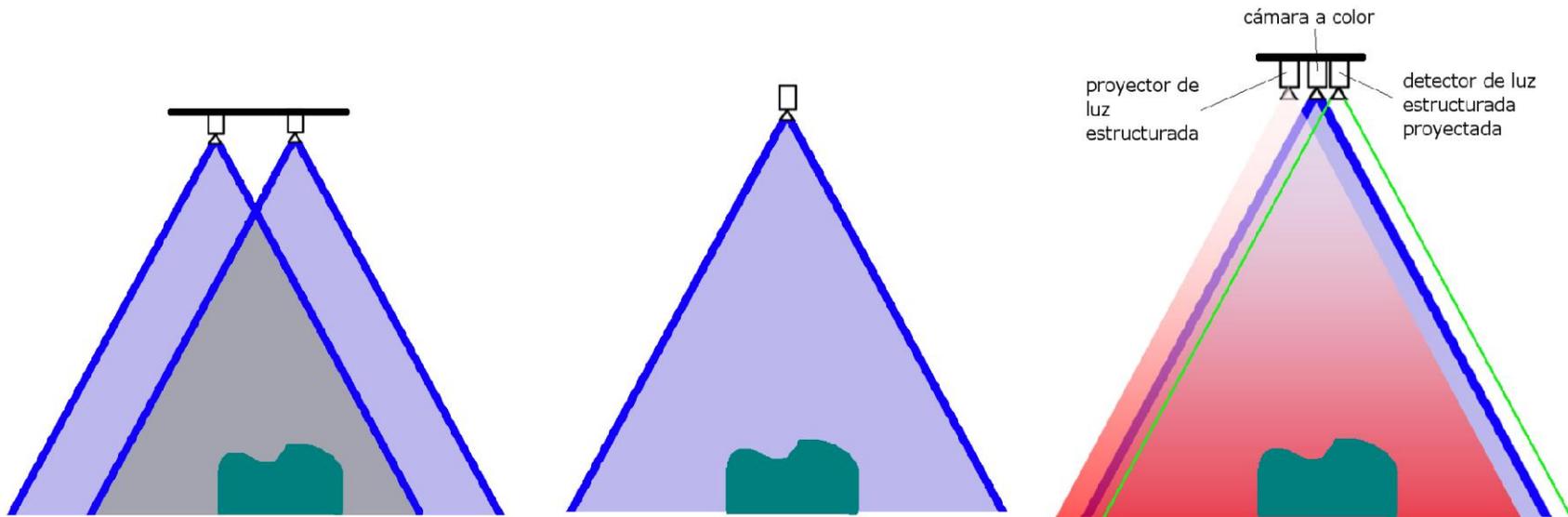
Nuevas tecnologías

Imágenes RGB



Nuevas tecnologías

Imágenes RGB-D



Estéreo



ToF



Luz estructurada



Objetivo

Generar una aplicación que realice la extracción de características visuales robustas, con una buena relación tiempo de ejecución-resultados, dicha metodología será utilizada para proponer un sistema que realice el mapeo y localización visual del entorno en el cual se desplaza la cámara.

METODOLOGÍA DE EXTRACCIÓN DE CARACTERÍSTICAS

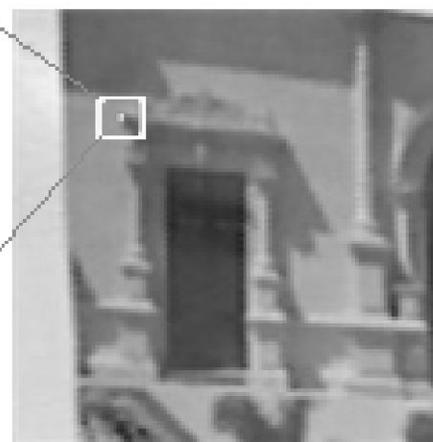
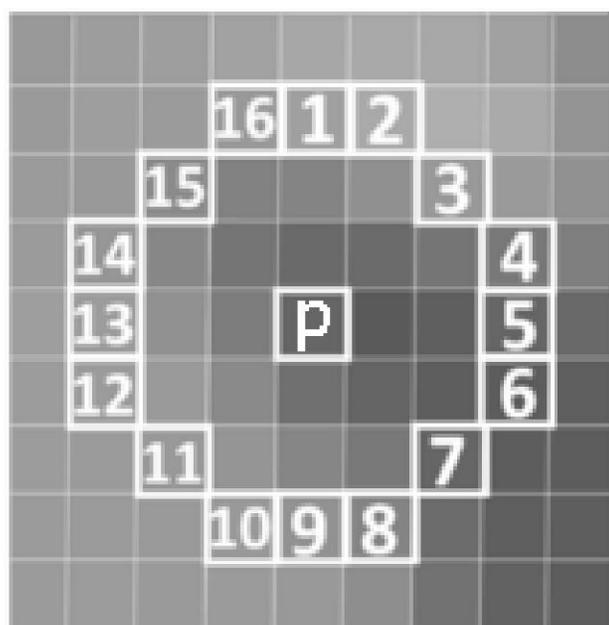


Metodología de extracción de características

FAST+3D

- Features from Accelerated Segmented Test (FAST)

Existen un conjunto de n pixeles contiguos en el perímetro de un círculo de radio 3 alrededor del mismo que tengan una intensidad superior a I_p+t o inferior a I_p-t



I_p valor de intensidad del punto p
 t es un umbral determinado
 n valores entre 9 y 12

Metodología de extracción de características FAST+3D

- Features from Accelerated Segmented Test (FAST)



Metodología de extracción de características

FAST+3D

- Features from Accelerated Segmented Test (FAST)
- Filtrado por vecindario inestable

¿Tamaño de vecindario?

Metodología de extracción de características

FAST+3D

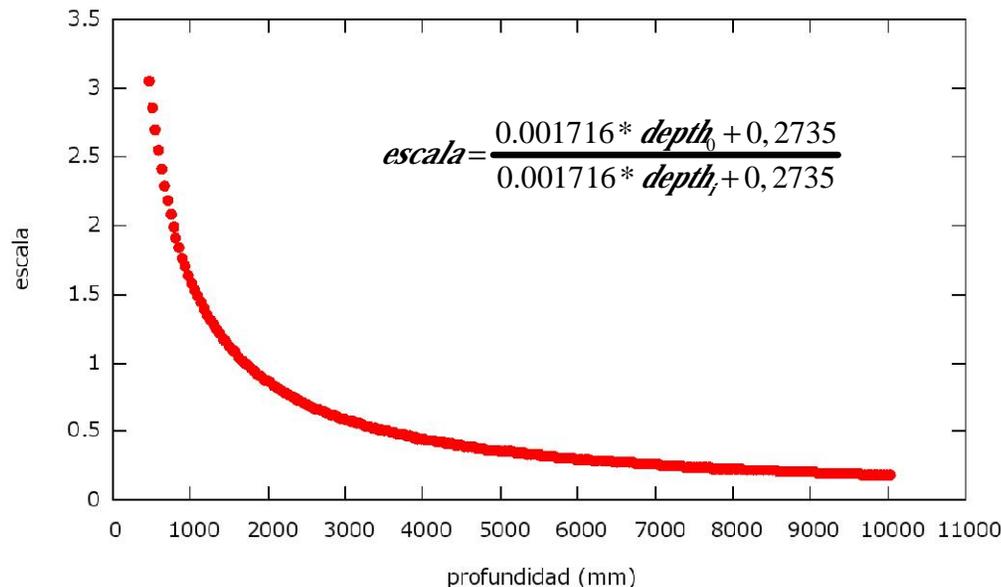
- Features from Accelerated Segmented Test (FAST)
- Filtrado por vecindario inestable

Objeto de dimensiones conocidas como entrada, colocado a diferentes profundidades desde la cámara

$$f=0.001716(\text{depth})+0.2735$$

f (mm/pixel) representa la dimensión del objeto a una profundidad depth (mm)

Función escala-profundidad



$depth_0$ profundidad referencia

$depth_i$ profundidad a la que se busca el factor de escala

Metodología de extracción de características

FAST+3D

- Features from Accelerated Segmented Test (FAST)
- Filtrado por vecindario inestable

La profundidad y tamaño de vecindario referencia escogida es 1700 mm y 25 pixeles respectivamente.

$$escala = \frac{0.001716 * 1700 + 0,2735}{0.001716 * depth + 0,2735}$$

el tamaño de vecindario a cualquier distancia puede ser calculado con

$$tam. vecindario = escala * 25$$

Metodología de extracción de características

FAST+3D

- Features from Accelerated Segmented Test (FAST)
- Filtrado por vecindario inestable

$$escala = \frac{0.001716 * 1700 + 0,2735}{0.001716 * depth_i + 0,2735}$$

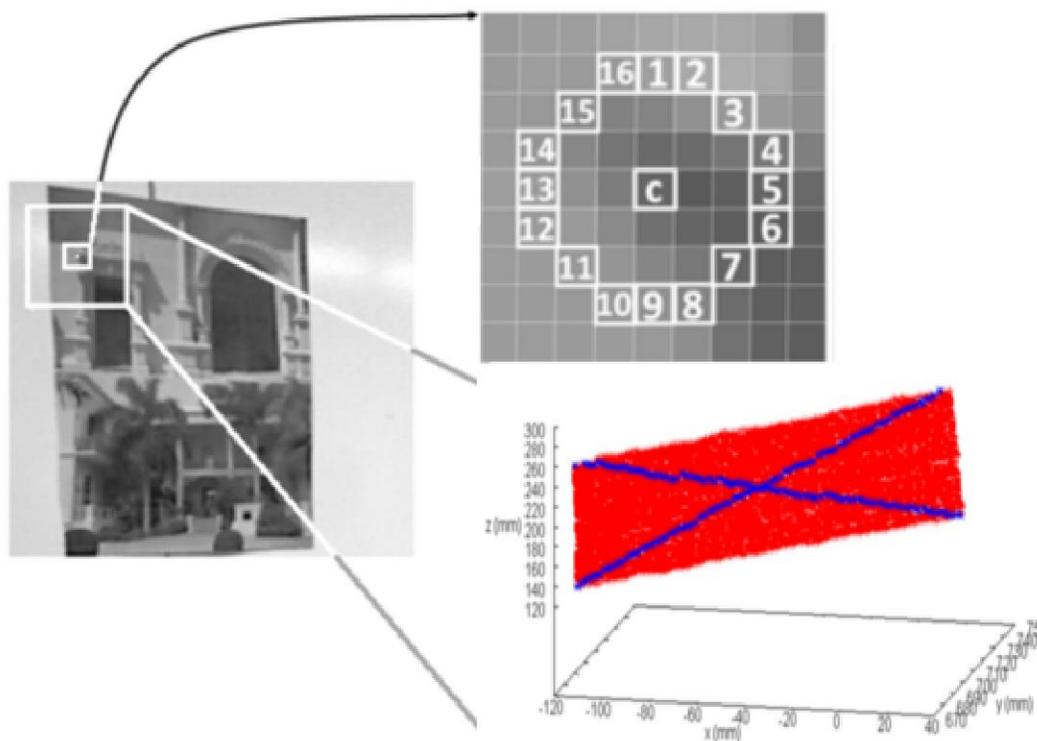
$$tam. vecindario = escala * 25$$

$$P_{FAST+3D} = \{ p \in P_{FAST} \mid gneig(p_i) = 1 \}$$

$$gneig(p_i) = \begin{cases} 0 & \text{si } diag(p_i) \geq \tau \\ 1 & \text{otros} \end{cases}$$

$$diag(p_i) = \frac{r}{2w_i + 1}$$

r es el número de píxeles en las diagonales del vecindario donde la diferencia de profundidades con el punto de interés (pixel central del vecindario) no es mayor que un umbral determinado



Metodología de extracción de características

FAST+3D

- Features from Accelerated Segmented Test (FAST)
- Filtrado por vecindario inestable

$$escala = \frac{0.001716 * 1700 + 0,2735}{0.001716 * depth_i + 0,2735}$$

$$tam. vecindario = escala * 25$$

$$P_{FAST+3D} = \{ p \in P_{FAST} \mid gneig(p_i) = 1 \}$$

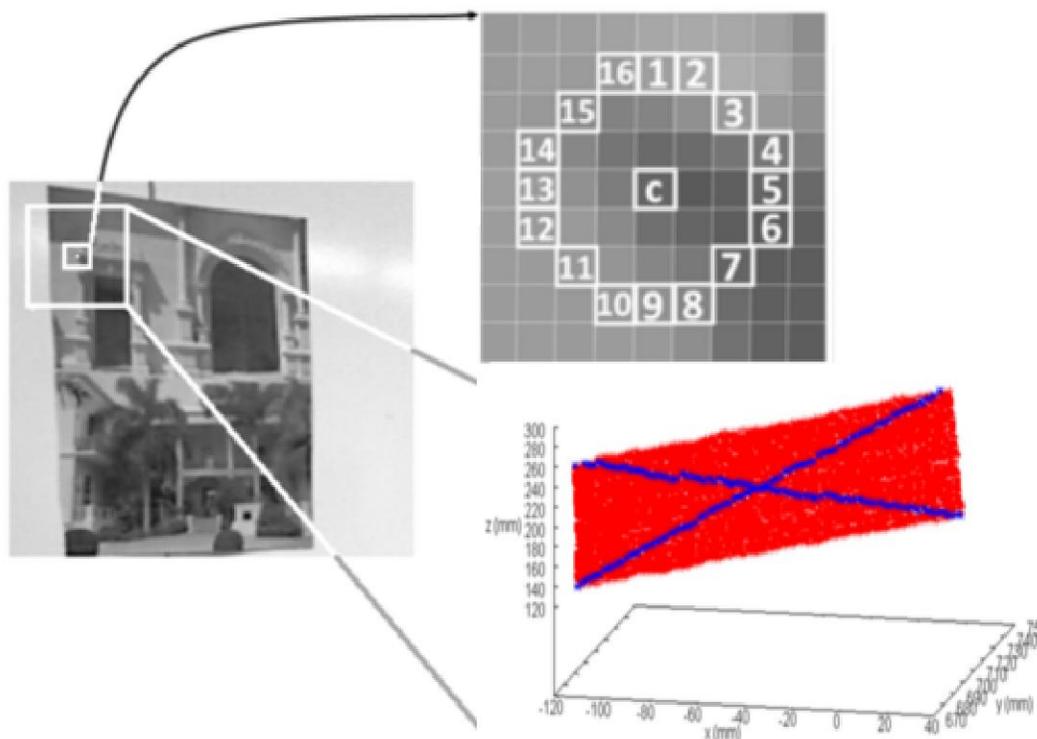
$$gneig(p_i) = \begin{cases} 0 & \text{si } diag(p_i) \geq \tau \\ 1 & \text{otros} \end{cases}$$

$$diag(p_i) = \frac{r}{2w_i + 1}$$

ρ

τ

es un umbral por medio del cual se determina cual es el porcentaje máximo de puntos en las diagonales que serán tolerados para determinar que un punto de interés sea eliminado del conjunto final.



Metodología de extracción de características

FAST+3D

- Features from Accelerated Segmented Test (FAST)
- Filtrado por vecindario inestable

$$escala = \frac{0.001716 * 1700 + 0,2735}{0.001716 * depth_i + 0,2735}$$

$$tam. vecindario = escala * 25$$

$$P_{FAST+3D} = \{ p \in P_{FAST} \mid gneig(p_i) = 1 \}$$

$$gneig(p_i) = \begin{cases} 0 & \text{si } diag(p_i) \geq \tau \\ 1 & \text{otros} \end{cases}$$

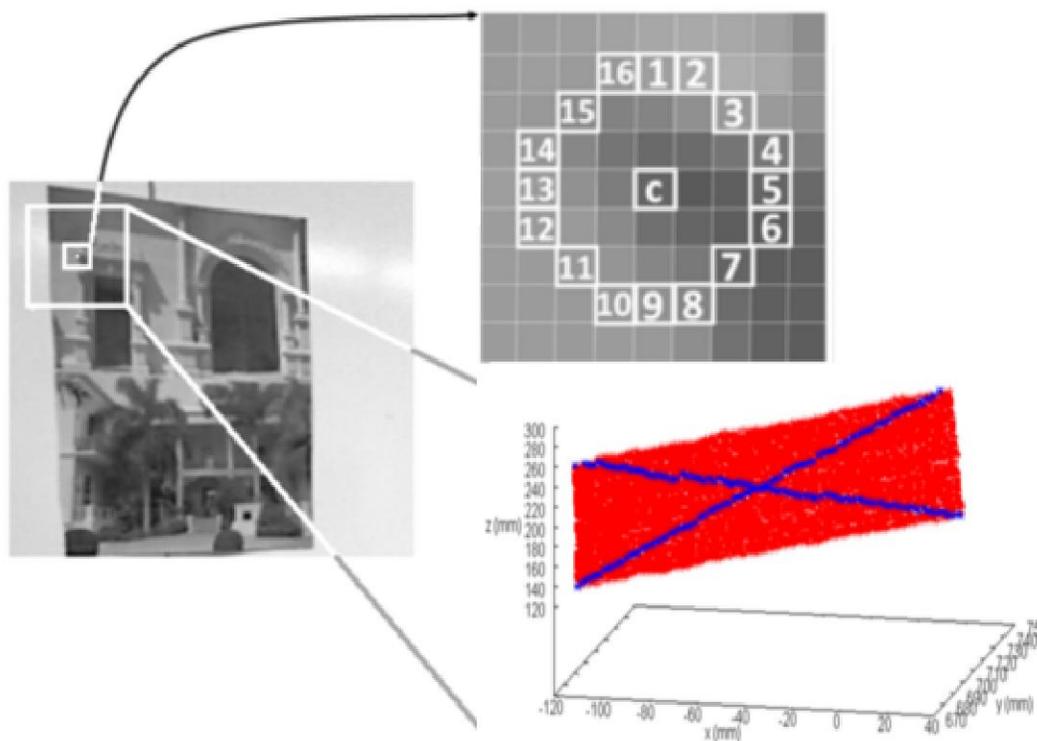
$$diag(p_i) = \frac{r}{2w_i + 1}$$

$$\rho = 200 \text{ mm}$$

$$\tau = 5\%$$

$$n = 9$$

$$t = 40$$



Metodología de extracción de características

FAST+3D

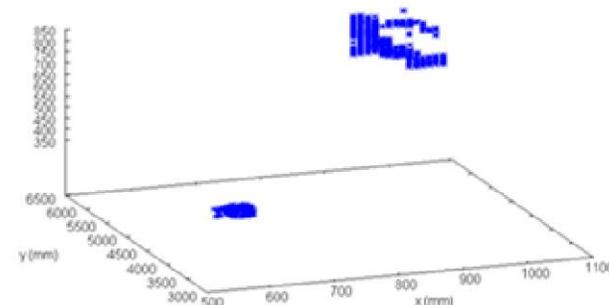
- Features from Accelerated Segmented Test (FAST)
- Filtrado por vecindario inestable



FAST point neighborhood



z (mm)



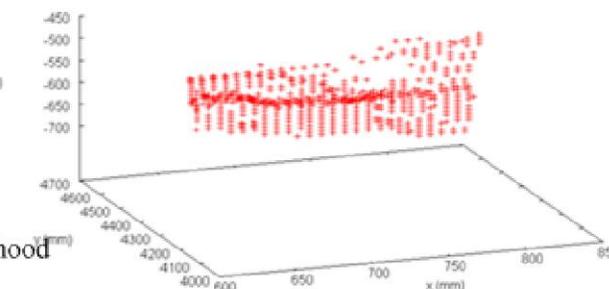
SAMPLE

3D information

FAST+3D point neighborhood



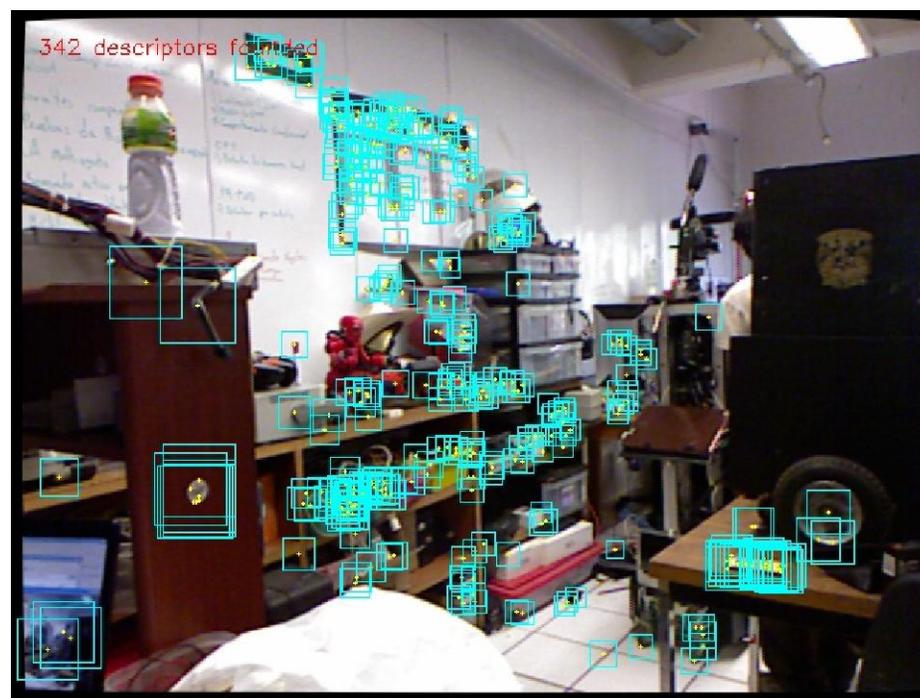
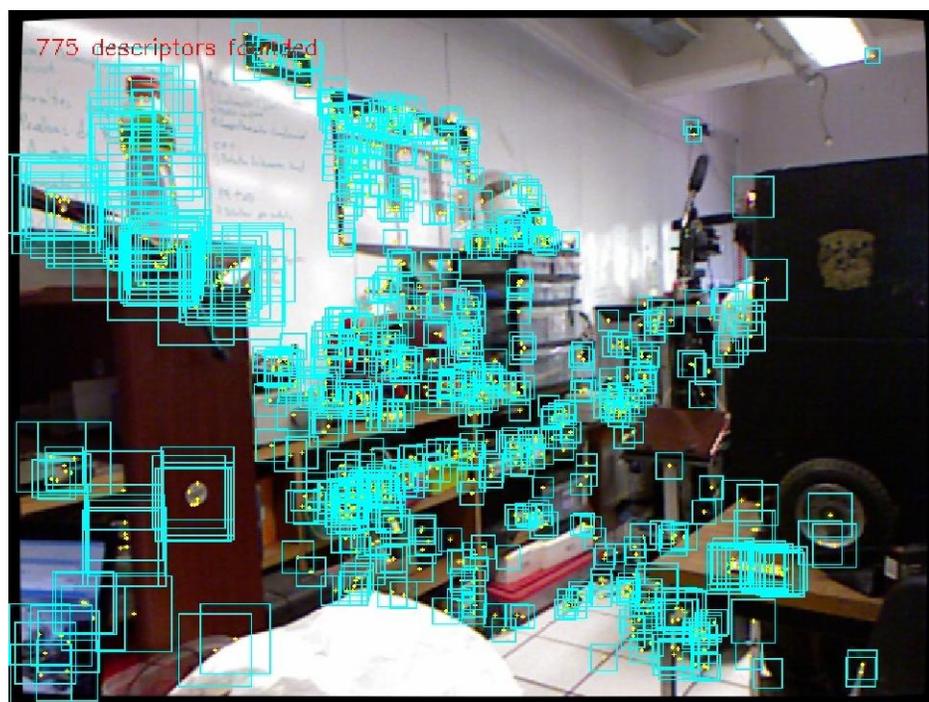
z (mm)



Metodología de extracción de características

FAST+3D

- Features from Accelerated Segmented Test (FAST)
- Filtrado por vecindario inestable



Metodología de extracción de características

Cambios de perspectiva

Plano en el espacio

$$Ax + By + Cz + D = 0$$

$$A = n_x$$

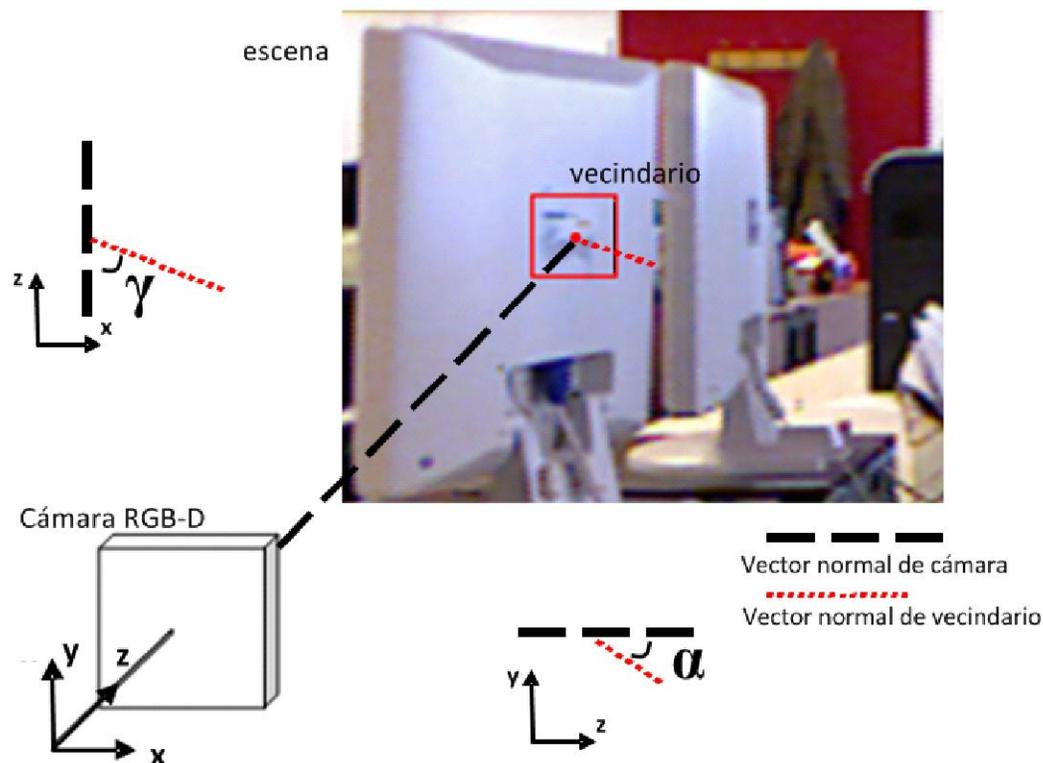
$$B = n_y$$

$$C = n_z$$

$$D = -(n_x x_0 + n_y y_0 + n_z z_0) = -n \cdot r_0$$

Random Sample Consensus (RANSAC)
para eliminar dichos elementos atípicos

Análisis de Componentes Principales



Metodología de extracción de características

Cambios de perspectiva (RANSAC)

PLANTEAMIENTO

- En análisis de datos en términos de modelos predefinidos

Estimación de parámetros. Computo de los parámetros libres del modelo seleccionado.

Clasico:

- Optimizan el ajuste de una descripción funcional (modelo) a todo los datos dados.

- **No** tienen mecanismos internos para detectar y rechazar errores graves .

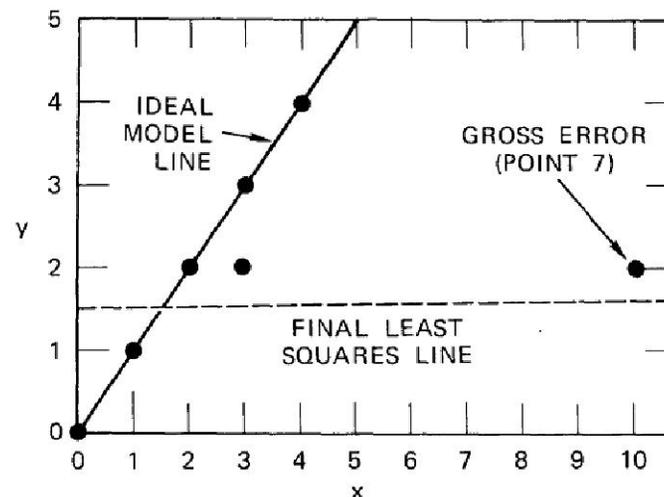
- Se basan en el supuesto de que la máxima desviación de cualquier dato del modelo asumido es una función directa del tamaño del conjunto de datos (**habrá siempre suficiente cantidad de datos para suavizar los efectos de desviaciones graves**)

RANSAC:

- Usa un conjunto inicial de datos tan pequeño como sea posible y lo hace cada vez más grande utilizando los datos que son consistentes con el mismo.

- Responde de manera correcta incluso si se tienen un número significativo de datos se presentan como graves.

POINT	x	y
1	0	0
2	1	1
3	2	2
4	3	2
5	3	3
6	4	4
7	10	2



Metodología de extracción de características

Cambios de perspectiva (Análisis de Componentes Principales)

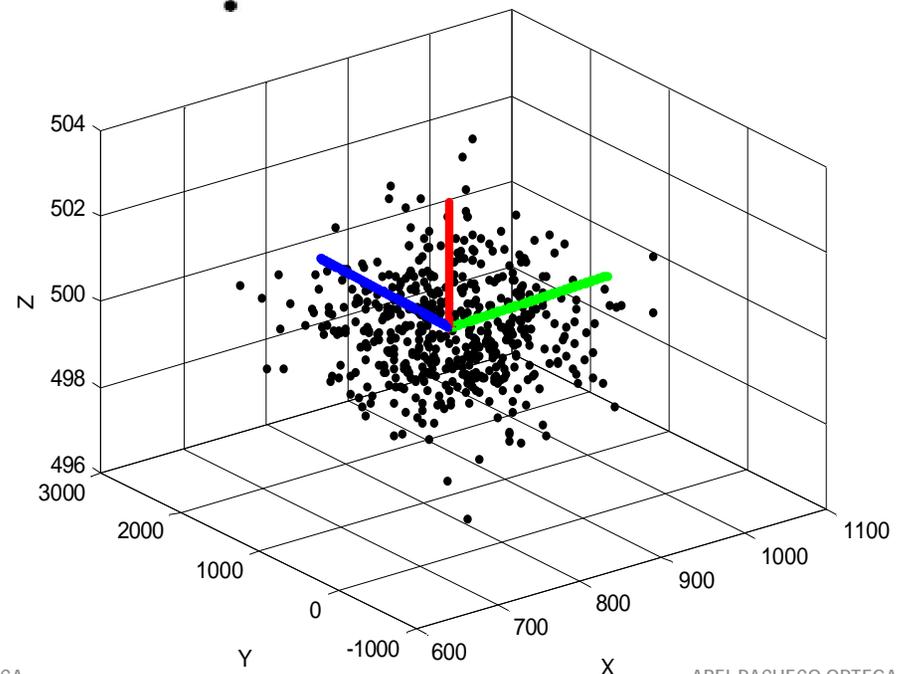
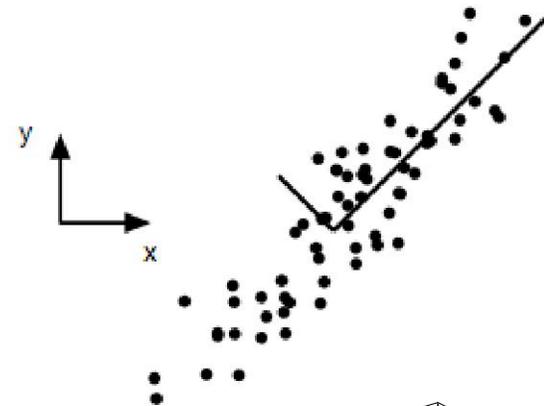
PLANTEAMIENTO

- Metodología utilizada para reducir la dimensionalidad de un problema
- Se basa en el hecho de: "Mayor información significa mayor variabilidad de datos o varianza"

Metodología

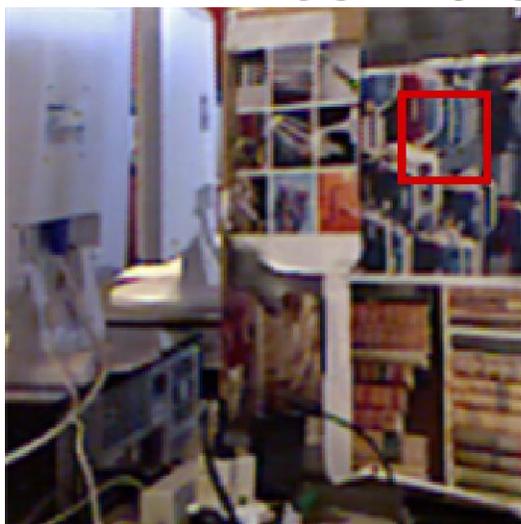
- Obtener la matriz de covarianza() del conjunto de puntos.
- Obtener los eigenvectores y eigenvalores de la matriz de covarianza

EL EIGENVECTOR ASOCIADO AL MENOR EIGENVALOR ES LA NORMAL DEL PLANO ASOCIADO

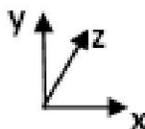


Metodología de extracción de características

Cambios de perspectiva



Rotaciones
eje y: 31.2°
eje x: -1.45°



Rotaciones
eje y: -81.4°
eje x: -1.76°



Metodología de extracción de características

Descripción de vecindarios normalizados

- Los vecindarios normalizados son redimensionados a 25x25 píxeles
- El descriptor generado por SIFT está basado en un análisis de gradientes y orientación de las regiones alrededor de los puntos de interés.

$$m(x, y) = \sqrt{(L(x+1, y) - L(x-1, y))^2 + (L(x, y+1) - L(x, y-1))^2}$$

$$\theta(x, y) = \tan^{-1} \frac{L(x, y+1) - L(x, y-1)}{L(x+1, y) - L(x-1, y)}$$

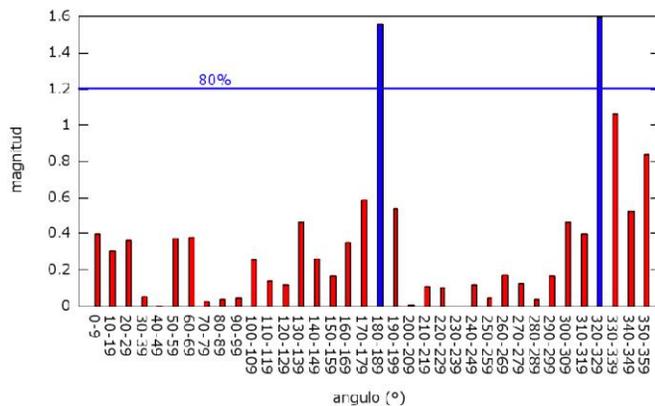
Metodología de extracción de características

Descripción de vecindarios normalizados

- Los vecindarios normalizados son redimensionados a 25x25 pixeles
- El descriptor generado por SIFT esta basado en un análisis de gradientes y orientación de la regiones alrededor de los puntos de interés.

$$m(x, y) = \sqrt{(L(x+1, y) - L(x-1, y))^2 + (L(x, y+1) - L(x, y-1))^2}$$

$$\theta(x, y) = \tan^{-1} \frac{L(x, y+1) - L(x, y-1)}{L(x+1, y) - L(x-1, y)}$$



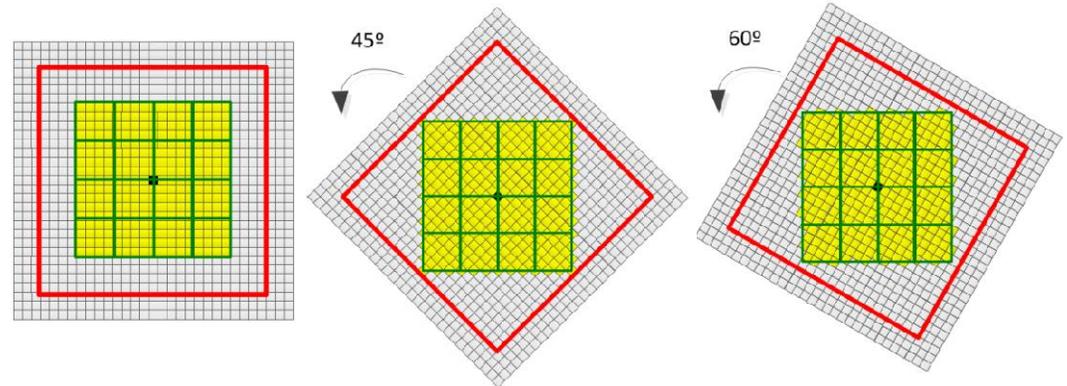
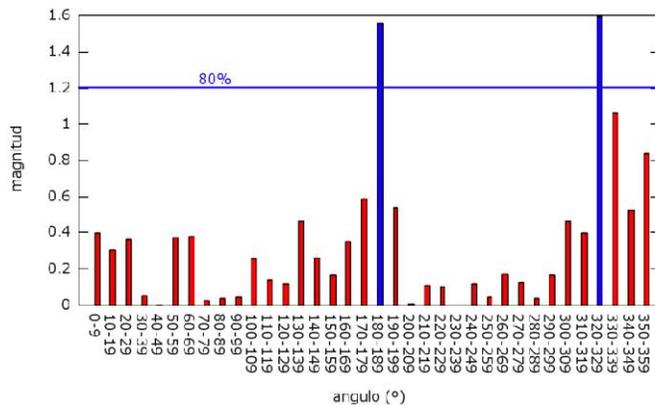
Metodología de extracción de características

Descripción de vecindarios normalizados

- Los vecindarios normalizados son redimensionados a 25x25 píxeles
- El descriptor generado por SIFT esta basado en un análisis de gradientes y orientación de la regiones alrededor de los puntos de interés.

$$m(x, y) = \sqrt{(L(x+1, y) - L(x-1, y))^2 + (L(x, y+1) - L(x, y-1))^2}$$

$$\theta(x, y) = \tan^{-1} \frac{L(x, y+1) - L(x, y-1)}{L(x+1, y) - L(x-1, y)}$$



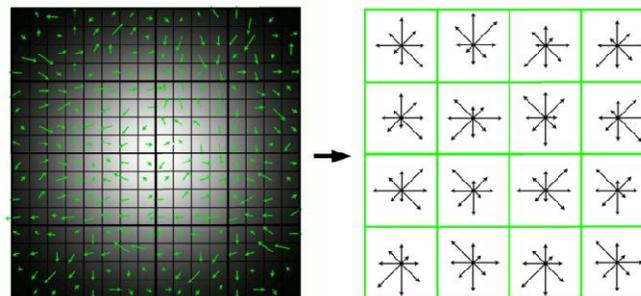
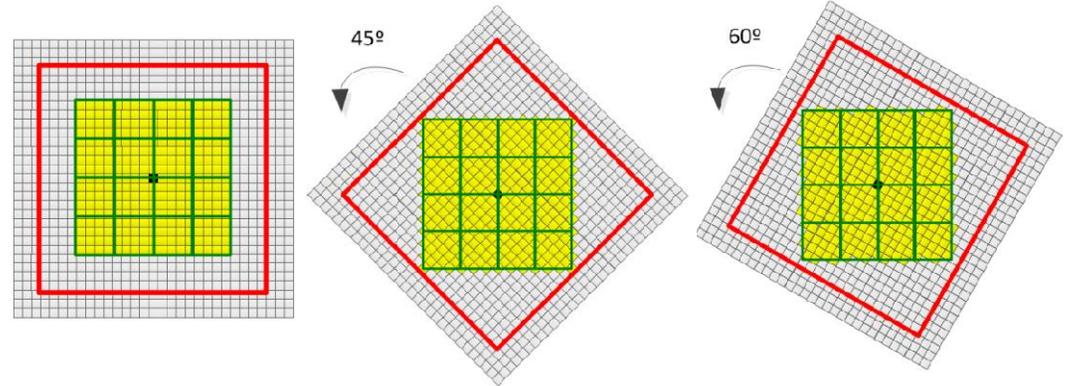
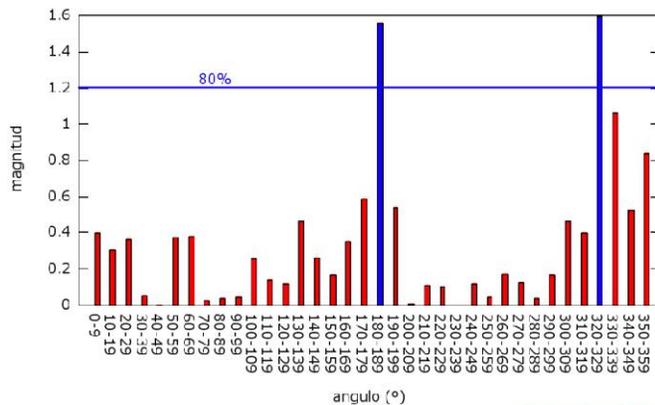
Metodología de extracción de características

Descripción de vecindarios normalizados

- Los vecindarios normalizados son redimensionados a 25x25 pixeles
- El descriptor generado por SIFT esta basado en un análisis de gradientes y orientación de la regiones alrededor de los puntos de interés.

$$m(x, y) = \sqrt{(L(x+1, y) - L(x-1, y))^2 + (L(x, y+1) - L(x, y-1))^2}$$

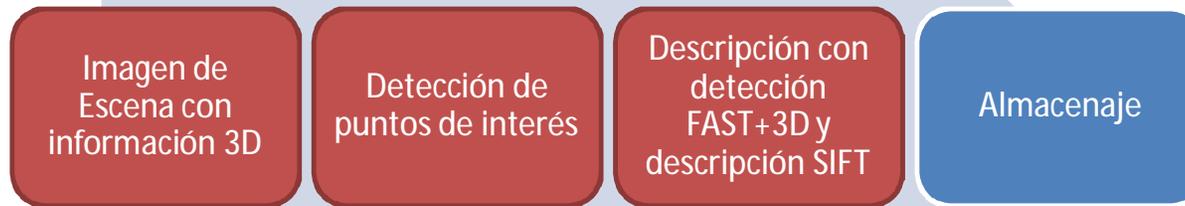
$$\theta(x, y) = \tan^{-1} \frac{L(x, y+1) - L(x, y-1)}{L(x+1, y) - L(x-1, y)}$$



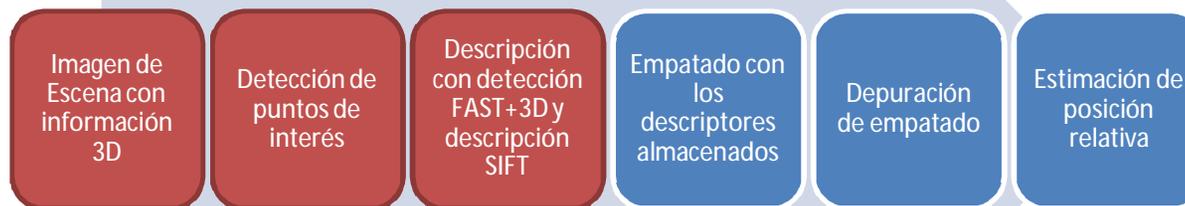
Dimensión de los descriptor
4x4x8 = 128

METODOLOGÍA DE LOCALIZACIÓN VISUAL

Caracterización



Empatado y localización



Metodología de localización visual

Caracterización (almacenaje)

- Arboles kd

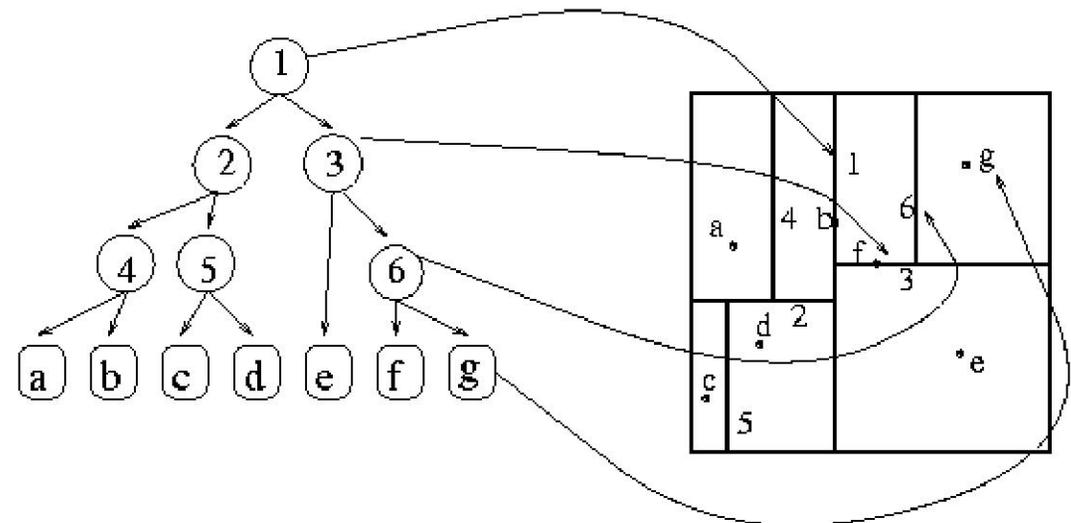
Dividen y organizan espacios multidimensionales con hiperplanos

- Información extra

Además del descriptor se almacenan la posición en la imagen y en el espacio

- Organización para búsquedas
Best Bin First

La división y almacenaje de el espacio se hace con hiperplanos que dividen de la manera mas eficiente.

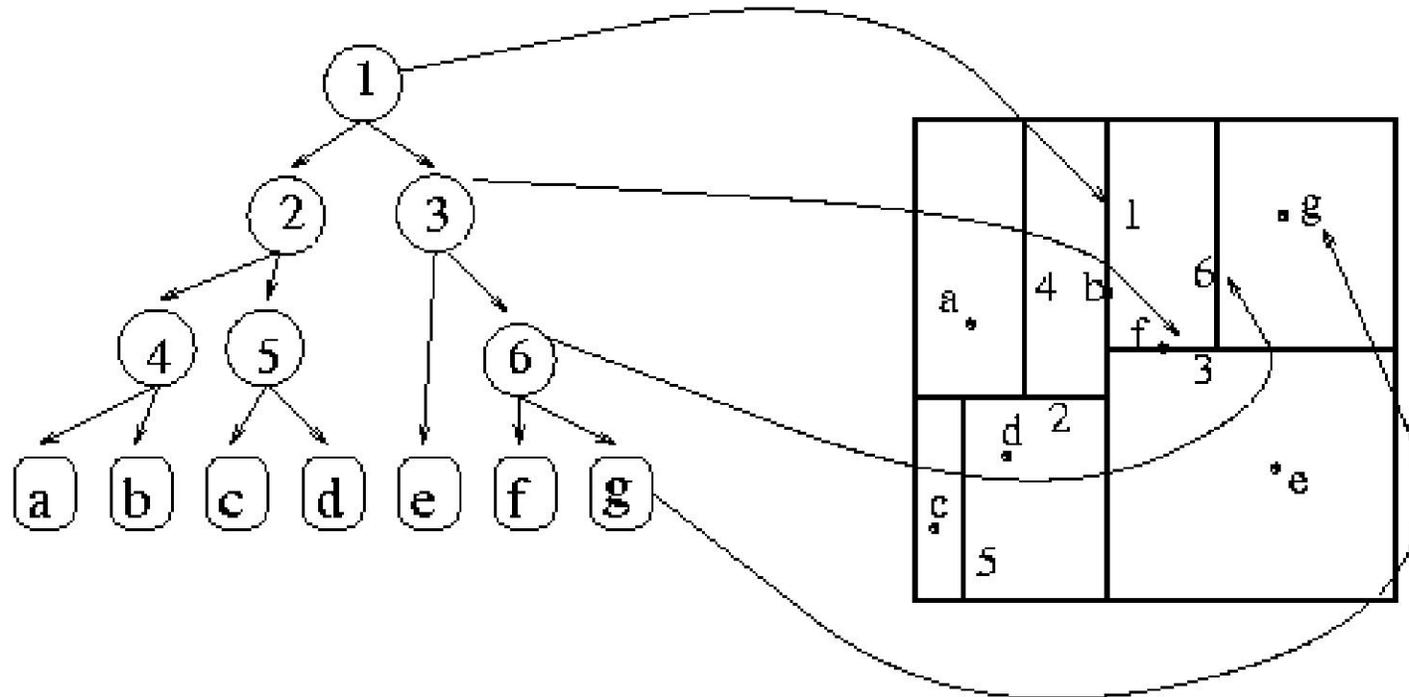


Metodología de localización visual

Empatado y localización

(empatado con descriptores almacenados)

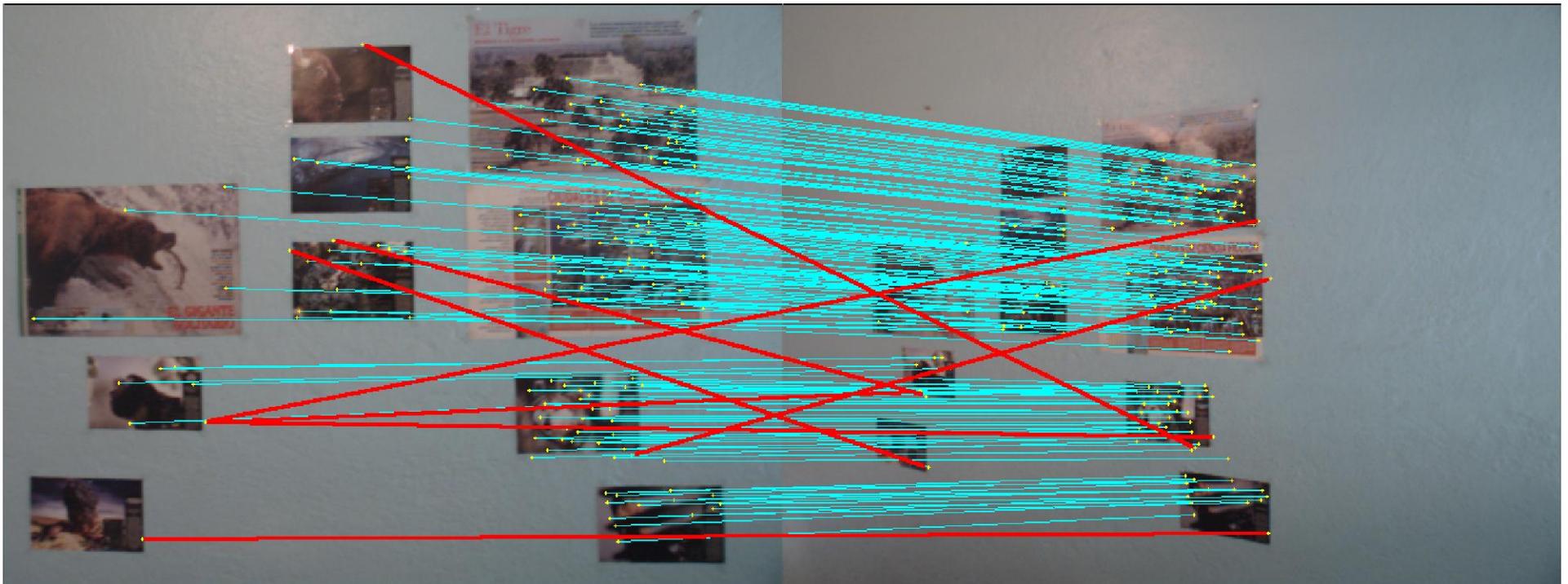
- Búsquedas *Best Bin First*
Con distancia euclidiana



Metodología de localización visual

Empatado y localización (empatado con descriptores almacenados)

- Búsquedas *Best Bin First*

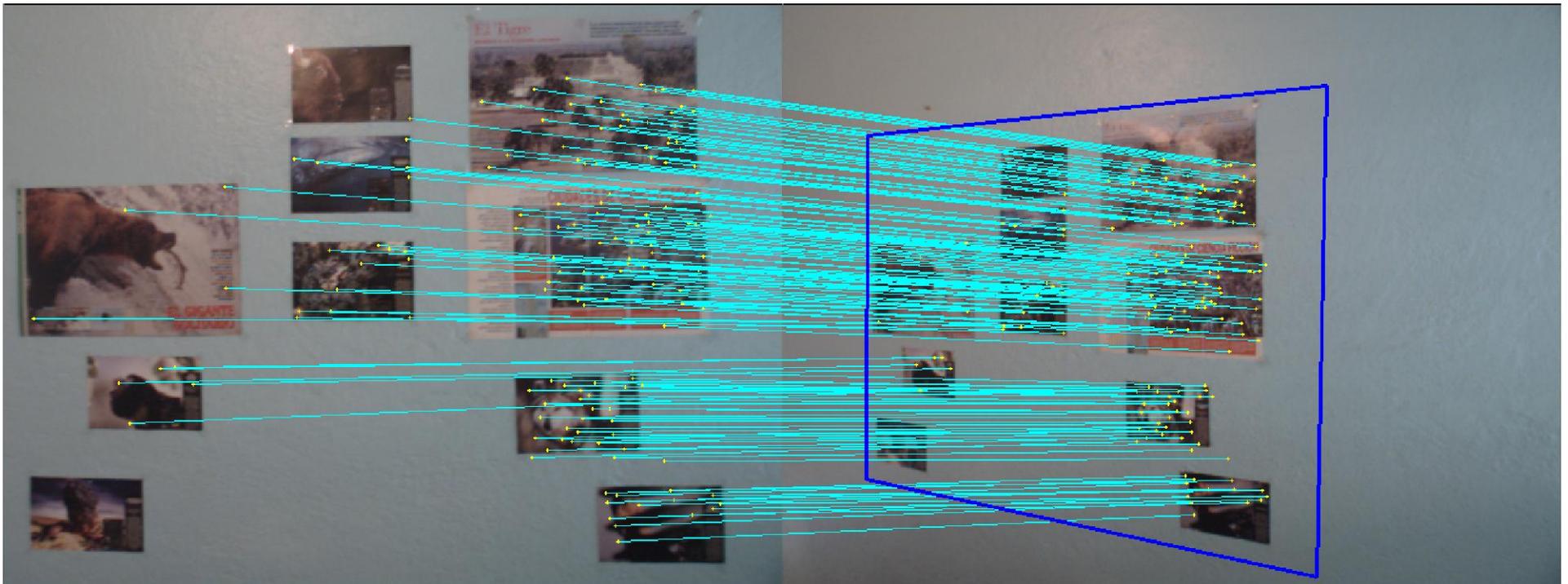


Metodología de localización visual Empatado y localización (depuración)

- RANSAC

El modelo utilizado es el de la transformación de homografía

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



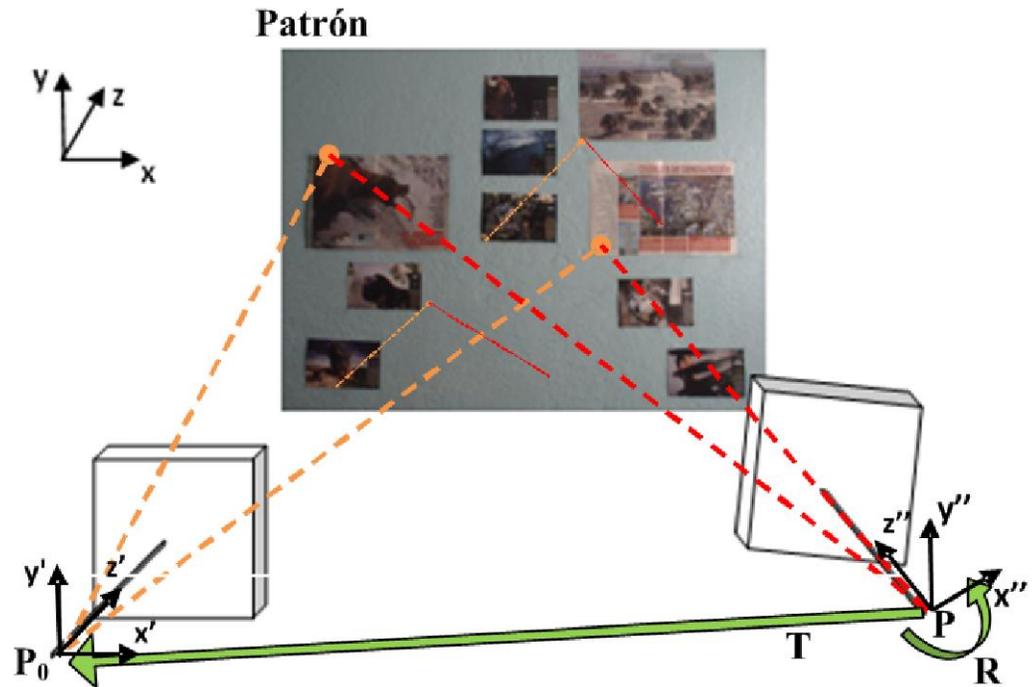
Metodología de localización visual

Empatado y localización (estimación de posición relativa)

- Orientación Absoluta
(fotogrametría)
Encontrar la relaciones de
translación T y rotación R entre dos
sistemas coordenados diferentes

$$p_{2,i} = \lambda R(p_{1,i}) + T$$

donde λ es un factor de escala

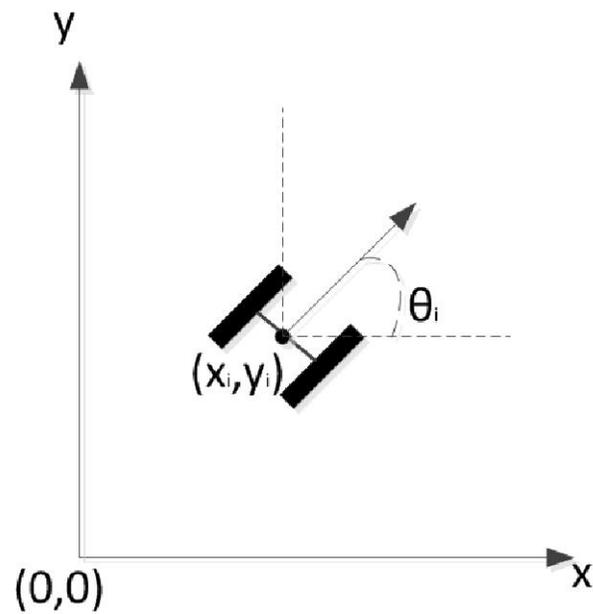


T es la relación de centroides

R es una matriz obtenida a partir sumas de productos y una análisis de eigenvectores aplicado a la misma

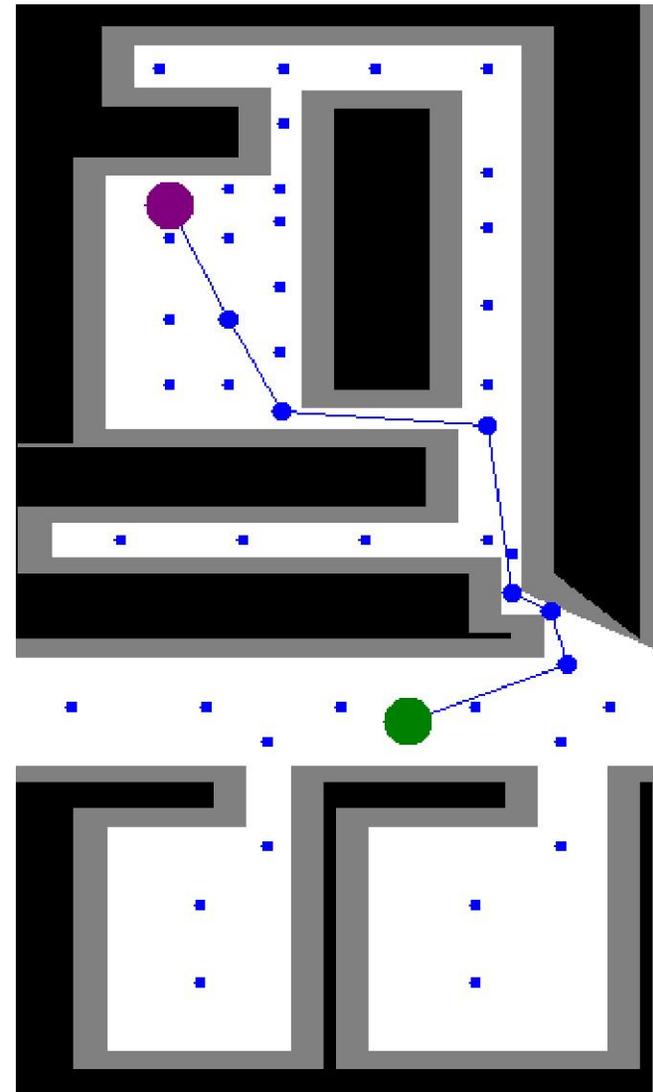
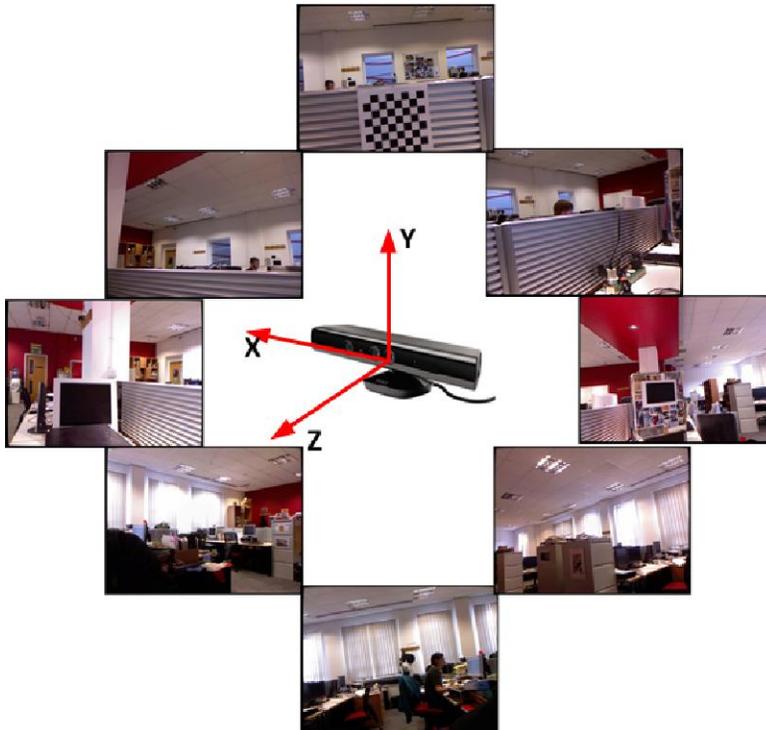
APLICACIÓN EN ROBÓTICA

- Navegación
 - Odometría



APLICACIÓN EN ROBÓTICA

- Navegación
 - Odometría
 - Mapas topológicos



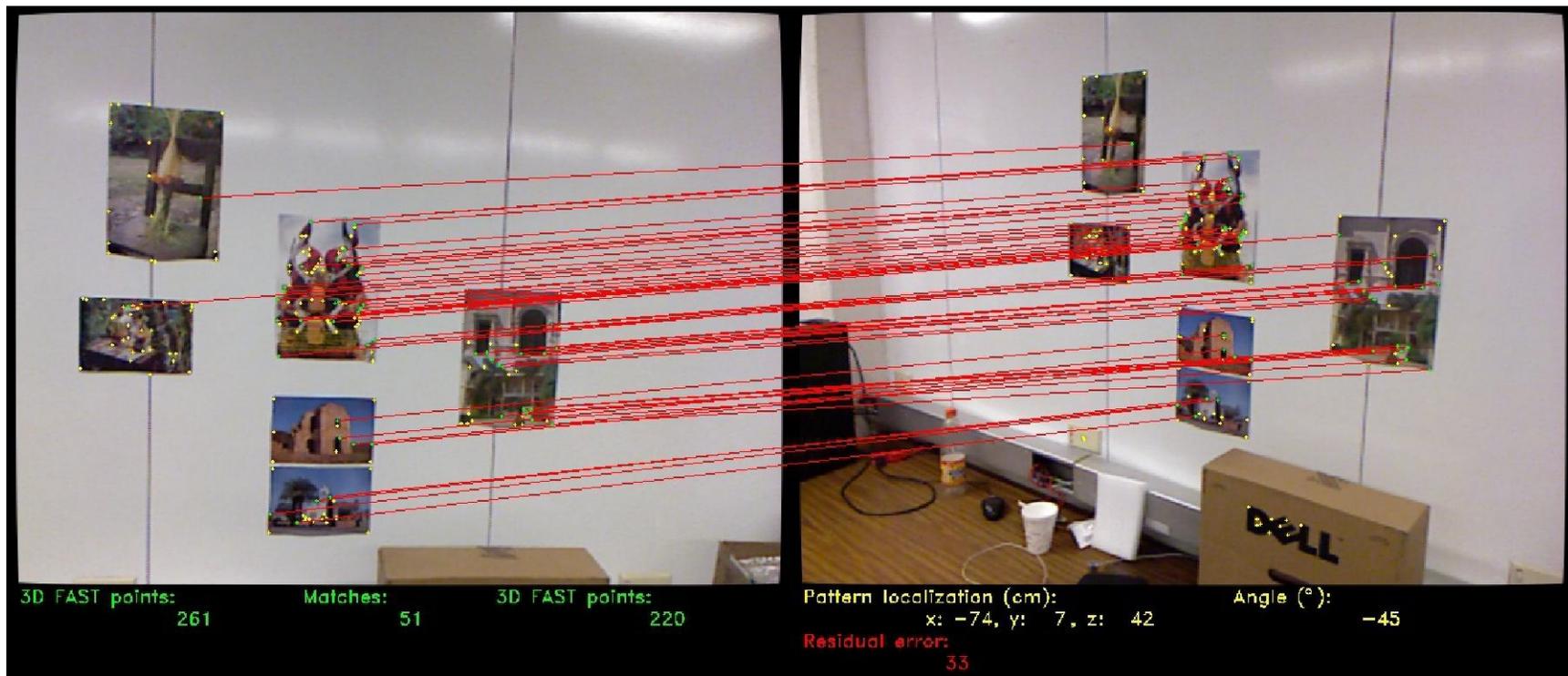
APLICACIÓN EN ROBÓTICA

- Navegación
 - Odometría
 - Mapas topológicos
- **Comunicación entre módulos**
 - oft_readMap
 - oft_createNode
 - oft_setNode
 - oft_saveMap
 - oft_getNode



PRUEBAS

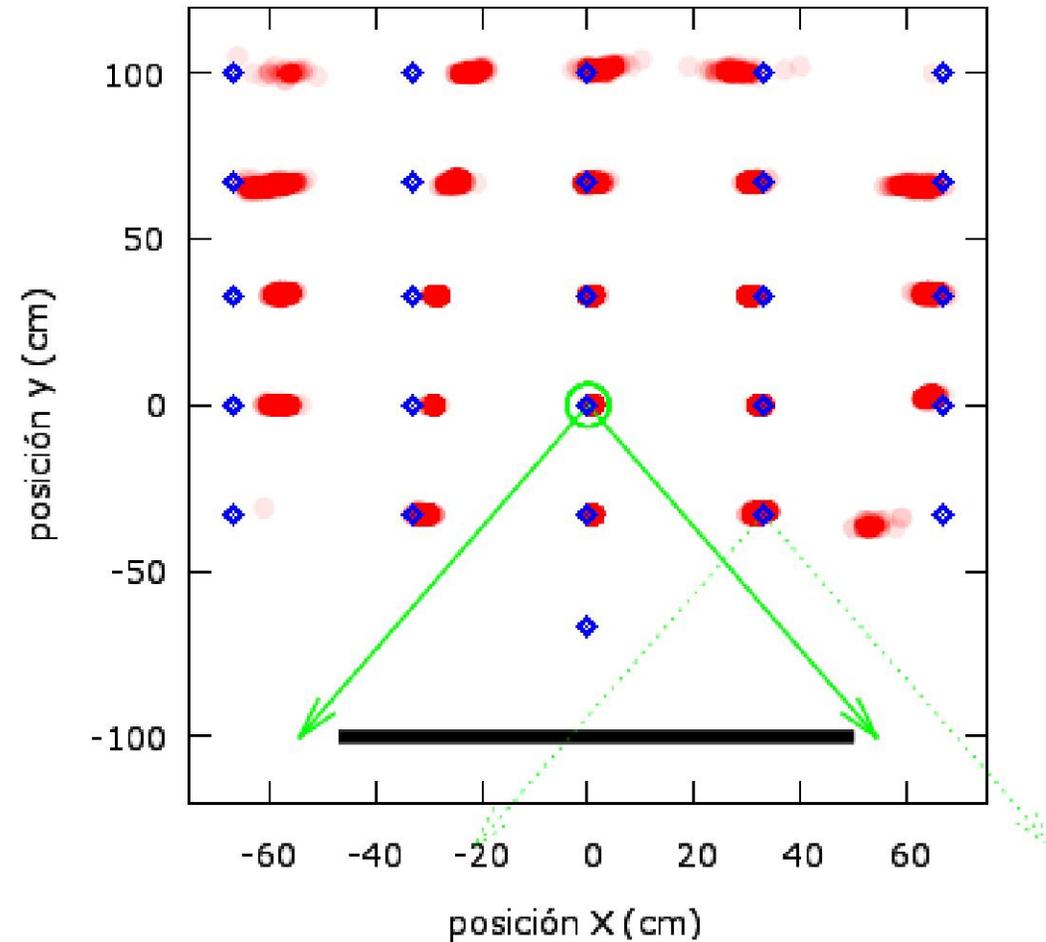
- Kinect de Microsoft como Cámara RGB-D
 - Imágenes con 640x480 pixeles
- Computadora portátil
 - Procesador doble núcleo a 2 GHz
 - 2GB de memoria



Pruebas

Posición real vs estimada

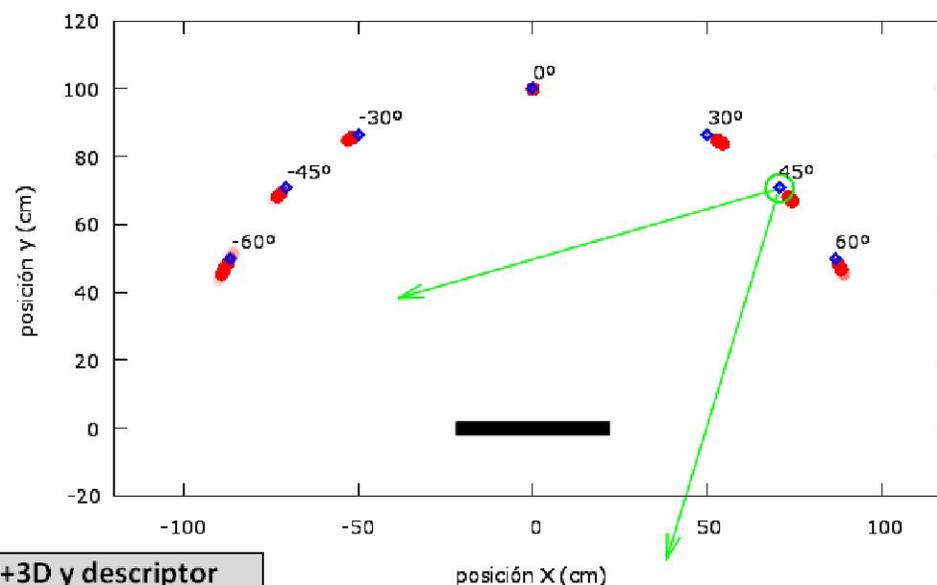
- Tiempo de ejecución 574 ms
- Error promedio
 - Horizontal = 3.75 mm
 - Vertical = 0.39 mm
- Peor posición estimada
 - (67,-33) [cm] \Rightarrow error = 17 mm



Pruebas

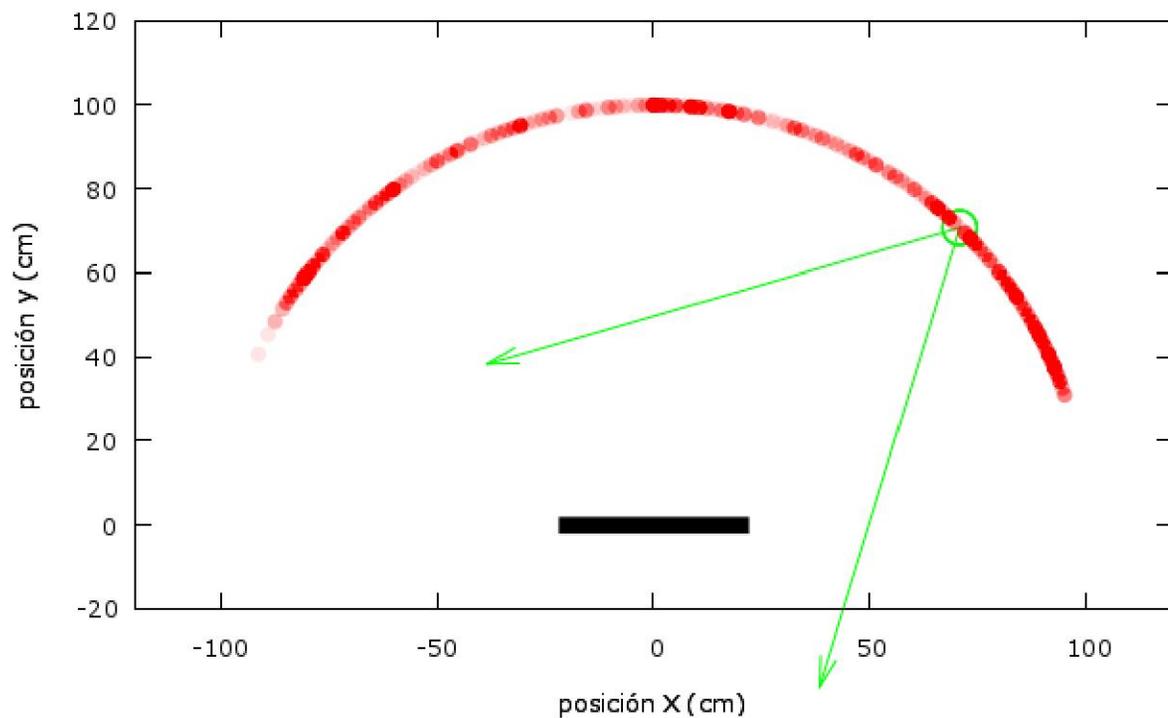
Angulo real vs estimado

- Error promedio: 1.76 [grados]
- Peor ángulo estimado:
 - 4[grados] \Rightarrow error: 4 [grados]

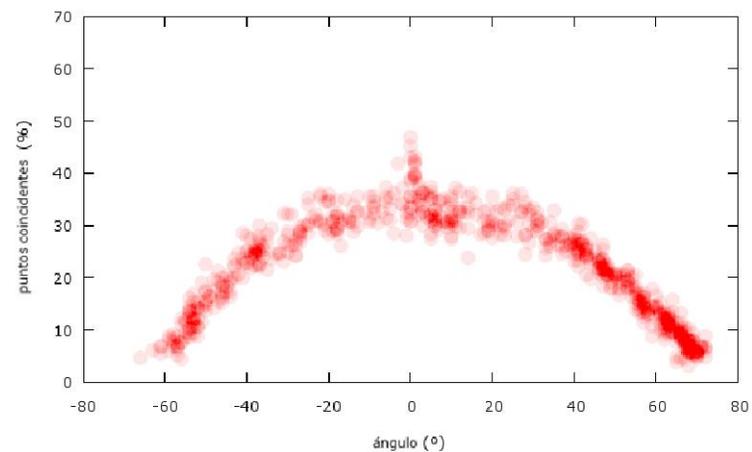


Ángulo	SIFT			Detector FAST+3D y descriptor SIFT		
	% empatado	Puntos detectados	Tiempo de ejecución [ms]	% empatado	Puntos detectados	Tiempo de ejecución [ms]
60	2	779	2077.51	100	225	705.53
45	100	860	2210.41	100	290	735.59
30	100	889	2331.35	100	297	728.80
0	100	846	2582.45	100	265	706.31
-30	100	867	2442.66	100	193	560.79
-45	100	880	2429.34	100	152	735.59
-60	3	826	2254.08	94	181	632.03
promedio		850	2332.54		228	686.38

Pruebas Angulo continuo

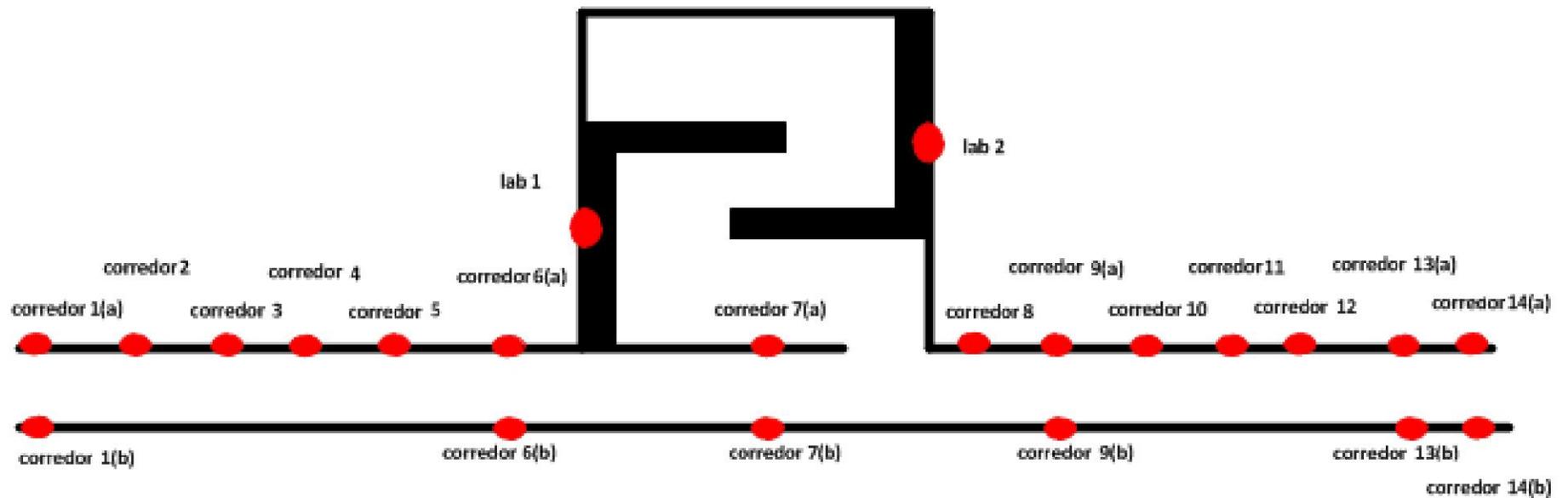


- Máximo ángulo estimado:
72 [grados]



Pruebas

Ambiente en condiciones no controladas



Pruebas

Ambiente en condiciones no controladas

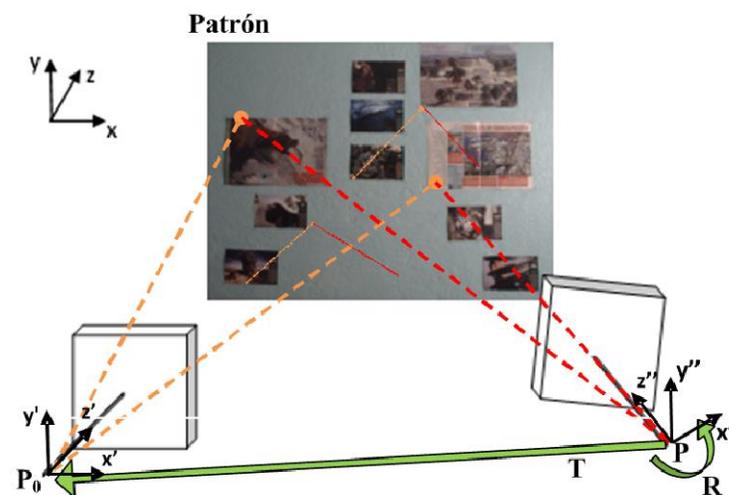
- Ejemplos



Pruebas

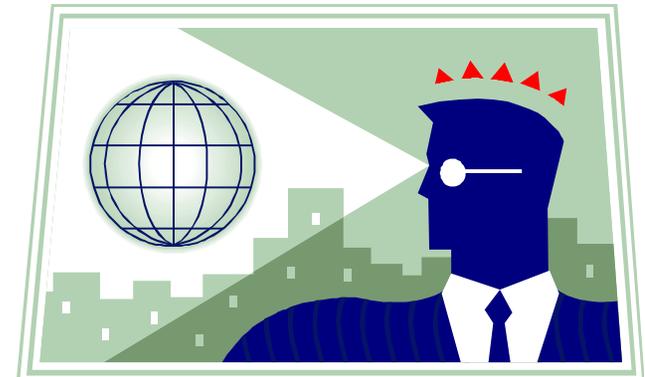
Ambiente en condiciones no controladas

Localización	# ejecuciones	tiempo de ejecución (ms)	% empatado	puntos FAST+3D	# puntos empatados
lab 1	368	180.08	8.70	155	7
lab 2	127	496.34	97.64	215	48
corredor 1	160	245.10	54.38	128	30
corredor 2(a)	87	57.43	0.00	0	0
corredor 2(b)	181	235.36	74.59	105	23
corredor 3	387	73.81	1.81	26	6
corredor 4	200	184.07	63.00	42	10
corredor 5	148	384.65	64.86	195	25
corredor 6(a)	181	211.62	63.54	87	16
corredor 6(b)	83	70.84	0.00	0	0
corredor 7(a)	170	230.36	53.53	92	17
corredor 7(b)	70	402.02	62.86	146	17
corredor 8	86	746.56	90.70	373	51
corredor 9(a)	60	161.10	43.33	63	18
corredor 9(b)	107	204.22	71.96	62	17
corredor 10	225	157.92	14.22	73	7
corredor 11	265	303.05	64.53	148	21
corredor 12	256	285.50	50.39	171	25
corredor 13(a)	32	1832.60	65.63	904	84
corredor 13(b)	233	135.49	33.91	43	8
corredor 14(a)	216	394.76	64.35	200	37
corredor 14(b)	74	272.34	70.27	115	21



CONCLUSIONES

- Generada una metodología de detección y descripción visual a partir de información 3D
- Aumento en robustez de descripción
- Adecuación de Orientación Absoluta para localización visual
- El desempeño es adecuado, realizándose 2 o 3 veces por segundo



TRABAJO FUTURO

- Repetibilidad en detección (Harris, Shi-Tomasi)
- Construcción de descriptores (GLOH)
- Disminuir dimensión de descriptores generados (PCA-SIFT)
- Paralelizar

