

Seguimiento de objetos en video mediante el método de emparejamiento de bloques

Andrés Ely Pat-Chan¹, Francisco Javier Hernández-López²,
Mario Renán Moreno-Sabido¹

¹ Instituto Tecnológico de Mérida,
Tecnológico Nacional de México,
México

² Consejo Nacional de Ciencia y Tecnología,
Centro de Investigación en Matemáticas A. C.,
México

{mg13080859, mario.ms}@merida.tecnm.mx,
fcoj23@cimat.mx

Resumen. El seguimiento de objetos en video (*video object tracking*) ha sido uno de los retos más relevantes en el campo de la visión computacional; actualmente, en la literatura existen cientos de métodos que realizan esta acción. El seguimiento de objetos en video contribuye a muchas aplicaciones que hoy en día son una necesidad en la vida del ser humano, como tratamiento de enfermedades crónicas, robótica, videovigilancia, la industria del entretenimiento, entre otras. En este artículo se presenta la implementación de un algoritmo para el seguimiento de objetos en video usando el método de emparejamiento de bloques. Se muestra el desarrollo de cinco versiones del algoritmo considerando estrategias para determinar si el objeto se está detectando de forma correcta, lo cual ofrece resultados interesantes ante los diferentes entornos y escenarios al que se pudiera enfrentar el algoritmo. Se utilizó un conjunto de datos de veinte secuencias de video para evaluar y comparar cada una de las versiones del algoritmo.

Palabras clave: Método de emparejamiento de bloques, seguimiento de objetos en video, umbral de aceptación.

Video Object Tracking Using Template Matching Method

Abstract. Video object tracking has been one of the most relevant challenges in computer vision; currently, hundreds of methods in the literature perform this action. Video object tracking contributes to many applications that are a necessity in human life today, such as the treatment of chronic diseases, robotics, video surveillance, the entertainment industry, and more. This article presents the implementation of an algorithm for video object tracking using template matching method. The development of five versions of the algorithm is shown considering strategies to determine if the object is being detected correctly, which offers interesting results in the different environments and scenarios that the

algorithm could face. A data set of twenty video sequences was used to evaluate and compare each version of the algorithm.

Keywords: Template matching method, video object tracking, acceptance threshold.

1. Introducción

El seguimiento de objetos en video ha sido uno de los retos más importantes en el campo de la visión computacional. A lo largo de los años, se ha trabajado para proponer diversos métodos que realizan esta acción. El seguimiento de objetos en video consiste en la acción de señalar por medio de diversos recursos visuales, tales como líneas, círculos o cuadros de colores, un objeto específico a lo largo de un video de manera automática [1].

Algunas aplicaciones del seguimiento de objetos en video que destacan hoy en día se encuentran en los sistemas inteligentes de videovigilancia, en la industria cinematográfica con las nuevas herramientas de edición de películas, en el campo de la medicina para detectar organismos celulares u otros patógenos, entre muchas más aplicaciones [2].

También en el área de la robótica se puede encontrar la implementación de seguimiento de objetos [3]. Al igual que en la medicina, con un sistema de seguimiento de movimiento para el tratamiento de pacientes con alguna enfermedad crónica [4].

Un problema en sistemas de videovigilancia ocurre cuando un operador debe revisar una cantidad enorme de videos para poder detectar objetos de interés, además debe seguir o rastrear a dichos objetos durante toda la secuencia de imágenes. Este trabajo es tedioso y conduce a que los operadores dejen pasar situaciones que pudieran ser importantes. Este trabajo se enfocó en la implementación de un algoritmo para contribuir en el desarrollo de sistemas de videovigilancia inteligente.

Actualmente, en la literatura existen un centenar de métodos de seguimiento de objetos, y cada uno realiza esta acción de manera distinta. Uno de los métodos de seguimiento de objetos en video elementales en la visión computacional es el de emparejamiento de bloques (*template matching*), del cual se hablará en este artículo.

El emparejamiento de bloques es un algoritmo, o método, elemental en el mundo del seguimiento de objetos en video; adopta el nombre de emparejamiento de bloques debido a que basta con tomar una plantilla del objeto a seguir para localizarlo en imágenes o fotogramas del video [5].

Algunos trabajos similares que se encuentran en la literatura utilizan el emparejamiento de bloques como fundamento para poner a prueba alguna nueva estrategia que consideren conveniente, con el fin de mejorar la acción de seguimiento de objetos en video [6].

Cabe mencionar que los retos en el seguimiento de objetos en video son interminables, y esa es la esencia y la inspiración para continuar con la investigación sobre el seguimiento de objetos en video. Por cada nuevo método propuesto para este fin, siempre surge alguna inquietud que propicie a seguir mejorando las técnicas y procesos que conllevan a encontrar un objeto a lo largo de un video.

De esta forma, se contribuye cada vez más a la información y, por ende, a las nuevas aplicaciones que estos, de alguna forma, puedan aportar.

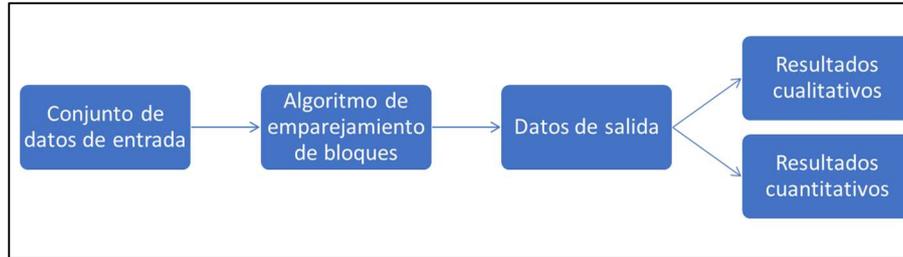


Fig. 1. Diagrama de la metodología propuesta.

El objetivo de este trabajo fue desarrollar un algoritmo para el seguimiento de objetos que automatice el trabajo humano de seguir el objeto a través de un video o una secuencia de imágenes.

En la sección 2 se describe la metodología que se utilizó para el desarrollo de este trabajo. En la sección 3 se muestran los resultados obtenidos en las pruebas. En la sección 4 se presentan las conclusiones a las que se llegaron, así como el trabajo a futuro.

2. Metodología

En la metodología propuesta están involucrados diversos elementos que en conjunto funcionan para lograr un objetivo en común, que en este caso es el de seguir un objeto a través de un video. En la Fig. 1 se observa un diagrama de bloques de la metodología propuesta. El primer paso es el establecimiento del conjunto de datos; actualmente, se cuentan con muchos recursos visuales que pueden servir para tales fines; algunos de ellos se pueden encontrar en internet, o incluso se pueden crear.

El segundo paso se refiere a aplicar un algoritmo de seguimiento de objetos; en el caso de este trabajo consistió en aplicar el método de emparejamiento de bloques. Posteriormente, se obtienen los datos de salida, los cuales se dividen en resultados cualitativos y resultados cuantitativos.

Los resultados cualitativos se obtuvieron al analizar de forma visual las cajas envolventes que devuelve el algoritmo sobre el objeto de interés a través de la secuencia de imágenes.

Los resultados cuantitativos se obtuvieron al comparar las cajas envolventes que devuelve el algoritmo con las cajas envolventes obtenidas de forma manual usando una métrica. A continuación, se describen de manera detallada los pasos de la metodología propuesta.

2.1. Conjunto de datos de entrada

En este trabajo, el recurso visual utilizado fue la secuencia de imágenes del conjunto de datos VOT [7]. Se decidió utilizar este conjunto de datos debido a que es muy conocido y recomendado para trabajos de seguimientos de objetos en video.

Tabla 1. Secuencia de imágenes del conjunto de datos de entrada.

Secuencia	Número de fotogramas	Breve descripción
Ball	603	Pelota en movimiento.
Board	698	Tarjeta de un circuito electrónico en movimiento.
Box	1161	Caja de cartón pequeña en movimiento.
Car	374	Auto saliendo de un estacionamiento.
Car_2	945	Camioneta blanca en calles.
Carchase	9928	Persecución policiaca.
Cup_on_table	1021	Una tasa sobre una mesa con la cámara moviéndose.
Dog1	1390	Peluche de perro en movimiento.
Gym	767	Gimnasta olímpica presentando su número.
Juice	404	Caja de jugo en una mesa con la cámara moviéndose.
Jumping	313	Persona saltando una cuerda.
Lemming	1336	Oso de peluche en movimiento.
Liquor	1741	Botellas en movimiento.
Mountain-bike	228	Motociclista saltando una rampa.
Person	948	Persona en movimiento.
Person_crossing	1018	Persona caminando por un parque.
Person_partially_occluded	306	Persona filmada desde diferentes ángulos.
Singer	351	Cantante presentando su número.
Sylvester	1345	Oso de peluche en movimiento.
Track_running	503	Corredora profesional presentando su número.

Otra ventaja es que cuenta con la información y etiquetas necesarias para trabajar con el desarrollo e implementación de algoritmos y métodos enfocados al seguimiento de objetos en video.

El conjunto de datos para este trabajo consistió en veinte secuencias de imágenes. En la Tabla 1 se presenta una breve descripción de las secuencias de imágenes que conforman el conjunto de datos de entrada.

En las diferentes secuencias de imágenes se pueden encontrar archivos de formatos JPEG y PNG que van desde los tamaños de 320x240 píxeles, hasta 640x480 píxeles. Cada secuencia posee información del objeto a seguir en un archivo de texto, lo cual facilita la evaluación de los algoritmos.

El número de fotogramas varía según la secuencia de imágenes; en este caso, van desde 228 hasta 9928 archivos de imágenes o fotogramas. Ahora bien, ya con los recursos visuales y la información del objeto a seguir, se procede a obtener la plantilla.

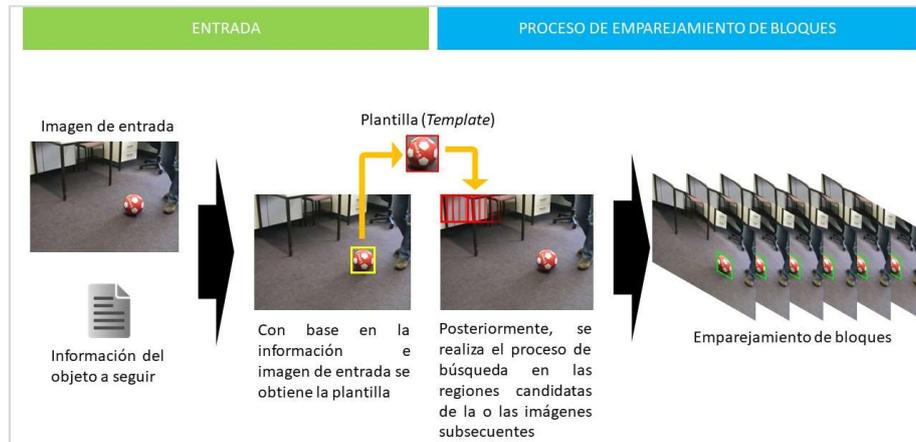


Fig. 2. Proceso general del emparejamiento de bloques.

Como primera instancia, se necesita una imagen de entrada que se obtiene de la secuencia de imágenes; por lo regular, siempre se utiliza el primer fotograma.

De igual forma, se necesita información sobre el objeto a seguir, la cual consiste en coordenadas (x, y) en la imagen, que indican donde se encuentra el objeto y las dimensiones que este abarca (ancho, alto). Posteriormente, con estos recursos se procede a ubicar y señalar en la imagen el objeto en cuestión. Con esa localización se obtiene la plantilla que es una imagen de dimensiones menores a la original, y que servirá como modelo para encontrarlo en los fotogramas subsecuentes.

2.2. Algoritmo de emparejamiento de bloques

Tomando el recurso visual, llámese video o secuencia de imágenes (*frames*), y la información sobre el objeto a seguir, la cual puede estar guardada en un archivo que proporciona el usuario, se puede obtener una plantilla con la cual se inicia el proceso de emparejamiento de bloques.

La plantilla es una imagen del objeto a seguir y está formada por todos los valores de color que tienen los píxeles en esa región seleccionada como plantilla. Con la plantilla establecida, lo que sigue es el proceso de barrido. Se le denomina de esta forma debido a que va recorriendo la imagen hasta encontrar la región que más se parezca a la plantilla [8]. El proceso de barrido lo hace dentro de una región específica de la imagen denominada región de búsqueda.

La región de búsqueda es un área que abarca los alrededores de donde probablemente se encuentre el objeto en la imagen subsecuente. Con esta estrategia se busca reducir el tiempo de procesamiento computacional y aumentar la velocidad de búsqueda. En la Fig. 2 se ejemplifica el método de emparejamiento de bloques de manera general.

Se realiza un cálculo matemático a la que se le denomina métrica, la cual sirve para determinar si la región analizada efectivamente es la más parecida a la plantilla.

Se ejemplificará lo anterior con lo siguiente: suponiendo que se quiere analizar una región de la imagen con la plantilla del objeto a seguir; para ello se necesita “algo” que indique que efectivamente ahí se encuentra dicho objeto [9].

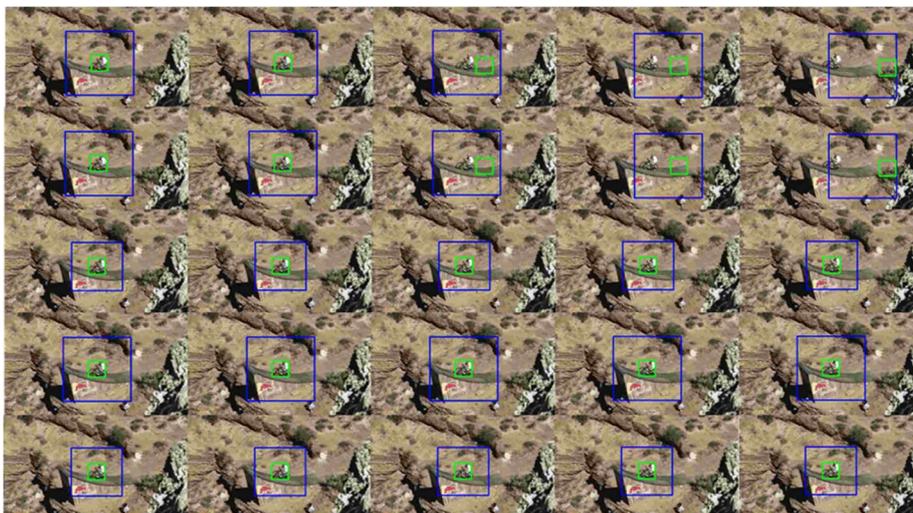


Fig. 3. Resultados de las estrategias procesando la secuencia de imágenes *Mountain-bike*. Cada fila de la figura representa los resultados de las estrategias V1, V2, V3, V4 y V5 respectivamente usando la métrica CCN.

Ese “algo” es un valor que es proporcionado por una de las métricas de las cuales se hablará más adelante en este artículo.

Lo antes descrito brinda una idea general de cómo se realiza el emparejamiento de bloques para encontrar el objeto en las imágenes subsecuentes.

Existen varias métricas para hacer el emparejamiento de bloques, sin embargo, para este trabajo se utilizó la suma de diferencias al cuadrado (SDC) y la correlación cruzada normalizada (CCN).

La SDC es un valor decimal que representa el error de analizar la región candidata contra la plantilla. Mientras más cercano a 0 sea este valor, significa que el error es mínimo y, por ende, en la región candidata analizada se encuentra el objeto.

Cabe mencionar, que de todas las regiones candidatas analizadas, siempre se debe considerar la que ofrezca el valor mínimo.

La SDC se denota de la siguiente forma:

$$SDC = \sum_{i=1}^M (I_{i,v} - T_i)^2, \quad (1)$$

donde M es el número de píxeles en la imagen, I es la imagen de entrada, T la plantilla o modelo del objeto a seguir, i denota un píxel en la imagen, y v denota una ventana del tamaño de T centrada en el píxel i .

La CCN, a diferencia de la SDC, tiene un rango acotado entre $[-1, 1]$, en donde si el valor resultante es cercano a 1, entonces la región analizada es parecida a la plantilla. Esta métrica se denota de la siguiente forma:

$$CCN = \frac{\sum_{i=1}^M (T_i - \bar{T})(I_{i,v} - \bar{I}_v)}{\sqrt{\sum_{i=1}^M (T_i - \bar{T})^2} \sqrt{\sum_{i=1}^M (I_{i,v} - \bar{I}_v)^2}} \quad (2)$$



Fig. 4. Comparación cualitativa de las métricas SDC (primera fila) y CCN (segunda fila) usando la estrategia V5 y procesando la secuencia de imágenes Singer.

donde \bar{I}_v es el promedio de los valores de los píxeles de la imagen en la ventana v y \bar{T} el promedio de los valores de los píxeles de la plantilla.

Con la ayuda de cualquiera de estas dos métricas es como se realiza el emparejamiento de bloques para encontrar la región donde se encuentra el objeto.

Técnicamente, con el paso del barrido se estaría concluyendo el método de emparejamiento de bloques, ya que una vez encontrada la región que más se le parece a la plantilla, el siguiente paso es marcar dicha región en la nueva imagen mediante un cuadro con un color que sobresalga.

Este mismo proceso se repite por cada nuevo fotograma. En este trabajo se realizaron algunas modificaciones que derivaron en cinco versiones (denotadas como V1, V2, V3, V4 y V5) de la implementación del algoritmo; estas versiones se denominaron estrategias y tienen como objetivo principal hacer más robusto el método de emparejamiento de bloques. A continuación, se explican estas estrategias.

2.3. Estrategias usando el método de emparejamiento de bloques

Una inquietud que surge con el método de emparejamiento de bloques (V1) es que siempre marcará una región en donde se considere que está el objeto. Existen casos en donde el objeto ya no está en la escena, o se encuentra ocluido por algún otro objeto; el método, por su naturaleza, siempre señalará en la imagen la región donde considere que está el objeto, aunque en realidad ya no exista en la imagen.

Tomando en cuenta estos casos, se realizaron algunas modificaciones al método que derivaron en versiones que mostraron mejoras significativas.

Una estrategia desarrollada fue la región de confianza (V2), que consiste en tomar los primeros n valores de la métrica que se esté utilizando, almacenarlos en un vector del tamaño de los primeros n valores, y calcular el promedio y desviación estándar para establecer umbrales de aceptación.

Entonces, si el próximo valor a analizar está dentro de estos umbrales, es aceptado y, por ende, se señala el objeto en la imagen en cuestión.

Posteriormente, se desarrolló la estrategia llamada plantilla dinámica (V3), que consiste en evaluar el valor de la métrica; si dicho valor está dentro de los umbrales de aceptación, entonces se señala al objeto en la nueva imagen, y con base en esta información se actualiza la plantilla. Con esto, se busca reducir el problema ocasionado en los casos en donde se tiene una plantilla fija y que no se actualiza a través del tiempo.

Otra estrategia (V4) consistió en combinar las estrategias V2 y V3 en busca de obtener mejores resultados. Por último, se desarrolló una estrategia que se denominó umbral de aceptación (V5), la cual consiste en utilizar las estadísticas de la estrategia

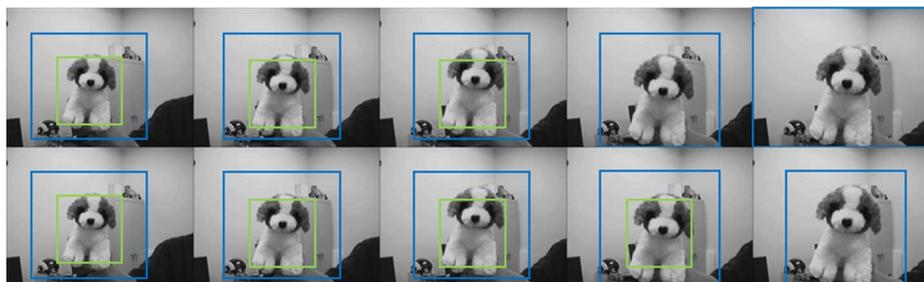


Fig. 5. Demostración cualitativa bajo cambios de escala del objeto a seguir. La primera fila corresponde a la métrica SDC y la segunda a la métrica CCN, usando la estrategia V5 y procesando la secuencia de imágenes Dog1.

V2, pero con la característica de que utiliza solamente valores positivos para crear dos umbrales de aceptación.

El primer umbral sirve para determinar si la región analizada es candidata para señalar la localización del objeto en la imagen. El segundo umbral determina si el objeto encontrado es adecuado para la actualización de la plantilla.

Los datos de salida de las diferentes estrategias, que son recursos visuales como videos con cuadros marcando el objeto, fueron sometidos a pruebas y evaluaciones que ofrecieron resultados interesantes, los cuales se describen a continuación.

3. Resultados

3.1 Resultados cualitativos

Al ejecutar cada una de las estrategias con las secuencias de imágenes se pudieron observar mejorías en cuanto al seguimiento del objeto. También se notó un buen ajuste en ciertas secuencias de video en la que la estrategia base (V1) no podía seguir el objeto.

En la Fig. 3. se observan los resultados de las estrategias procesando la secuencia de imágenes *Mountain-bike* del número de fotograma 15 hasta 19. La fila de la parte superior corresponde a los resultados de la estrategia base (V1), le sigue la fila de la estrategia V2, y así sucesivamente hasta llegar a la estrategia V5. Todos estos usando la métrica CCN. Es evidente como a partir de la estrategia V3, el objeto es seguido con precisión, mientras que en la estrategia V1 hay un desfase considerable.

Las métricas para el emparejamiento (SDC y CCN) jugaron un papel importante para encontrar una mejora a este algoritmo de objetos en video; La métrica CCN ofrece una cierta ventaja sobre la métrica SDC en cuanto a cambios de iluminación. En la Fig. 4 se muestran los resultados de la estrategia V5 procesando la secuencia de imágenes *Singer*, en donde se puede notar que ocurren cambios de iluminación en la escena.

Se observa que al usar la métrica SDC (primera fila de secuencia de imágenes) hay un desfase de la caja que envuelve al objeto, y posteriormente, la caja desaparece perdiendo al objeto. Por otro lado, al usar la métrica CCN (segunda fila de secuencia de imágenes) se muestra un seguimiento más preciso.



Fig. 6. Demostración cualitativa bajo oclusión del objeto a seguir. La primera fila corresponde a la métrica SDC y la segunda a la métrica CCN, usando la estrategia V5 y procesando la secuencia de imágenes *Person_partially_occluded*.

En la Fig. 5 se muestran los resultados de la estrategia V5 al procesar la secuencia de imágenes *Dog1* usando las métricas SDC (primera fila de secuencia de imágenes) y CCN (segunda fila de secuencia de imágenes).

Se observa que el algoritmo pierde al objeto cuando este presenta un cambio de escala, es decir, cuando el objeto está más cerca o más lejos de la cámara.

En la Fig. 6 se muestra una prueba de oclusión usando la estrategia V5, en donde se observa que las estrategias son capaces de encontrar el objeto después de tener una oclusión, siempre y cuando el objeto no cambie su tamaño a través de toda la secuencia de imágenes. La fila superior le corresponde a la métrica SDC, y la inferior a la métrica CCN, ambas de la estrategia V5, donde el algoritmo fue capaz de encontrar el objeto después de sufrir una oclusión.

3.2 Resultados cuantitativos

Para observar de manera cuantitativa la precisión y la robustez de un método de seguimiento de objetos en video ante diversos escenarios se emplean medidas o métricas existentes en la literatura [10].

En este trabajo se empleó la medida de exhaustividad (E) para obtener una ponderación de las diferentes versiones implementadas (derivadas del método original) usando los conjuntos de datos de entrada (secuencia de imágenes).

Para implementar la E se requiere de información verdadera e información estimada. La información verdadera se obtiene del conjunto de datos de entrada en donde se mencionó sobre la información y etiquetas del objeto a seguir; esta información viene incluida en el conjunto de datos de entrada, y son los cuadros o cajas envolventes reales del objeto.

Posteriormente, se necesitan los datos de salida de las diferentes versiones del algoritmo. Esta información es la del objeto que siguió el algoritmo a lo largo del video; normalmente son cuadros o cajas envolventes. Esta información que provee el algoritmo como datos de salida es la información estimada.

La E es la medida de rendimiento para una secuencia; esta medida (a veces denominada porcentaje de seguimiento correcto) indica cuántos cuadros de la salida del algoritmo o rastreador satisface el requisito en la superposición de los cuadros cuando el objeto fue visible; esta superposición se da cuando se comparan los cuadros de la

Tabla 2. Resultados de la E de cada estrategia considerando las métricas SDC|CCN.

Secuencia	V1	V2	V3	V4	V5
Ball	0.06 0.08	0.08 0.09	0.12 0.76	0.12 0.12	0.12 0.12
Board	0.01 0.15	0.17 0.05	0.18 0.16	0.18 0.18	0.18 0.18
Box	0.02 0.05	0.07 0.14	0.82 0.15	0.82 0.82	0.82 0.82
Car	0.17 0.10	0.10 0.18	0.18 0.55	0.18 0.18	0.18 0.18
Car_2	0.02 0.23	0.23 0.29	0.53 0.85	0.53 0.60	0.53 0.60
Carchase	0.04 0.01	0.01 0.04	0.04 0.16	0.04 0.03	0.04 0.03
Cup_on_table	0.02 0.08	0.08 0.13	1.00 0.88	1.00 1.00	1.00 1.00
Dog1	0.03 0.07	0.09 0.11	0.86 0.78	0.86 0.86	0.86 0.86
Gym	0.01 0.49	0.50 0.18	0.36 0.59	0.36 0.36	0.36 0.36
Juice	0.05 0.33	0.33 1.00	0.96 0.48	1.00 1.00	1.00 1.00
Jumping	0.29 0.80	0.80 0.99	0.99 1.00	1.00 0.99	1.00 1.00
Lemming	0.01 0.32	0.33 0.04	0.30 0.50	0.30 0.28	0.30 0.28
Liquor	0.02 0.04	0.04 0.17	0.25 0.53	0.25 0.25	0.35 0.25
Mountain-bike	0.07 0.06	0.06 0.26	0.88 0.07	0.98 0.99	0.99 1.00
Person	0.03 0.06	0.06 0.10	0.55 1.00	0.57 0.57	0.57 0.57
Person_crossing	0.01 0.07	0.07 0.06	0.37 0.82	0.37 0.41	0.37 0.41
Person_partially_occluded	0.07 0.18	0.18 0.53	0.94 1.00	1.00 0.98	1.00 1.00
Singer	0.20 0.20	0.20 0.17	0.46 0.54	0.46 0.47	0.46 0.47
Sylvester	0.01 0.03	0.03 0.05	0.92 0.35	0.94 0.98	0.94 0.98
Track runing	0.05 0.03	0.04 0.28	0.44 0.89	0.44 1.00	0.44 1.00
Promedio	0.06 0.17	0.18 0.24	0.56 0.60	0.57 0.60	0.58 0.61

información verdadera con los cuadros de información estimada. Se denota de la siguiente forma:

$$E = \frac{TP}{TP + FN}, \quad (3)$$

donde TP (*True Positive*) son los verdaderos positivos, y FN (*False Negative*) son los falsos negativos de la superposición.

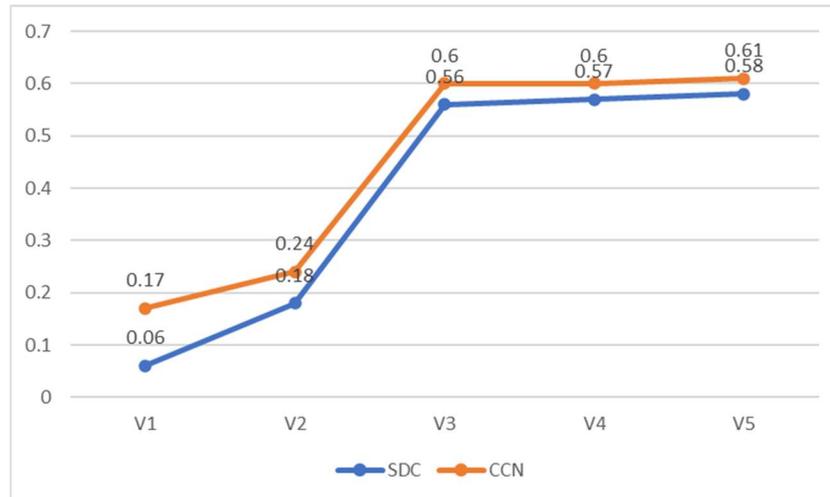


Fig. 7. Gráfica de resultados de la medida E .

Teniendo en cuenta lo anterior, y para tener una medida más objetiva, se decidió utilizar E como referente para mostrar los resultados. En la Tabla 2 se muestran los resultados de la E usando las diferentes estrategias.

En la Fig. 7 se presenta la gráfica que expresa los resultados de la E . En el eje horizontal se tienen las diferentes estrategias; para este trabajo se implementaron cinco.

En el eje vertical se tiene el resultado de la E promedio considerando todas las secuencias de video en cada una de las estrategias; entre más cercano sea este valor a uno, mejor es el método. La línea azul indica la métrica SDC, mientras que la línea naranja se usa para la métrica CCN.

4. Conclusiones y trabajo a futuro

Como se pudo observar en los resultados cuantitativos, se puede concluir que al aplicar la métrica CCN, ésta ofrece una mejora con respecto a la métrica SDC; esto se atribuye a que la primera es más robusta a los cambios de iluminación en comparación con la segunda.

También se concluye que las estrategias implementadas para el caso de la métrica SDC tienden a ser favorables para el seguimiento de objetos debido a que la línea azul apunta hacia arriba.

De manera cualitativa se observó que cuando se utilizó la métrica SDC en algunas secuencias de video con cambios de iluminación, en la escena se perdía el objeto cuando había cambio de iluminación, mientras que cuando se implementó la métrica CCN, éstos se reducían significativamente.

Las estrategias implementadas lograron una mejora en el método de emparejamiento de bloques para el seguimiento de objetos. Además, el uso de la métrica CCN mejoró los resultados.

Las limitantes que se encontraron en este trabajo son, en primera instancia, que se debe definir el objeto a seguir al inicio de la secuencia de video.

Otra limitante se presenta cuando los objetos cambian de escala; en este caso, el método ya no puede seguirlos.

Como trabajo a futuro se desarrollarán métodos de seguimiento de objetos basados en puntos característicos y descriptores de la imagen con el fin de acelerar el proceso de seguimiento y que el método sea robusto a oclusiones.

También se realizarán comparaciones de resultados con otros métodos de seguimiento de objetos, como lo son los basados en aprendizaje automático y puntos característicos.

Referencias

1. Pernici, F., Bimbo, A.: Object tracking by oversampling local features. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 12, pp. 2538–2551 (2014) doi: 10.1109/TPAMI.2013.250
2. Maggio, E., Cavallaro, A.: *Video tracking: Theory and practice*. Wiley (2010) <https://www.wiley.com/en-sg/Video+Tracking:+Theory+and+Practice-p-9780470749647>
3. Petrović, E., Leu, A., Ristić-Durrant, D., Nikolić, V.: Stereo vision-based human tracking for robotic follower. *International Journal of Advanced Robotic Systems*, vol. 10 (2013) doi: doi.org/10.5772/5612
4. Rolland, J. P., Fuchs, H.: Optical versus video see-through head-mounted displays in medical visualization. *Presence: Teleoperators and Virtual Environments*, vol. 9, no. 3, pp. 287–309 (2000) doi: 10.1162/105474600566808
5. Bovik, A.: *The essential guide to video processing*. Academic Press (2009) doi: 10.1016/B978-0-12-374456-2.X0001-1
6. Briechle, K., Hanebeck, U. D.: Template matching using fast normalized cross correlation. In: *Proceedings of SPIE the International Society for Optical Engineering*, vol. 4387, pp. 95–102 (2001) doi: 10.1117/12.421129
7. VOT Visual Object Tracking: VOT challenge (2013) <https://www.votchallenge.net/>
8. Corke, P.: *Robotics, vision and control: Fundamental algorithms in MATLAB*. Springer Cham (2017) doi: 10.1007/978-3-319-54413-7
9. Burger, W., Burge, M.: *Principles of digital image processing: Core algorithms*. Springer (2009) doi: 10.1007/978-1-84800-195-4
10. Padilla, R., Netto, S. L., Silva, E.: A survey on performance metrics for object-detection algorithms. In: *2020 International Conference on Systems, Signals and Image Processing (IWSSIP)*, pp. 237–242 (2020) doi: 10.1109/IWSSIP48289.2020.9145130