

Análisis Comparado de Métodos de Reconocimiento de Landmarks para la Navegación Autónoma en Entornos de Interior mediante Mapas Topológicos Visuales. Reconocedor Vecino Más Cercano con Umbral Adaptativo plus Histogramas y Reconocimiento Sintáctico-estructural Basado en la Representación de Imágenes mediante Ristas Simbólicas.

Autora: Andrea Velarde Chávez

Tutor: Darío Maravall Gómez-Allende

**Máster en Inteligencia Artificial
ETS de Ingenieros Informáticos de la Universidad Politécnica de Madrid**



Julio 2022

Índice

1 Introducción

- Objetivos

2 Representación de imágenes mediante ristras simbólicas

- Representación mediante ristras simbólicas asociadas a lenguajes de tipo natural

- Escenarios

3 Módulo de detección de objetos YOLO

- Entorno
- Objetos
- Pruebas Estáticas
- Pruebas Dinámicas

4 Análisis comparativo

■ Reconocedor “vecino más Próximo” (VMP) plus histograma

- Landmarks del entorno
- Histogramas
- VMP
- Pruebas estáticas
- Pruebas dinámicas

■ Reconocedor de objetos con distancia de Hamming plus ristras simbólicas

- Localización de objetos
- Pruebas estáticas
- Distancia de Hamming
- Pruebas dinámicas

5 Conclusión



Objetivos

Objetivo global:

- Desarrollo y prueba de un reconocedor de Landmarks en un entorno de interior.

Objetivos específicos:

- Representación de imágenes digitales mediante ristras(strings) simbólicas.
- Módulo de detección y reconocimiento de objetos en imágenes digitales mediante redes convolucionales y aprendizaje profundo: YOLO.
- Análisis comparado de eficiencia de 2 métodos en navegación autónoma con mapas topológicos visuales: reconocedor “Vecino Más Próximo” (VMP) con la distancia euclídea plus histograma VS reconocedor con la distancia de Hamming plus ristras simbólicas basado en la librería YOLO.
- Realizar unos experimentos en un entorno de interior de validación del sistema más óptimo y analizar resultados obtenidos.



Definiciones formales 1

Definición formal de un lenguaje de tipo natural: todos los lenguajes tienen asociados complementos básicos del lenguaje.

Vocabularios del lenguaje espacial utilizado:

- $V_{objetos} = \{ < objeto_1 >, \dots, < objeto_n > \}$, referente a todos los objetos físicos del universo de trabajo del sistema y en donde $< objeto_i >$ representa el símbolo o vocablo asociado al objeto i
- $V_{localizacion} = \{ < localizacion\ espacio > \} = \{ ARI, ARC, ARD, I, C, D, ABI, ABC, ABD \}$



arriba a la izquierda	arriba en el centro	arriba a la derecha	=	ari	arc	ard
a la izquierda	centro	a la derecha		i	c	d
abajo a la izquierda	abajo en el centro	abajo a la derecha		abi	abc	abd

Disposición de localizaciones en la escena y sus correspondientes siglas identificativas.

Definiciones formales 2

Reglas de producción:

$Sentence \rightarrow < object_1 > < localizacion\ espacio >, \dots, < object_n >$
 $< localizacion\ espacio >$

$< object > ::= \text{an object} \in V_{objects}$

$< localizacion\ espacio > ::= \text{una localización espacial} \in V_{localizacion}$

$f_{verdad}(\text{sentencia}) = f_{verdad}(O_1) \times f_{verdad}(L) \times f_{verdad}(O_2); \text{ en donde}$
 $f_{verdad}(\cdot) = 1 \text{ si el predicado } (\cdot) \text{ es cierto.}$



Escenario 1

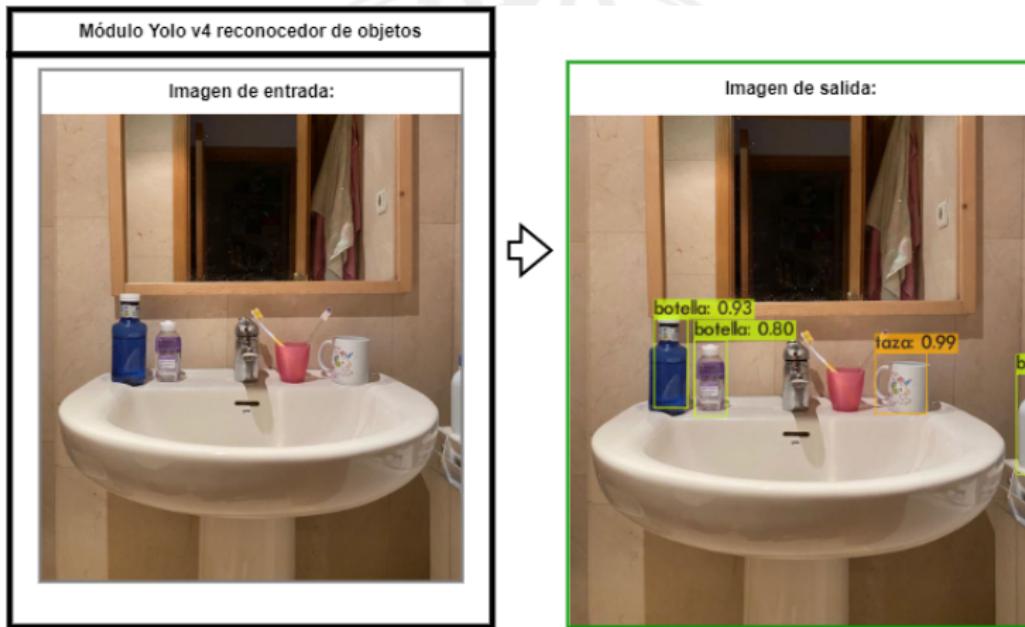
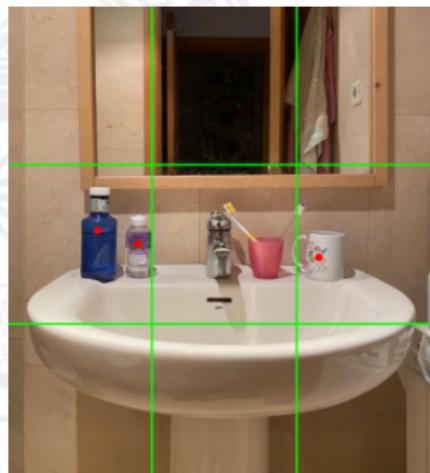
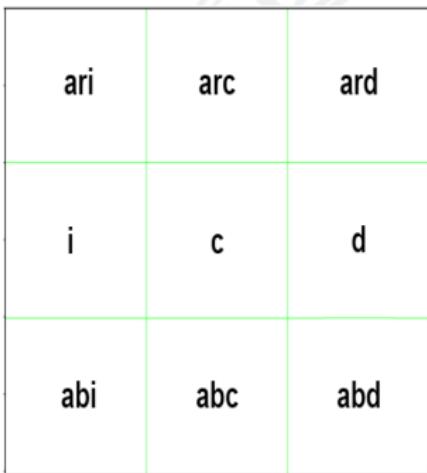


Gráfico representativo del proceso de la detección de objetos a partir del módulo Yolo v4. A la izquierda, imagen original de baño lleno de objetos proporcionada al detector. A la derecha, resultado con las cajas de recuadro de los objetos y sus clases resultado de la predicción



Escenario 1- Objetos y Localización

$$V_{objetos} = \{ \text{botella, taza} \}$$



A la izquierda, imagen en blanco dividida en 9 cuadrantes (líneas verdes) definidos como posiciones en el espacio. A la izquierda, imagen de entrada dividida en dichos cuadrantes (líneas verdes), los puntos (color rojo) se representan en el centro de masa de cada objeto.



Escenario 1- Sentencias y Ristra simbólica

Aplicando nuestro método hemos obtenido la siguiente sentencia:

**“taza a la derecha, botella a la izquierda,
botella a la izquierda, botella a la
derecha”**

Para convertir dicho resultado en ristras simbólicas, se trataría la sentencia de la siguiente manera y dando como resultado:

*taza a la derecha = TD
botella a la izquierda = BI
botella a la derecha = BD
“TDBIBIBD”*



Escenario 2

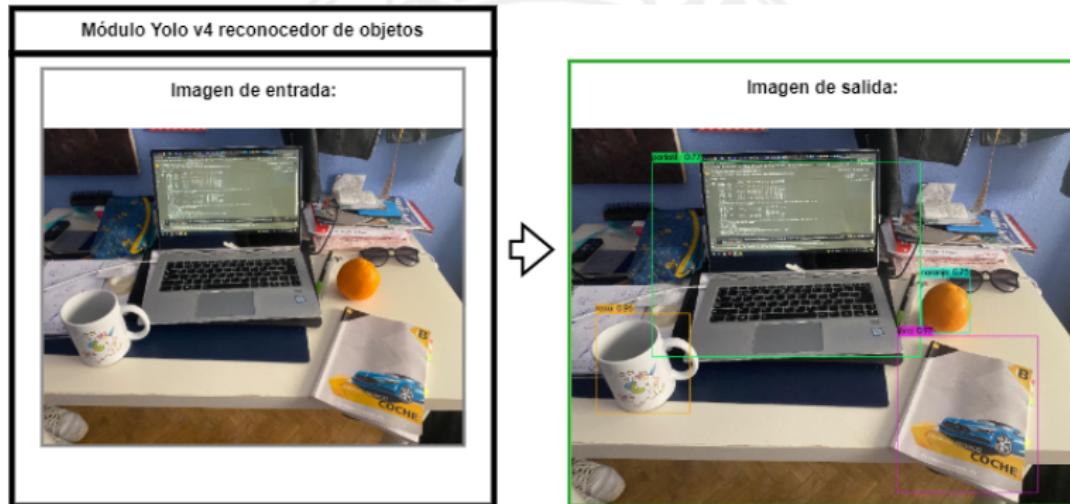


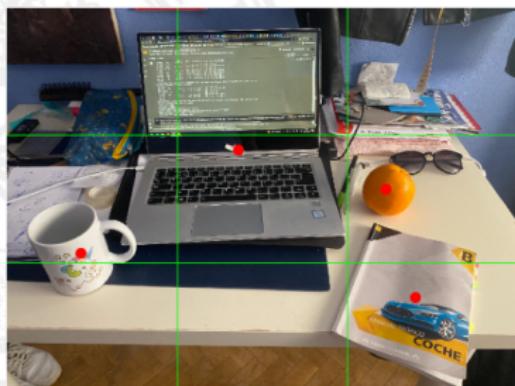
Gráfico representativo del proceso de la detección de objetos a partir del módulo Yolo v4. A la izquierda, imagen original de escritorio lleno de objetos proporcionada al detector. A la derecha, resultado con las cajas de recuadro de los objetos y sus clases resultado de la predicción.



Escenario 2- Objetos y Localización

$$V_{objetos} = \{ \text{taza, portatil, naranja, libro} \}$$

ari	arc	ard
i	c	d
abi	abc	abd



A la izquierda, imagen en blanco dividida en 9 cuadrantes (líneas verdes) definidos como posiciones en el espacio. A la izquierda, imagen de entrada dividida en dichos cuadrantes (líneas verdes), los puntos (color rojo) se representan el centro de masa de cada objeto.



Escenario 2- Sentencias y Ristra simbólica

Aplicando nuestro método hemos obtenido la siguiente sentencia:

"taza a la izquierda, portatil en el centro, naranja a la derecha, libro abajo a la derecha"

Para convertir dicho resultado en ristras simbólicas, se trataría la sentencia de la siguiente manera y dando como resultado:

*taza a la izquierda = TI
portatil en el centro = PC
naranja a la derecha = ND
libro abajo a la derecha = LABD
“TIPCNDLABD”*



Conclusiones del método de ristras

Hemos visto que el método funciona para dos escenarios complejos multiobjeto lo cual lo valida como método para su posterior aplicación al reconocimiento de los landmarks del entorno experimental de navegación. Lo utilizaremos posteriormente.



Landmarks del entorno



Distribución de Landmarks en el plano



Disposición de objetos



Mapa topológico de landmarks basado en el espacio del plano del domicilio



Objetos



Objeto vasija



Segundo objeto taza



Primer objeto taza



Objeto libro



Objetos



Primer objeto botella



Objeto paraguas



Segundo objeto botella



Objeto naranja



Objetos



Objeto cepillo de dientes



Objeto portátil



Objeto ratón



Pruebas Estáticas

Ratios de acierto (RA) y error (RE)

$$RA = \frac{Nº \text{ objetos correctamente detectados y reconocidos}}{Nº \text{ casos totales}} \quad (1)$$

$$RE = \frac{Nº \text{ objetos incorrectamente detectados y reconocidos}}{Nº \text{ casos totales}} \quad (2)$$

Objeto	Ratio de error	Ratio de acierto
Vasija	0/10=0	10/10=1
Taza	0/23=0	23/23=1
Libro	0/11=0	11/11=1
Botella	0/19=0	19/19=1
Paraguas	0/8=0	8/8=1
Naranja	0/11=0	11/11=1
Cepillo dental	0/13=0	13/13=1
Portátil	0/8=0	8/8=1
Ratón	0/8=0	8/8=1

Tabla: Ratios de error y acierto para todas las clases.



Pruebas Dinámicas

	Detectados correctamente	Detectados incorrectamente	Total	% final	Ratio de error	Ratio de acierto
Vasija	100	0	100	100 %	0/100=0	100/100=1
Taza	191	9	200	95,5 %	9/200=0.045	191/200=0.955
Libro	85	15	100	85 %	15/100=0.15	85/100=0.85
Botella	200	0	200	100 %	0/200=0	200/200=1
Paraguas	100	0	100	100 %	0/100=0	100/100=1
Naranja	79	21	100	79 %	21/100=0.21	79/100=0.79
Cepillo dental	100	0	100	100 %	0/100=0	100/100=1
Portátil	100	0	100	100 %	0/100=0	100/100=1
Ratón	100	0	100	100 %	0/100=0	100/100=1
Media			1100	95,9 %	0.04090	0.95909

Tabla: Ratios de error y acierto para todas las clases.

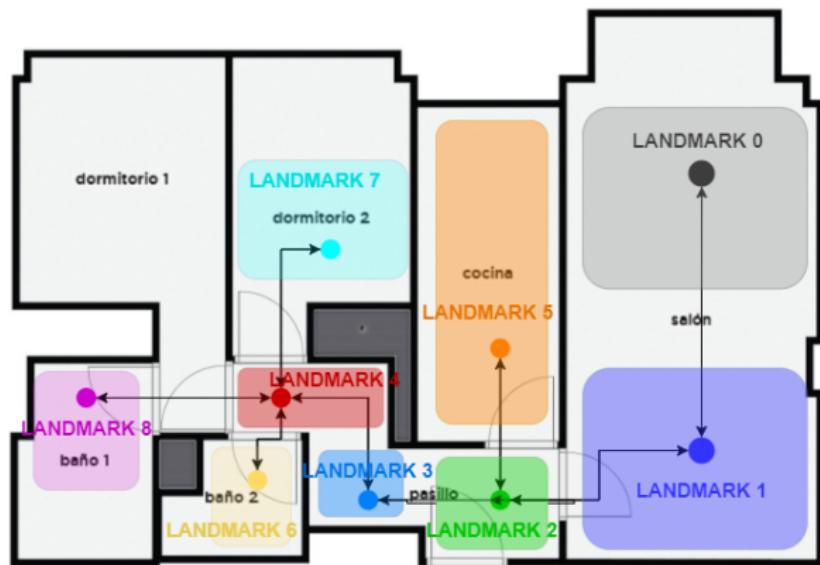


Conclusiones de validación de pruebas

Podemos dar validado este módulo en estática y dinámica. Los resultados en las pruebas estáticas dan de media de acierto para todos los objetos un 1 (la máxima), y de media de error 0 (la mínima). En las pruebas dinámicas la media general de acierto es de 0.95909 y de error 0.04090. Puede ser una herramienta muy interesante para comprobar la navegación del detector.



Landmarks del entorno



Distribución de Landmarks en el plano



Histogramas de contraste

Histograma de contraste de imagen

A modo de ejemplo, desde la siguiente imagen en escala de grises se calcula el histograma de su único canal de color. A la derecha, el histograma resultante de la imagen , en el eje “x” los posibles valores que pueden tomar los pixels [0,255] y en el eje “y” su frecuencia.

Histograma de imagen del Landmark 0 (escala de grises)

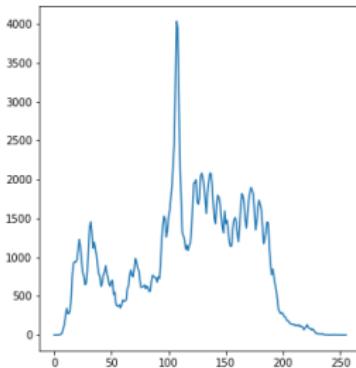


Imagen e histograma Landmark 0



Histogramas de contraste



Histograma de imagen del Landmark 1 (escala de grises)

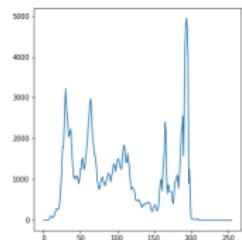


Imagen e histograma Landmark 1



Histograma de imagen del Landmark 2 (escala de grises)

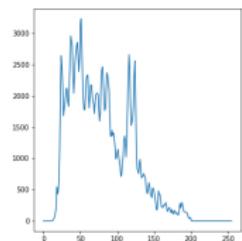


Imagen e histograma Landmark 2



Histograma de imagen del Landmark 3 (escala de grises)

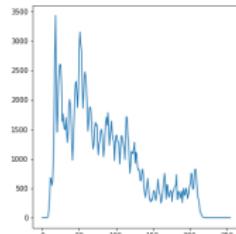


Imagen e histograma Landmark 3

Histograma de imagen del Landmark 4 (escala de grises)

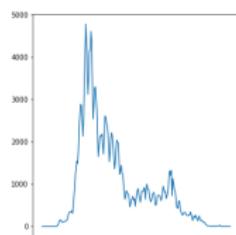


Imagen e histograma Landmark 4

Histogramas de contraste



Histograma de imagen del Landmark 5 (escala de grises)

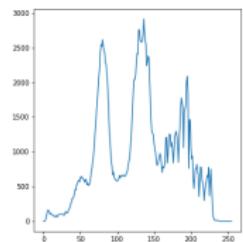


Imagen e histograma Landmark 5



Histograma de imagen del Landmark 6 (escala de grises)

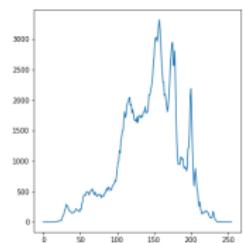


Imagen e histograma Landmark 6



Histograma de imagen del Landmark 7 (escala de grises)

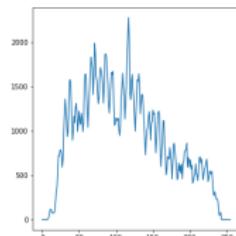


Imagen e histograma Landmark 7

Histograma de imagen del Landmark 8 (escala de grises)

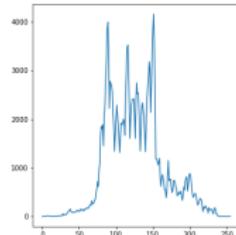
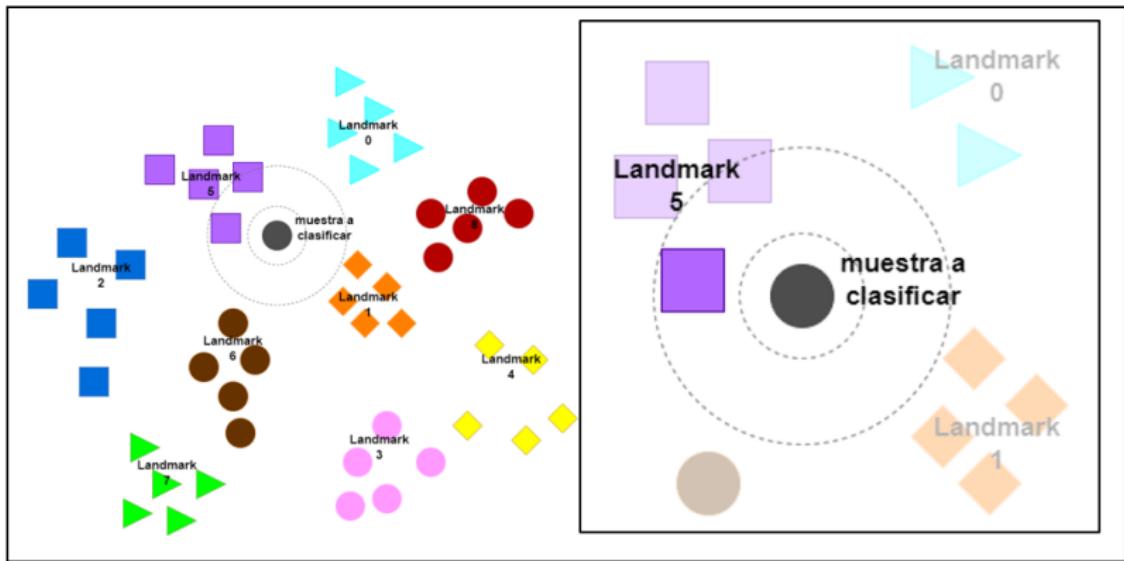


Imagen e histograma Landmark 8

Vecino más próximo



A la izquierda, representación del espacio característico de 9 clases y una muestra a clasificar. A la derecha, vecino más cercano de la muestra a clasificar.



Pruebas estáticas

C	precision	recall	f1-score	support
nodo 0	1.00	1.00	1.00	18
nodo 1	1.00	1.00	1.00	18
nodo 2	1.00	1.00	1.00	18
nodo 3	1.00	1.00	1.00	18
nodo 4	1.00	1.00	1.00	18
nodo 5	1.00	1.00	1.00	18
nodo 6	1.00	1.00	1.00	18
nodo 7	1.00	1.00	1.00	18
nodo 8	1.00	1.00	1.00	18
accuracy			1.00	162
macro avg	1.00	1.00	1.00	162
weighted avg	1.00	1.00	1.00	162

Resultados de la validación del reconocedor del vecino más proximo entrenado con histogramas.



Pruebas dinámicas

	L0	L1	L2	L3	L4	L5	L6	L7	L8	Total
L0	167	7	1	0	0	0	0	0	0	175
L1	0	0	41	0	28	12	36	0	0	117
L2	6	0	94	0	27	0	0	19	0	146
L3	0	0	9	72	1	0	3	32	0	117
L4	0	0	50	1	38	0	1	27	0	117
L5	0	0	0	0	0	131	57	0	74	262
L6	0	0	0	0	0	0	114	27	5	146
L7	0	8	42	0	0	0	0	183	0	233
L8	0	0	0	0	0	0	0	44	73	117

	Ratio de error	Ratio de acierto
L0	$8/175 = 0.0457$	$167/175 = 0.9542$
L1	$117/117 = 1$	$0/117 = 0$
L2	$52/146 = 0.3561$	$94/146 = 0.6438$
L3	$45/117 = 0.3846$	$72/117 = 0.6153$
L4	$79/117 = 0.6752$	$38/117 = 0.3247$
L5	$131/262 = 0.5$	$131/262 = 0.5$
L6	$32/146 = 0.2191$	$114/146 = 0.7808$
L7	$50/233 = 0.2145$	$183/233 = 0.7854$
L8	$44/117 = 0.3760$	$73/117 = 0.6239$

Matriz de confusión pruebas dinámicas

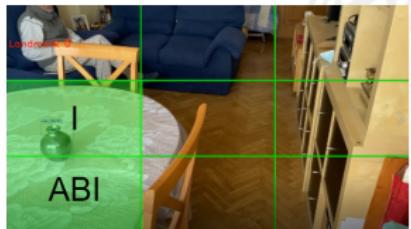
Ratios de error y de acierto

Media ratio de error: 0.41902...

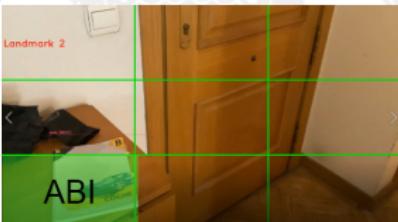
Media ratio de acierto: 0.60979...



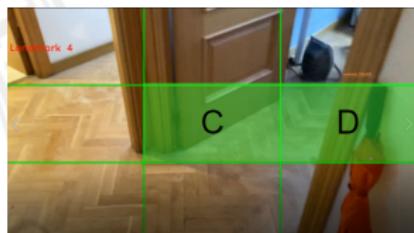
Reconocedor de objetos con distancia de Hamming plus ristras simbólicas



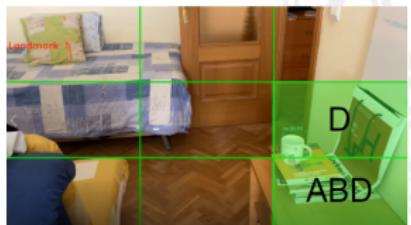
Localización objeto Vasija en L0



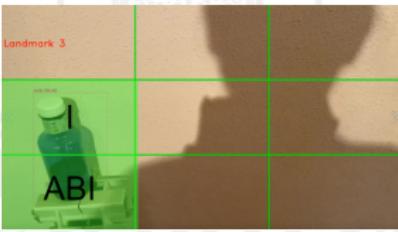
Localización objeto Libro en L2



Localización objeto Paraguas en L4



Localización objeto Taza en L1



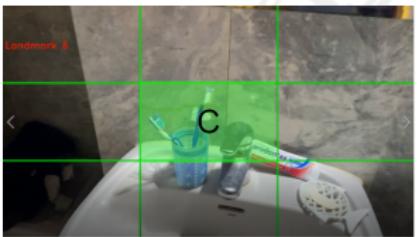
Localización objeto Botella en L3



Localización objeto Naranja en L5



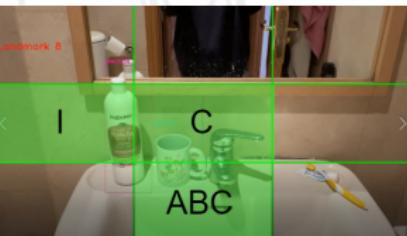
Reconocedor de objetos con distancia de Hamming plus ristras simbólicas



Localización objeto Cepillo en L6



Localización objetos Portátil y Ratón en L7



Localización objetos Taza y Botella en L8



Pruebas estáticas

	Nº frames detectados correctamente	Nº total de frames	% de acierto
Landmark 0	242	243	99.58847 %
Landmark 1	184	193	95.33678 %
Landmark 2	202	202	100 %
Landmark 3	217	217	100 %
Landmark 4	231	231	100 %
Landmark 5	214	229	93.44978 %
Landmark 6	234	236	99.15254 %
Landmark 7	285	285	100 %
Landmark 8	214	229	93.44978 %
Total	2023	2065	97.96610 %

Resultados de pruebas estáticas por Landamark



Distancia de Hamming

Utilidad de la Distancia de Hamming

Sólo acepta una diferencia igual a 0, es decir, la ristra simbólica obtenida a partir del reconocedor debe coincidir de principio a fin con alguna de las posibles ristras de un Landmark.

- Distancia de hamming entre ristra L8 "TC"y ristra L3 "BAB": 1.0
- Distancia de hamming entre ristra L8 "TC"y ristra L5 "NI": 1.0
- Distancia de hamming entre ristra L8 "TC"y ristra L6 "CC": 0.5
- Distancia de hamming entre ristra L8 "TC"y ristra L1 "TD": 0.5
- Distancia de hamming entre ristra L8 "TC"y ristra L1 "TABD": 1.0
- Distancia de hamming entre ristra L8 "TC"y ristra L8 "BCTC": 1.0
- Distancia de hamming entre ristra L8 "TC"y ristra L8 "BITC": 1.0



Pruebas dinámicas

Vídeo #1			
Landmark	Aciertos	Muestras	Precisión de acierto
0	50	87	57.47126436781609
1	115	117	98.29059829059828
2	112	117	95.72649572649573
5	170	175	97.14285714285714
3	117	117	100.0
4	117	117	100.0
6	134	146	91.78082191780823
7	262	262	100.0
8	97	118	82.20338983050848
Total	1174	1256	93.47133 %

Resultados vídeo # 1

Vídeo #2			
Landmark	Aciertos	Muestras	Precisión de acierto
0	84	94	89.36170212765957
1	78	88	88.63636363636364
2	48	59	81.35593220338984
5	173	175	98.85714285714286
3	116	117	99.14529914529915
4	88	88	100.0
6	117	117	100.0
7	167	175	95.42857142857143
8	103	126	81.74603174603175
Total	974	1039	93.74398 %

Resultados vídeo # 2

Vídeo #3			
Landmark	Aciertos	Muestras	Precisión de acierto
0	82	87	94.25287356321839
1	57	59	96.61016949152543
2	38	59	64.40677996101694
5	144	146	98.63013698630137
3	114	117	97.43589743589743
4	30	30	100.0
6	117	117	100.0
7	175	175	100.0
8	137	144	95.13888888888889
Total	894	934	95.71734 %

Resultados vídeo # 3

Resultados totales

Landmark	Aciertos	Muestras	Precisión de acierto
0	216	268	80.59701492537313
1	250	264	94.69696969696969
2	198	235	84.25531914893617
5	487	496	98.18548387096774
3	347	351	98.86039886039886
4	235	235	100.0
6	368	380	96.84210526315789
7	604	612	98.69281045751633
8	337	388	86.85567010309278
Total	3042	3229	94.20873 %

Resultados totales de los 3 vídeos



Vídeos

■ **Vídeo demostración 1 pruebas dinámicas:**

[https://drive.google.com/file/d/
1pM-gC6zAnzSWnFgucu6U30aYzoEbgCub/view?usp=sharing](https://drive.google.com/file/d/1pM-gC6zAnzSWnFgucu6U30aYzoEbgCub/view?usp=sharing)

■ **Vídeo demostración 2 pruebas dinámicas:**

[https://drive.google.com/file/d/
1Lonn2gUqvxFpqERRa1WwD9e54qFVufXr5/view?usp=sharing](https://drive.google.com/file/d/1Lonn2gUqvxFpqERRa1WwD9e54qFVufXr5/view?usp=sharing)

■ **Vídeo demostración 3 pruebas dinámicas:**

[https://drive.google.com/file/d/
1BCGizG6myCbTD13C8ShfLVsGXaY71q-I/view?usp=sharing](https://drive.google.com/file/d/1BCGizG6myCbTD13C8ShfLVsGXaY71q-I/view?usp=sharing)



Conclusiones generales

- Hemos experimentado con un nuevo método de reconocimiento de imágenes de Landmarks basado en su representación mediante ristras simbólicas, se han obtenido buenos resultados en las pruebas dinámicas alcanzando un **94,20 % de precisión de acierto general** comparado con el sistema de reconocimiento de imágenes basado en histogramas cuyo rendimiento fue de **60,97%**.
- El sistema es **sensible a** factores como la **iluminación**, remarcará o ensombrecerá los objetos característicos.
- Se realizó un **análisis de 2 sistemas de reconocimiento** de Landmarks y escogido el más efectivo. Trabajando en técnicas de visión por computador, sistemas YOLO y técnicas en campos de la investigación aplicada a la imagen digital, reconocimiento basado en histogramas de contraste.
- Personalmente he disfrutado del proyecto, del conocimiento adquirido en uno de los campos de la IA y de tener la **posibilidad de aplicar dicho conocimiento en áreas laborales a futuro.**



Gracias por la atención

Contacto:

andrea.velarde.chavez@alumnos.upm.es

