UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

POSGRADO EN CIENCIA E INGENIERÍA DE LA COMPUTACIÓN

DESCRIPCIÓN Y DETECCIÓN DE OBJETOS EN 3D POR MEDIO DE CÁMARAS RGB-D

PRESENTA:

CECILIA GÓMEZ CASTAÑEDA

TUTOR:

DR. JESÚS SAVAGE CARMONA

FACULTAD DE INGENIERÍA

Índice

1. Introducción
2. Marco teórico
   - Definiciones: pixel, imágenes, nube de puntos
   - Deteccion y descripción de objetos
   - Detectores y descriptores 2D
   - Detectores y descriptores 3D
   - Algoritmos de alineación de nubes de puntos
3. Solución propuesta
   - Detección, descripción y reconocimiento
4. Pruebas y resultados
5. Conclusiones

Robots de servicio y visión computacional

Un **robot de servicio** personal es construido para asistir al humano en tareas de la vida cotidiana. El robot de servicio domestico realiza diversas tareas que requieren la adquisición de información sobre su entorno.
La detección y reconocimiento de objetos es una capacidad necesaria para este tipo de robots. Lo cual es estudiado en el área de **visión computacional** donde se estudian procesos computacionales que producen descripciones útiles a partir de imágenes

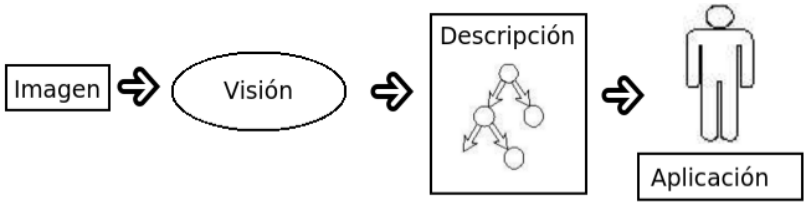


Figura : Justina: Robot de servicio, laboratorio de bio-robotica UNAM

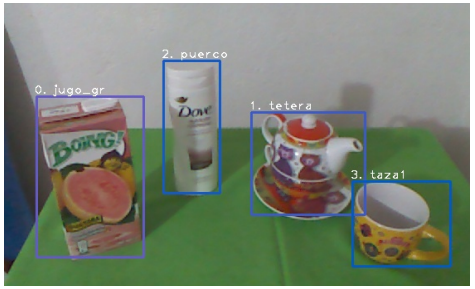


Figura : Aplicación de visión computacional

Objetivos

Objetivo general: Desarrollar un sistema de detección, entrenamiento y reconocimiento de objetos basado en la información que proveen las cámaras RGB-D, para lo que se plantean los siguiente objetivos particulares:

- Implementar un procedimiento de entrenamiento, que incluye realizar el registro 3D a partir de diversas vistas, la generación de descriptores 3D y 2D y el almacenamiento de los objetos.
- Implementar un método de identificación de objetos, utilizando descriptores 3D y 2D, que reduzca el tiempo de búsqueda y mejore la tasa de reconocimiento.

Definiciones

Pixel: unidad mínima de información en una imagen.

Imagen digital: arreglo bidimensional $I[x, y]$ en el cual un pixel es almacenado en cada celda.

En una imagen multiespectral cada pixel es un vector de valores. Por ejemplo, una **imagen RGB** está formada por pixeles de tres canales, cada canal representa la luz en diferentes longitudes de ondas:
$I[x, y] = [rojo(x, y), verde(x, y), azul(x, y)]$

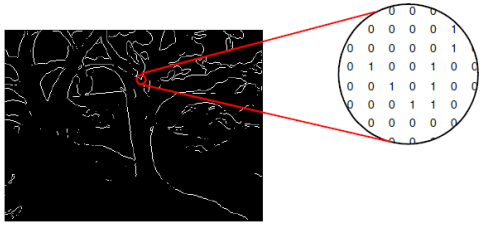


Figura : Imagen monocromática

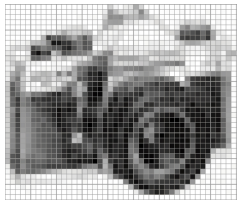


Figura : Imagen escala de grises

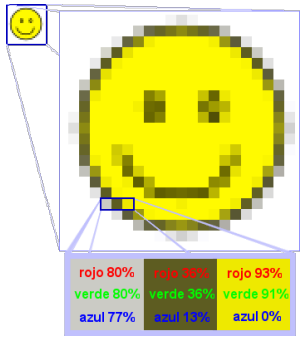


Figura : Imagen RGB

Definiciones

Imagen RGB-D: A cada pixel de la imagen le es asociada información tanto de color como de profundidad.

$$I[x, y] = [r(x, y), v(x, y), z(x, y), d(x, y)]$$

$$d(x, y) = [x_d, y_d, z_d]$$

Nube de puntos: Conjunto de n puntos, $P = \{p_1, p_2, ..., p_n\}$, donde cada p_i es una colección de valores, por ejemplo $p_i = [x, y, z]$.

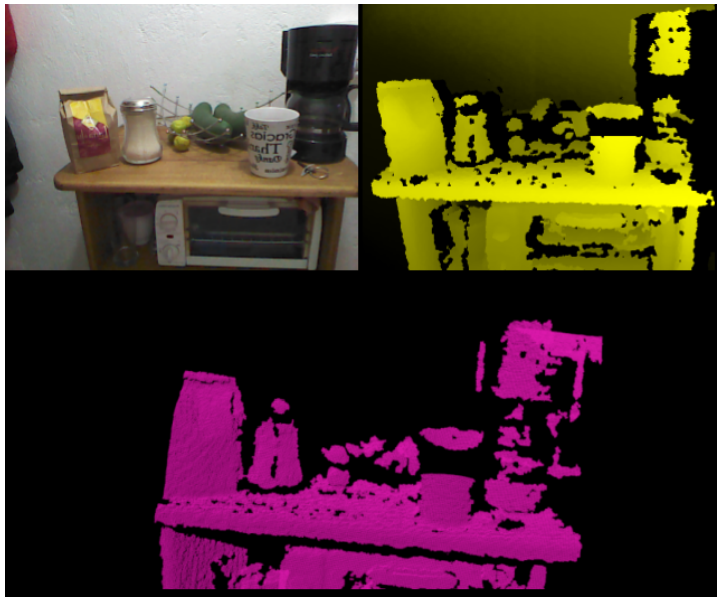


Figura : Imagen RGB-D y su respectiva nube de puntos

Deteccion y descripción de objetos

La deteccion e identificación de objetos a partir de imágenes es un tema especialmente importante en visión comunicacional.

Segmentar área de interés: Procesar la imagen para obtener una nueva imagen que sólo contiene información útil de la escena.

Extracción de puntos de interes: Localización de características que sean encontradas en diferentes vistas de un sólo modelo, es decir, que sean repetibles.

Descripción de puntos de interes: Generar un descriptor, con la vecindad de cada punto de interés detectado, el cual sea distintivo y robusto al ruido y transformaciones.

Empatado de descriptores: Comparación entre descriptores basada en distancia Euclidiana, distancia de Mahalanobis, búsquedas de correspondencias utilizando arboles kd-tree, etc.

.

Detectores y descriptores 2D

SIFT - SURF

SIFT y SURF son extractores de características y descriptores de imágenes 2D invariantes a rotaciones, escalamientos y parcialmente invariantes a cambios de iluminación.

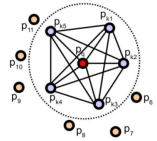
	SIFT	**SURF**
Presentado	por Lowe en 1999	por Bay en 2006
Detección	Diferencia Gaussiana	Determinante de la matriz Hessiana
Descripción	Gradiantes de la vecindad	Respuestas de una ventana de Harr
Tamaño de descriptor	128 elementos	64 elementos

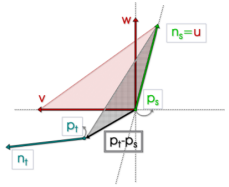
Detectores y descriptores 3D

PFH, FPFH, RSD, SHOT

	Descripción
PFH	A partir de p_s, p_t, n_s, y n_t se calculan:1

$$\left.\begin{array}{lcc} f_1 & = & \langle v, n_t \rangle \\ f_2 & = & ||p_t - p_s|| \\ f_3 & = & \langle u, p_t - p_s \rangle / f_2 \\ f_4 & = & atan(\langle w, n_t \rangle, \langle u, n_t \rangle) \end{array}\right\} idx = \sum_{i=1}^{i \leq 4} step(s_i, f_i) \cdot 2^{i-1}$$

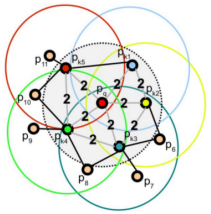




FPFH

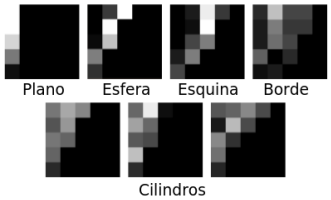
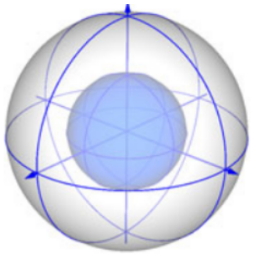
Para cada punto p_i sólo se calculan las relaciones entre éste y sus vecinos obteniendo un Histograma Característico Simplificado (SPFH).

Posteriormente para cada uno se vuelven a evaluar sus k vecinos tomando los valores de SPFH para asignarle un peso al histograma final FPFH de p_i:



Detectores y descriptores 3D

PFH, FPFH, RSD, SHOT

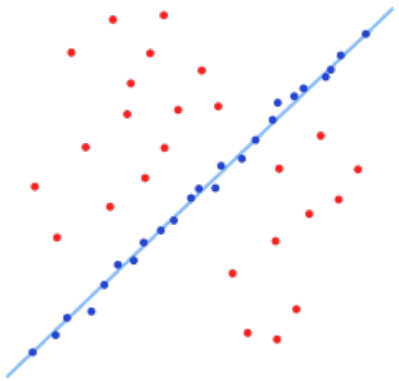
	Descripción
RSD	Se describe la geometría de un punto p_i estimando el radio de las curvaturas que caen en su vecindad. El valor característico de cada punto p_i es conformado por las curvaturas máxima y mínima, las cuales se obtienen calculando la distancia entre el punto y cada punto de su vecindad formada por la siguiente relación: $$d(\alpha) = \sqrt{2r}\sqrt{1 - cos\alpha} \quad (2)$$ 
SHOT	La vecindad de cada punto p_i es definida por los puntos que caen dentro de su división en la malla, para cada p_i el histograma local acumula en sus casilleros valores obtenidos a partir de la función $\cos\theta_i = n_u \cdot n_{v_i}$ del ángulo formado entre la normal de cada punto perteneciente a la división de la malla, n_{v_i}, y la normal en el punto característico n_u. 

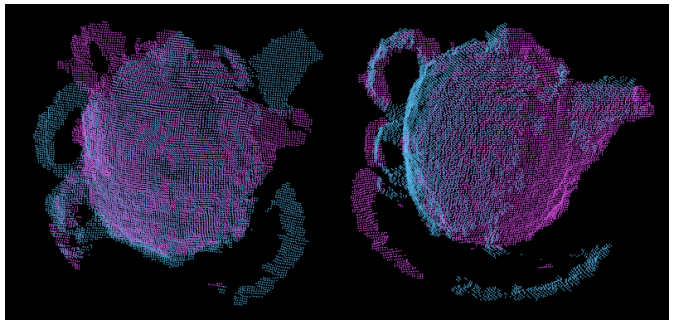
Algoritmos de alineación de nubes de puntos

RANSAC e ICP

RANSAC: Estima los parámetros de un modelo matemático a partir de un conjunto de datos el cual contiene datos que no pertenecen al modelo.

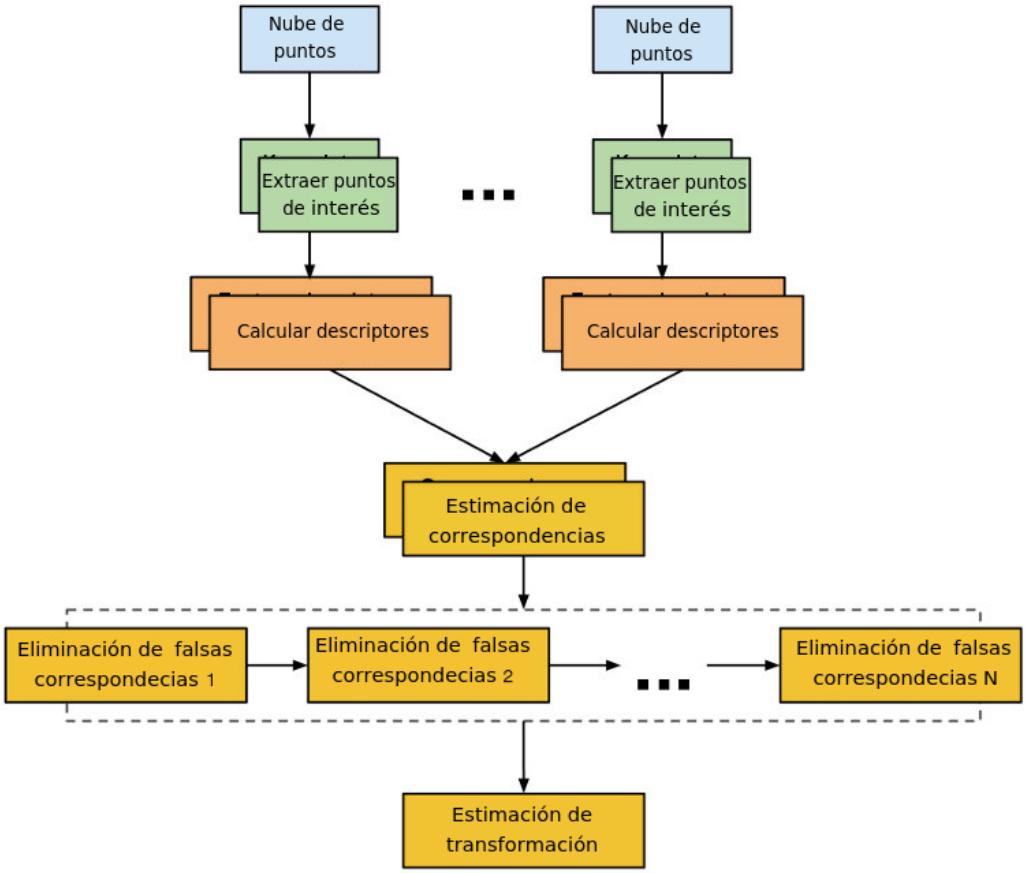
ICP: Dado un par de nubes de puntos, una nube se queda fija, mientras la otra, es acercada buscando iterativamente una transformación (rotación y traslación) que aplicada a esta minimice la distancia entre ambas nubes de puntos





Registro 3D - Alinenación de nubes de puntos

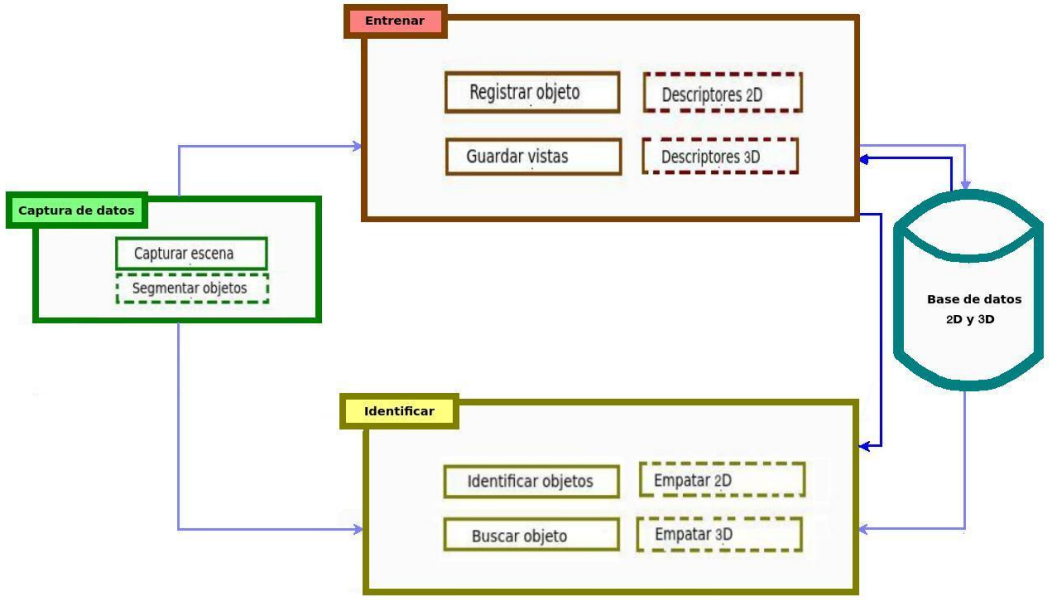
El registro de un objeto consiste el formar un modelo de un objeto a partir de un conjunto de nubes de puntos que contienen diferentes vistas del objeto.



El sistema propuesto de reconocimito de objetos consta de tres modulos principales:

- Captura de datos
- Entrenamiento
- Identificación

y la base de datos donde los objetos son almacenados



Captura de datos

Adquisición y segmentación de escenas

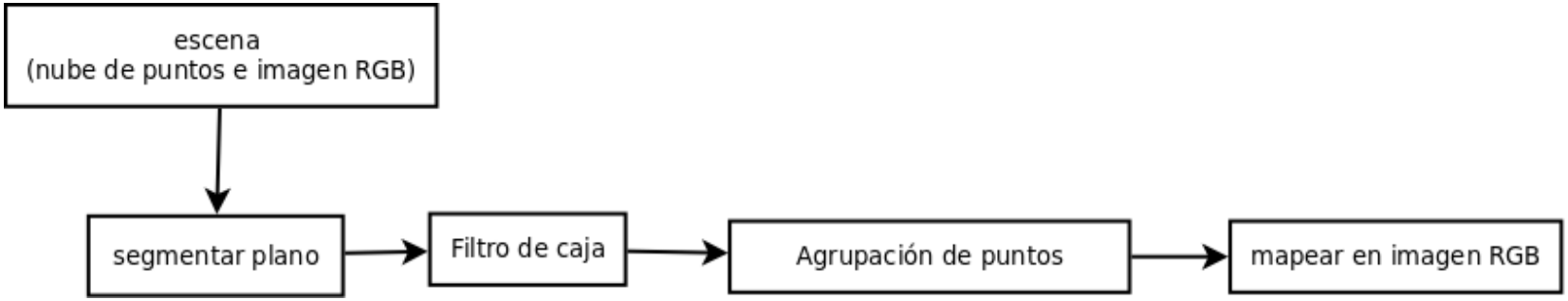
Escena: Se captura la escena a partir del Kinect, dejando en la nube aquellos puntos que estén a un distancia determinada.

Segmentación: Con RANSAC se busca el plano más grande.

Filtro de caja: A partir del plano se define el área de interés aplicando un filtro de caja.

Agrupación de puntos: Los grupos de puntos son separados.

Mapeo en imagen RGB: cada punto de cada cumulo detectado es mapeado a la imagen RGB, para obtener una mascara del objeto.

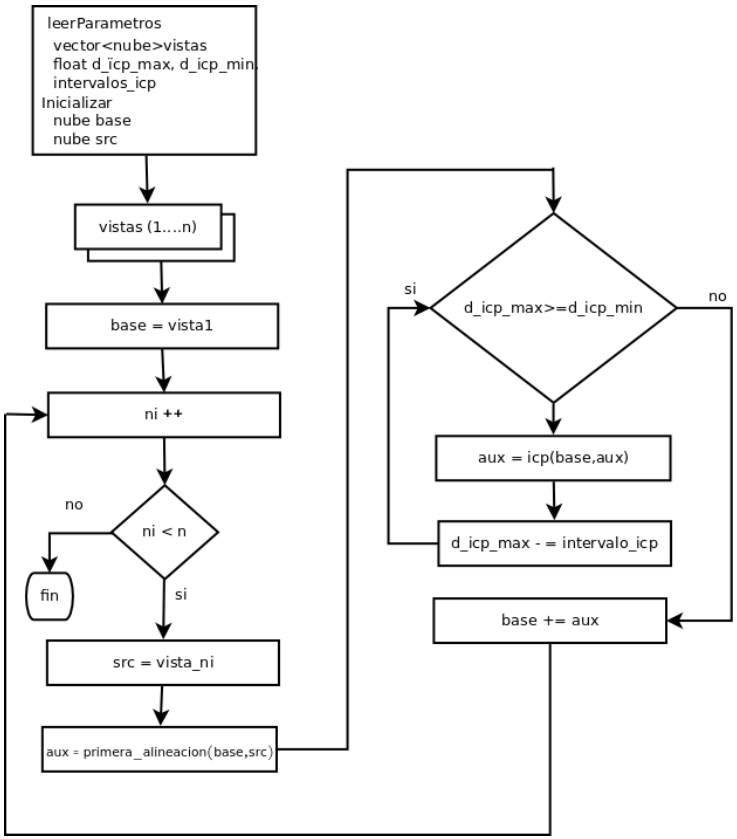


Entrenamiento y registro de objetos

Cada vista capturada es almacenada bajo la siguiente estructura

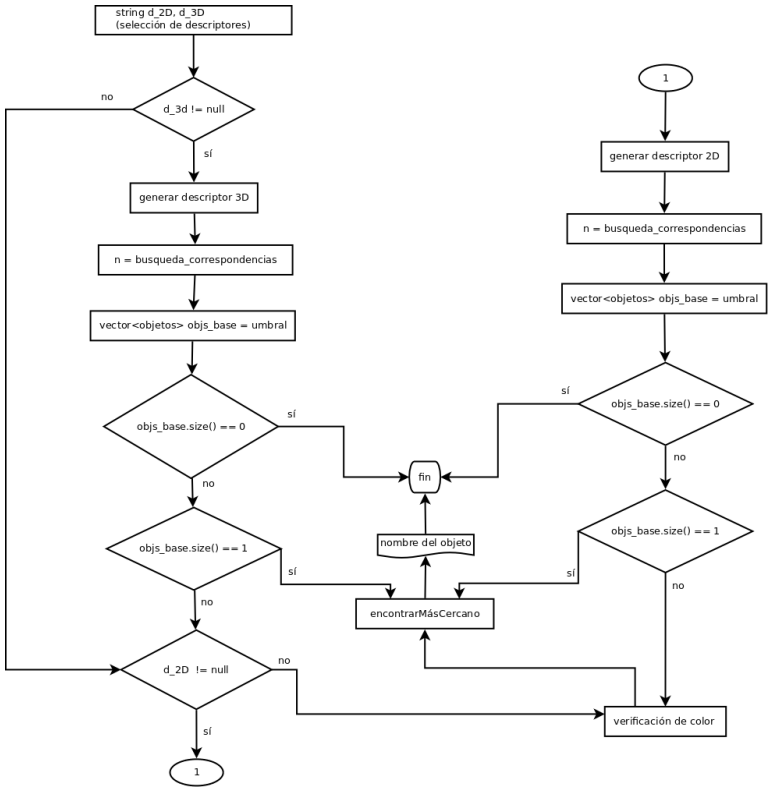
Entrenamiento y registro de objetos

Con las nubes de puntos de un objeto almacenado, se genera un único modelo del objeto, esperando que este sea lo más completo posible.



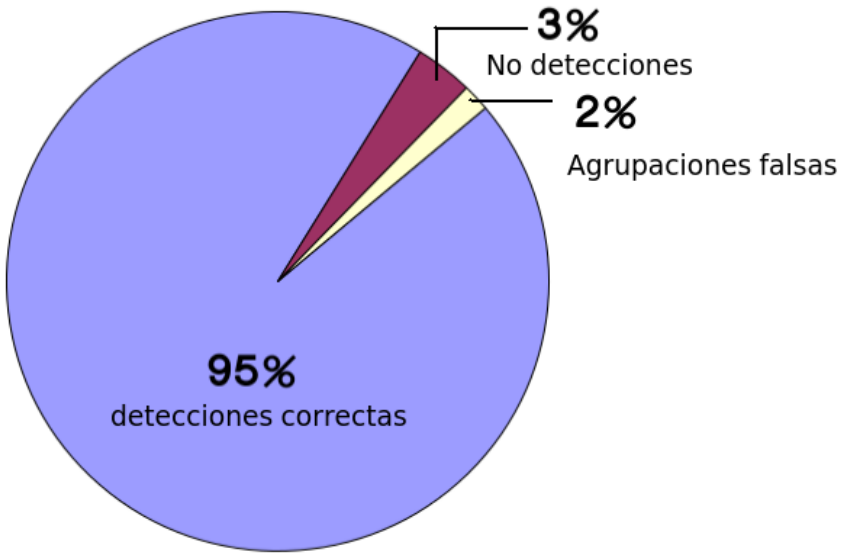
Identificación de objetos

Con las nubes de puntos de un objeto almacenado, se genera un único modelo del objeto, esperando que este sea lo más completo posible.



Detección de objetos

Se analizaron 101 escenas en las cuales se esperaba detectar 492 objetos, de los cuales se detectaron correctamente

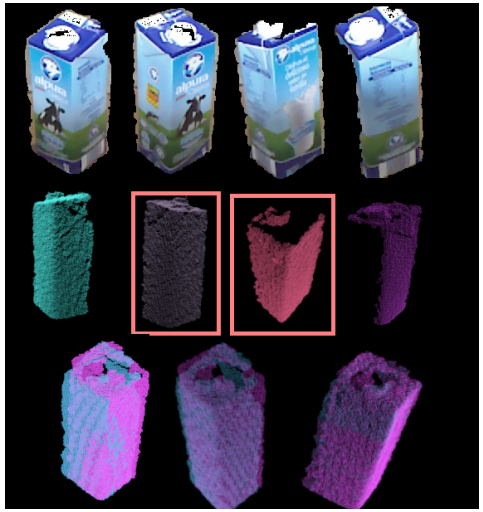


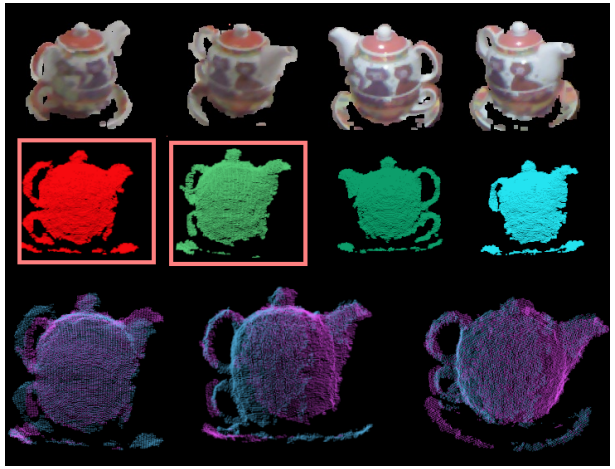
Los principales problemas que se detectaron, fueron

- Objetos no detectados
- Objetos acumulados detectados como un sólo objeto.

Entrenamiento de objetos

Almacenamiento y registro

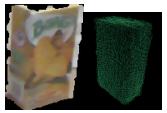




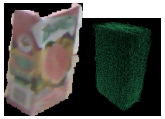
Pruebas y resultados

Objetos entrenados

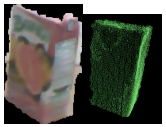
1. Jugo manzana



2. Jugo de mango



3. Jugo de guayaba



4. Jugo grande

5. Taza 1

6. Taza 2

7. Refresco(lata)

8. Tequila(lata)

9. Leche

10. Puerco

11. Jarra

12. Tetera

Cuadro : Objetos registrados

Pruebas y resultados

Pruebas de reconocimiento de objetos

Prueba 1: Escenario con objetos conocidos desconocidos, con el dispositivo a 1m

Prueba 2: Escenario con objetos conocidos desconocidos, con el dispositivo a 1m

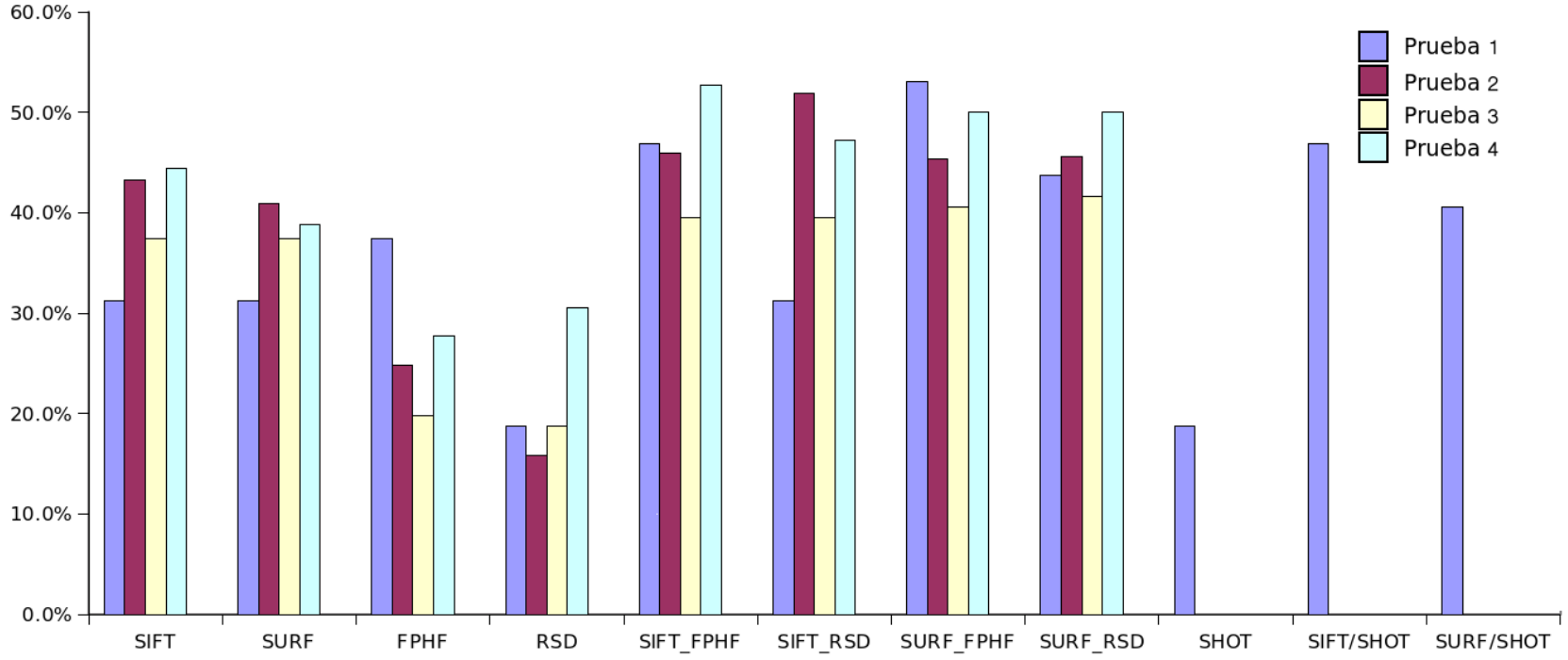
Prueba 3: Escenario con objetos conocidos desconocidos, con el dispositivo a .5m

Prueba 4: Escenario con objetos conocidos, con el dispositivo a .5m

imagenes

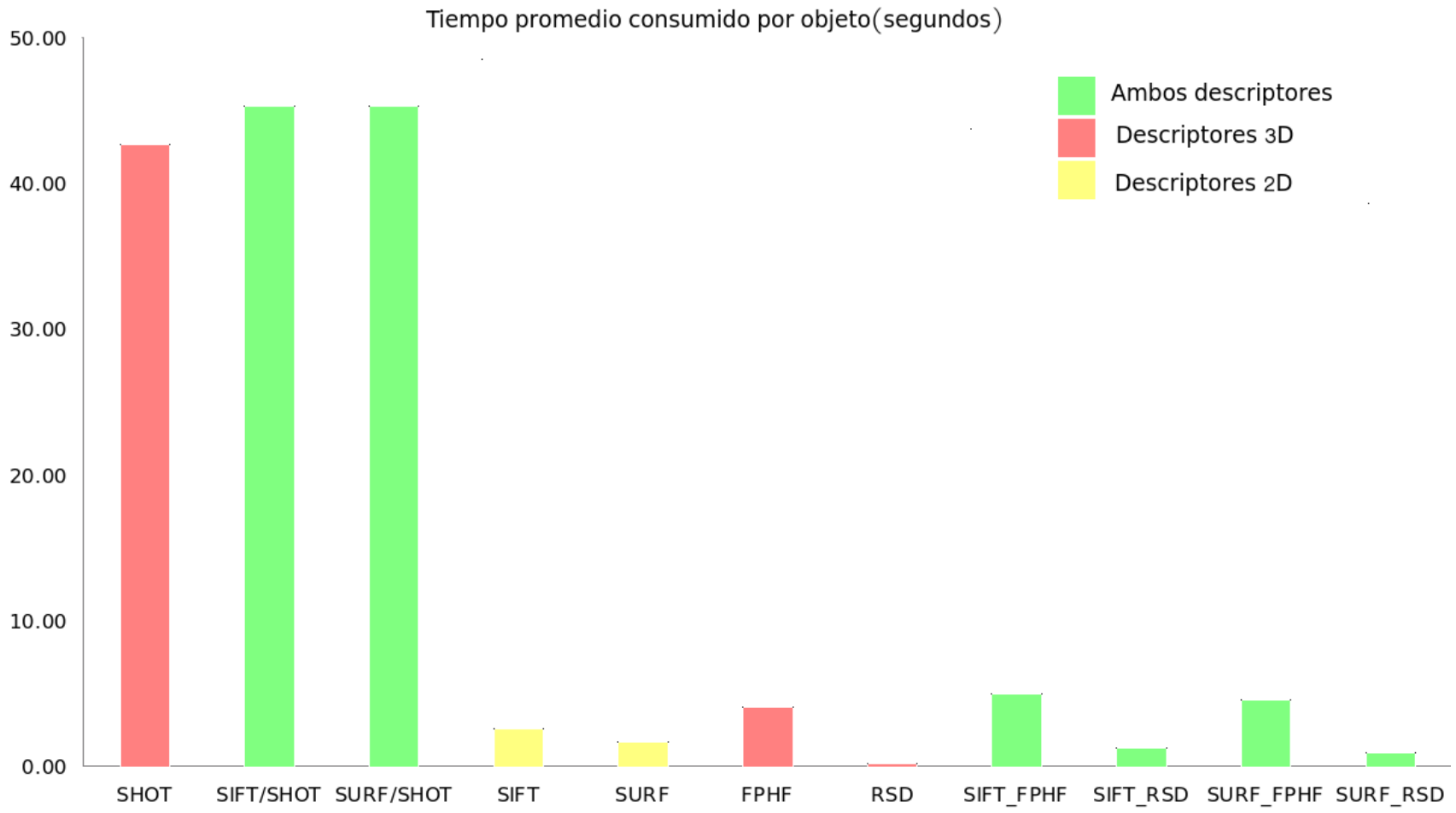
Pruebas y resultados

Pruebas de reconocimiento de objetos



Pruebas y resultados

Pruebas de reconocimiento de objetos



Conclusiones

La visión por computadora es un área que impacta directamente en la ... de un robot de servicio. Aunque existen diferentes ... cada problema se debe atacar de diferentes maneras.
La implementación de una arquitectura que permita una fácil adapactacion a diferentes porblemas es necesaria por lo que la s pricipaes ventajs del sitema propuesta es la que se pueden insertar nuevas funciones sin tenr que modificar demasidao codigo, y que de las funciones programasdas se pueden realizar diversas pruebas modificando los archivos de configuracion.
Respecto a procedimiento propuesto en cascada para realizar la identificación de objetos, se puede decir que, el procesar por etapas utilizando diferentes mejora la tasa de reconocimiento y acelera e tiempo de análisis de una escena.
También hay que menciona que le pipeline propuesto para identificación

References I