Volumen 5 (2019)
XXV Verano de la Ciencia
ISSN 2395-9797

Agrupación de puntos de interes para la detección de objetos en movimiento

Gónzalez Blanco José Salvador (1), Guzmán Valades María José (1), Rodríguez Salazar Jonathan Humberto (1), Ibarra Manzano Mario Alberto (2), Almanza Ojeda Dora Luz (2)

1 [Licenciatura en Ingeniería en Sistemas Computacionales, Universidad de Guanajuato] | Dirección de correo electrónico: {jose.blanco, mj.guzmanvalades, jh.rodriguezsalazar}@ugto.mx
2 [Departamento de Ingeniería Electrónica, División de Ingenierías, Campus Irapuato - Salamanca, Universidad de Guanajuato] | Dirección de correo electrónico: {ibarram, dora.almanza}@ugto.mx

Resumen

La detección y seguimiento de objetos móviles es una tarea indispensable que requerir alta eficiencia computacional y un mínima tiempo de desempeño para lograr la funcionalidad de un sistema autónomo. Un sensor RGBD, como el Kinect©, es un dispositivo que responde a las necesidades de alto desempeño y económico que permite resolver tareas de autonomía en robots de forma eficiente. Este trabajo se enfoca en la detección de objetos móviles para lo cual se configura un sistema de visión RGB-D, para detectar los principales puntos de interés y su correspondiente valor en profundidad en la secuencia adquirida de imágenes. Los puntos detectados en la primera imagen de la secuencia son localizados a lo largo de 5 imágenes consecutivas por medio de flujo óptico, para obtener los datos de velocidad y formar un vector de características de 4 dimensiones. Un modelo probabilista basado en la teoría a contrario permite encontrar los puntos que pertenecen a objetos móviles, analizando su posición y velocidad desde un árbol jerárquico generados con los puntos característicos. Los resultados experimentales muestran la separación de los objetos detectados del resto de la escena.

Abstract

The mobile objects detection and tracking is an indispensable task that requires high computational efficiency and the minimum performance time to achieve the functionality of an autonomous system. A RGBD sensor, like Kinect©, is a device that responds to the needs of high performance and economic cost solving autonomy tasks performance in robots efficiently. This work focuses on the detection of mobile objects for which an RGB-D vision system is configured, based on the interesting points detection and their corresponding depth value in the image sequence. The points detected in the first image of the sequence are located along 5 consecutive images using optical flow, then the velocity data is obtained to form a vector of characteristics in 4D. A probabilistic model based on the a contrario theory allows to find the points that belong to moving objects, analyzing their position and velocity from a hierarchical tree of characteristic points. Experimental results show the detected objects highlighted from the rest of the scene.

Palabras Clave
Palabras Clave: agrupamiento; Kinect ONE; flujo óptico; detección de objetos.

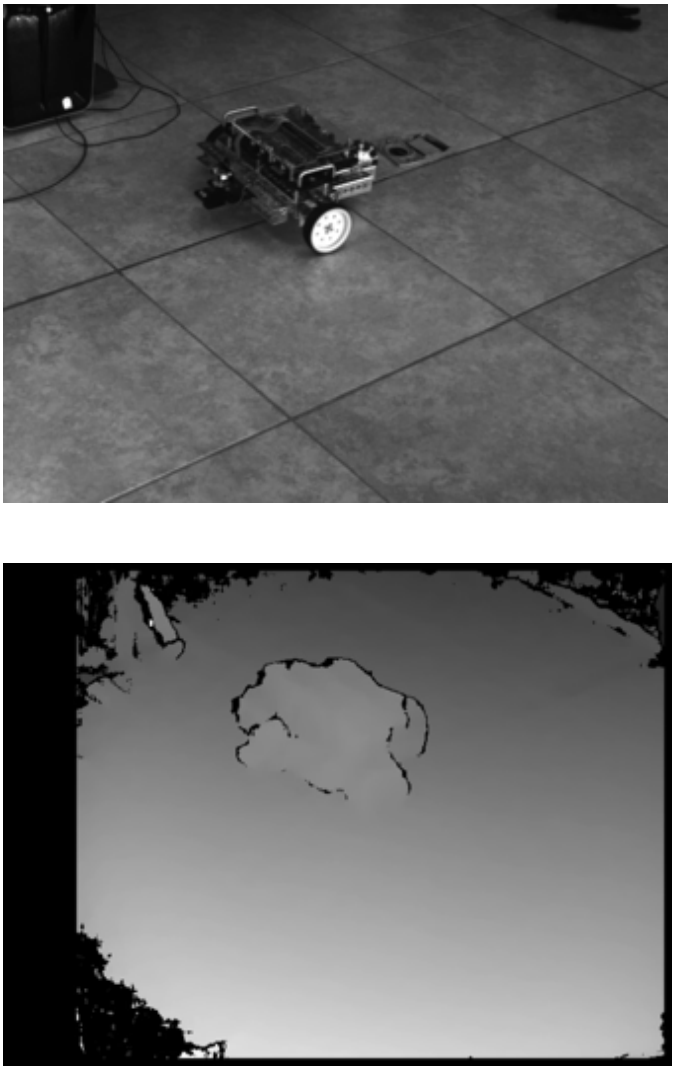

IMAGEN 1: Capturas con el kinect ONE, imagen de color RGB (izquierda), imagen de profundidad (derecha). profundidad

En las imágenes color se realiza una técnica de detección de puntos de interés basado en el seguidor de puntos KLT [5]. Esta técnica calcula el gradiente en x y y de la imagen en el tiempo t, esto es, en I(t) y mediante un análisis de los eigenvalores obtiene los N puntos más sobresalientes en la imagen. Una vez obtenidos los puntos en la imagen se realiza un seguimiento de ellos en la siguiente imagen I(t+1) de la secuencia, garantizando que el punto encontrado es el mismo que el previo en I(t). Se hace uso de las posiciones de los puntos en I(t) para inicializar y calcular el desplazamiento del punto en I(t+1). La Imagen 2 muestra un ejemplo de los puntos detectados y su seguimiento durante 4 imágenes consecutivas para formar nuestro vector característico. En la primera imagen de la figura en el tiempo I(t) se obtienen primeros N puntos de interés, que en nuestro caso fue indicado como N=150 puntos. Estos puntos son mostrados en la segunda imagen, donde, los puntos azules, son los puntos que fueron encontrados de I(t) hacia I(t+1). Los puntos rojos representan los puntos que estaban en I(t) pero que ya no fueron encontrados en I(t+1). Cuando un punto ya no es encontrado, se reemplaza por uno nuevo en la imagen I(t+1) el cual no se muestra en la imagen actual sino hasta la siguiente imagen I(t+2) en color verde. En la imagen I(t+2) se muestra en azul los puntos que siguen siendo encontrados desde I(t), en rojo los que se van perdiendo y en verde los que siguen siendo encontrados pero que no estaban desde el principio en I(t). La imagen I(t+3) es la última imagen que sigue procesando ese conjunto de puntos desde I(t) y al encontrarlos, es posible generar el vector de características espacio-temporal de los puntos. Esto es, dado el conjunto de puntos de interés entregado por la técnica KLT, se genera un grupo de puntos llamado V(x,y,vx,vy,t) de longitud N.
Los valores de x y y corresponden a las coordenadas en pixeles de la imagen y los valores vx, vy son los valores de velocidad, respectivamente, para cada coordenada obtenidos como se explicó en el párrafo anterior. Entonces, es posible almacenar un conjunto de puntos a lo largo de 4 imágenes consecutivas que representen la "estela" de movimiento de los puntos característicos en la secuencia, es decir las posiciones de los puntos desde I(t) hasta I(t+3). Este vector es agrupado en un árbol de datos, utilizando la técnica de enlace simple, la cual se explica a continuación.

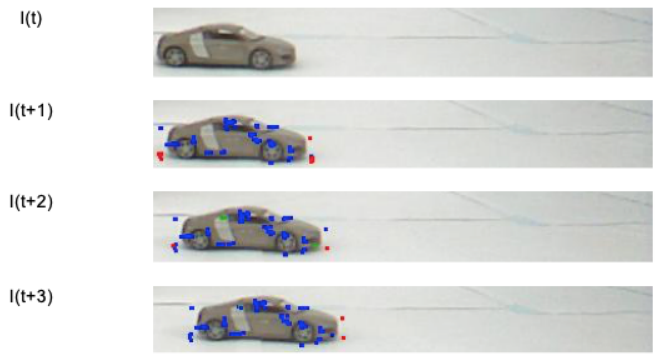

IMAGEN 2: Detección y seguimiento de puntos de interés en 4 imágenes consecutivas utilizando KLT

Métrica de enlace y formación del árbol de datos

En este trabajo es necesario utilizar una métrica de enlace para conocer la similitud entre el conjunto de características obtenido V(x,y,vx,vy,t) que consiste de la posición y velocidad de los puntos. Al aplicar esta métrica se obtendrá un árbol jerárquico ordenado. El correcto ordenamiento de este árbol es esencial para lograr un alto desempeño en el método de agrupamiento. En el estado del arte podemos encontrar diferentes métricas de enlace, como son: el enlace simple, el enlace doble, el enlace promedio, el enlace centrado, entre otros. En nuestro caso, se utiliza la técnica de enlace simple para generar nuestro árbol jerárquico de los puntos. La IMAGEN 3 muestra un ejemplo sintetizado de cómo se genera un árbol jerárquico y que representa. El grupo más arriba en el árbol GXXX se le conoce como el nodo raíz del árbol. Este nivel representa todos los puntos del conjunto a analizar y las tres posiciones indicadas por las X indican el número de niveles de profundidad del árbol, que son 3 para este ejemplo. En el segundo nivel, la raíz se divide en dos conjuntos, aquí indicados con el subíndice G1XX y G2XX. El número 1 y 2, solo indican el número de grupo del nivel y al colocarse en la primera posición del subíndice, indica que son los grupos del primer nivel. Es decir, las siguientes dos posiciones son indicadas con XX aun porque solo se ha explorado el nivel 1 correspondiente a la primera posición del subíndice. Continuando nuestro recorrido en el árbol de la imagen, el siguiente nivel tiene 4 grupos, generados de dividir cada grupo anterior en dos. En este nivel, la segunda posición del subíndice indica el número de grupo del nivel y la primera posición indica el grupo del nivel anterior del cual proviene. Finalmente, el tercer nivel vuelve a dividir en dos cada grupo del nivel anterior, para en este caso generar 8 grupos. Note que, la tercera posición de los sub-índices del grupo indican, nuevamente, el número de grupo por nivel y las dos posiciones anteriores, los grupos de los cuales provienen en los niveles anteriores.

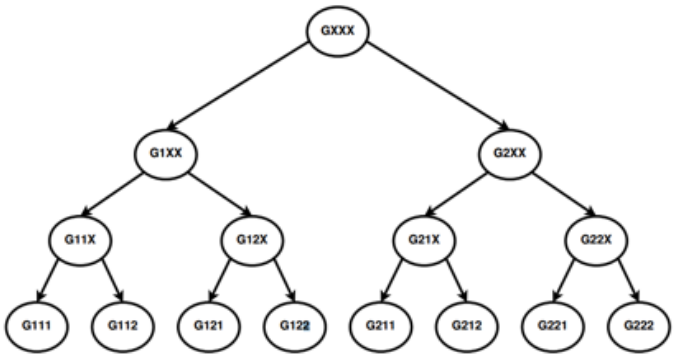

IMAGEN 2: Detección y seguimiento de puntos de interés en 4 imágenes consecutivas utilizando KLT

El vector de características V(x,y,vx,vy,t) forma un árbol con 2N-1 grupos y N-1 niveles, por lo que, para el caso de N=150, se formarán 299 grupos y 149 niveles. Cada nivel del árbol representa los grupos de puntos con las características más similares posibles, de tal manera que el siguiente nivel después de la raíz, logra separar los puntos más disimilares. Aún dentro de cada grupo, pueden existir puntos con grandes diferencias, por ello debe descenderse a lo largo de los niveles del árbol hasta encontrar un equilibrio entre el número de grupos, la separabilidad entre ellos en el mismo nivel y su valor de similitud de los puntos característicos al interior del grupo. Con el fin de realizar esta validación se utiliza el método de agrupamiento basado en la teoría a contrario, la cual permitirá entregar los “mejores” grupos, esto es los grupos que más se alejan del modelo propuesto, por ello se llama a contrario.

Método de agrupamiento

La técnica de agrupamiento utilizada en este trabajo es el método A contrario [6][7], basada en la teoría de Gestalt, la cual es utilizada para formar grupos evaluando una o varias características de sus elementos. La técnica propone un modelo de movimiento del fondo, el cual es aleatorio, de tal manera que, si encontramos algún grupo de punto que coincida con el modelo de movimiento de fondo propuesto, será un objeto móvil que pertenezca al fondo. Entonces, la técnica se llama a contrario por que se busca que los grupos en el árbol jerárquico no sigan este modelo de fondo y buscar solo los que indiquen lo contrario. De acuerdo con esto, la técnica a contrario evalúa cada grupo del árbol binario para encontrar cual contradice al modelo de fondo aleatorio propuesto. El modelo de movimiento aleatorio representa

movimientos como las hojas de los árboles que se mueven, pero solo en un vaivén o de forma aleatoria. La evaluación consiste en calcular el valor de métrica llamada Número de Falsas Alarmas, indicada en la ecuación 1 y cumplir con la condición de que ese valor sea NFA ≤1 [7].

$$NFA(G) = N^2|H| \min_{\substack{X \in G \\ H_X \in H \\ G \subset H_X}} B(N-1, n-1, p(H_X)) \qquad (1)$$

donde N representa el tamaño de V, |H| es la cardinalidad de las regiones y n es el número de elementos en cada grupo de prueba G. El término B(.) en la ecuación 1 representa la ley acumulada binomial e indica la probabilidad de que al menos n puntos de G, incluyendo X, estén dentro de las diferentes regiones de prueba indicadas en H. Las regiones de prueba H son zonas de interés de diferente tamaño que indican el "rango" de los datos y la cercania con otros grupos. Así, mientras más pequeña sea la zona que contiene a todo el grupo de puntos en G, la probabilidad Binomial de ese grupo es muy baja, ya que implica que se evalúe la en un tamaño de region pequeña, por lo que se entregará un valor de NFA bajo. Debido a que la condicion para validar los grupos está indicado por NFA ≤1, entonces, esto permite elegir ese grupo como candidato a no seguir el modelo propuesto de fondo con velocidad y posición aleatoria. Esta métrica NFA indica todos los grupos que son significativos, los cuales pueden llegar a ser muchos, por lo que es posible realizar una segunda evaluación y "podar" los grupos aún no tan representativos.

La técnica a contrario, arroja los grupos de puntos más significativos, que son los que representan las caracteristicas más unidas. Sin embargo, una última validación es realizada con los grupos de puntos entregados como significativos o representativos, que es el análisis del valor de profundidad del grupo. Para realizar esta prueba, unicamente se compara que el promedio del valor de intensidad de los puntos del grupo sea similar, o no difiera por más de 5 cm, que fue tomado como el cambio minimo que podriamos detectar en el fondo por un sensor Kinect. Cabe mencionar que la sensibilidad del sensor kinect es menor de 5 cm para detectar un cambio en el nivel de gris, pero decide dejarse en 5 para garantizar que solo los grupos que pertenezcan a objetos muy alejados entre sí, serán separados. La IMAGEN 4 muestra la salida del clasificador para la secuencia de imágenes presentada en la IMAGEN 2, esto es cuando ya se formo el cluster de salida.

IMAGEN 4: En azul claro los grupos entregados por el clasificador desde una camara RGB-D

Los puntos en azul más claro en la IMAGEN 4 ya fueron validados por la técnica a contrario e indica que son los grupos que no siguen el modelo de fondo con forma adyacente y además su valor de profundidad no varía demasiado. Antes de presentar los resultados y discusión de la técnica, en la siguiente subsección se describirá brevemente las librerías utilizadas para programar los módulos y todo este sistema de detección de movimiento en escenas dinámicas.

Resultados Y Discusión

Las pruebas experimentales fueron realizadas en el Laboratorio de Procesamiento Digital de Señales conectando un sensor Kinect ONE a 2 m del objeto en movimiento, varios objetos fueron utilizados, pero en este caso se presentan los resultados obtenidos con un carrito sobre la escena. El carrito fue acondicionado para moverse en línea recta a velocidad constante al pasar de forma ortogonal frente al campo de visión del Kinect. Las diferentes pruebas realizadas nos permitieron encontrar varios grupos de puntos móviles a lo largo de la secuencia y que solo pertenecían al objeto móvil. La IMAGEN 5 muestra en un recuadro color cyan la región de interés que envuelve al grupo de puntos móviles encontrados en la escena después de procesar 4 imágenes de forma consecutiva. Note que la foto 3 a la 5 de la IMAGEN 5, contienen dos recuadros, esto se debe a que en ocasiones el método no le es posible fusionar por similitud un par de objetos agrupados por la técnica a contrario. Lo que sucede de frecuentemente es que los centroides de los puntos no se encuentran muy cercanos entre sí y es por ello que la técnica los maneja como grupos separados a pesar de que son los mismos.

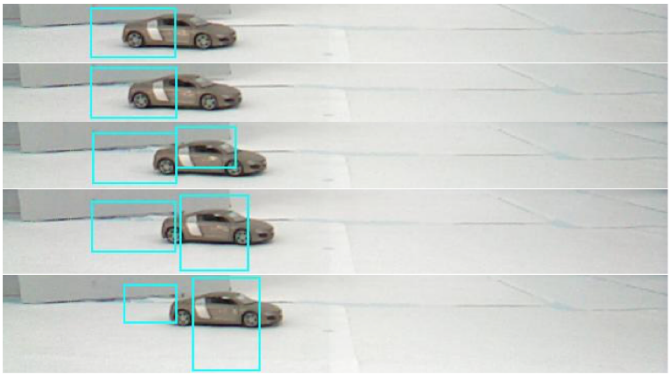

IMAGEN 4: Los recuadros muestran los grupos entregados por el agrupamiento de puntos de intéres.

CONCLUSIONES

En este trabajo se realizó la detección de puntos de interés en una imagen y se acumuló su posición y velocidad a lo largo de 4 imágenes consecutivas. Estos puntos característicos fueron suficientes para formar un árbol jerárquico y encontrar los grupos de puntos que mejor contradicen a un modelo aleatorio de fondo establecido. Los resultados de agrupamiento son bastante buenos, esto es, en los experimentos no hay omisión de algún objeto móvil que entre a la escena, sin embargo, una de las mejoras que pretendemos realizar a esta técnica es agilizar su tiempo de computo, ya que se requieren muchas operaciones y demanda de recursos en la PC para evaluar la técnica a contrario y entregar resultados.

AGRADECIMIENTOS

Los autores expresan su agradecimiento al Departamento de Ing. Electrónica de la División de Ingenierías del Campus Irapuato Salamanca por el espacio asignado para llevar a cabo este proyecto.

REFERENCIAS

[1] Baig, Q., Perrollaz, M., Laugier, C. (2014). A Robust Motion Detection Technique for Dynamic Environment Monitoring: A Framework for Grid-Based Monitoring of the Dynamic Environment. Robotics Automation Magazine, IEEE, vol. 21, pp. 40–48.

[2] Meisener J. (2013). Collaboration, expertise produce enhanced sensing in Xbox One. The Official Microsoft Blog. News & Perspectives. 22 de Julio de 2019.Recuperado de https://-blogs.technet.microsoft.com/microsoft_blog/2013/10/02/collaboration-expertise-produce-enhanced-sensing-in-xbox-one/

[3] Murillo A., Marcal M. (2013). Características Kinect 2. Kinect for Developers. 24 de Julio de 2019. Recuperado de http://www.kinectfordevelopers.com/es/2014/01/28/caracteristicas-kinect-2/

[4] Rodríguez Salazar, J.H., Almanza-Ojeda D.L. (2018). Clasificación de actividades por sistemas autónomos. Revista de divulgación científica: Jóvenes en la ciencia, vol. 4(1), pp. 1-5. ISSN: 2395-9797.

[5] Shi, J., Tomasi, C. (1994). Good Features to Track. IEEE Conference on Computer Vision and Pattern Recognition, pp. 593–600. IEEE Press.

[6] Almanza-Ojeda D.L., Ibarra-Manzano M.A. (2011). 3D visual information for dynamic object detection and tracking during robot mobile navigation”,
Recent Advances in Mobile Robotics, Dr. Andon Venelinov Topalov, (Ed.), ISBN-978-953-307-909-7, cap. 1, pp. 3-24, InTech.

[7] Cao, F., Delon, J., Desolneux, A., Muse, P., Sur, F.(2007). A Unified Framework for Detecting Groups and Application to Shape Recognition. Journal
of Mathematical Imaging and Vision, vol. 27, pp. 91–119.