La fusión de imágenes a nivel características para la detección y seguimiento de objetos

Janeth Cruz Enriquez, Leopoldo Altamirano Robles

[1] INAOE, Departamento de Ciencias Computacionales, Luis Enrique Erro No. 1, Sta. Ma. Tonantzintla, Puebla, 72840, México
jcruze@ccc.inaoep.mx, robles@inaoep.mx

Resumen. En este trabajo se presenta un algoritmo para la detección y seguimiento de un objeto mediante la fusión de datos proporcionados por dos sensores pasivos, un sensor infrarrojo y un sensor visible. La fusión se realiza a nivel características, las cuales se extraen de las imágenes de cada sensor para formar una imagen fusionada con las regiones resultantes que representan a los posibles objetos de interés sobre la escena. Las tres características locales seleccionadas se basan en la diferencia de contraste existente entre los objetos de interés y el fondo. El seguimiento del objeto en movimiento se mantiene mediante la correlación de un patrón de la imagen de muestra sobre los posibles objetos usando una ventana del tamaño del objeto, la cual se obtiene mediante la extracción adaptiva del fondo.

1 Introducción

Actualmente existe un gran número de algoritmos de detección y seguimiento de objetos. La mayoría de estos realizan la detección con los datos que provienen de un solo sensor. El uso de un solo sensor ocasiona desventajas que reducen la eficiencia del sistema, entre ellas: blancos falsos, susceptibilidad a contramedidas, incertidumbre que puede existir en los datos, deficiencia en cambios de escenarios, etc.

Para solucionar esta problemática se presentaron resultados a principios de los 80's con la creación de un nuevo modelo conocido como fusión de datos multisensor, el cual se relaciona con la asociación, correlación y combinación de datos e información de múltiples fuentes de información para obtener posiciones exactas y estimaciones de identidad, así como una completa valoración de situaciones y riesgos [4].

La fusión surge con la necesidad de llevar a cabo entre otras cosas la detección, identificación y seguimiento de blancos aéreos en el campo militar [9]. Muchas son las ventajas con el uso de múltiples sensores, donde cada sensor puede ser usado para complementar los datos de otros sensores, proveer una cobertura mayor y eficiente en la estimación del estado de blancos y decisiones [1]. Así como en el monitoreo cooperativo de blancos que no pueden ser visualizados por un simple sensor [2], [7].

A. Gelbukh, M. Hernández Cruz (Eds.) Avances en la ciencia de la computación en México, CORE-2003, pp. 209–219, 2003.

La desventaja de usar un sensor para realizar el seguimiento de objetos con una operatividad diurna-nocturna y la eliminación de blancos falsos fueron parte de motivación de este trabajo, en el cual se plantea la fusión adaptiva de los datos obtenidos de diferentes sensores, un sensor visible y un sensor infrarrojo, en la detección y seguimiento de objetos. El presente trabajo forma parte de un proyecto mas amplio, cuyo objetivo es la fusión adaptiva de sensores bajo una arquitectura fusión distribuida en aplicaciones de seguimiento de objetos.

Los sensores usados en este trabajo operan en diferentes rangos del espectro electromagnético pero ambos son sensores pasivos, es decir, estos sensores sólo reciben la radiación emitida por los objetos, a diferencia de los sensores activos, como radar que envía una señal y la recibe después de ser reflejada por el objeto. Un sensor visible capta la cantidad de luz que emiten o reflejan los objetos en la escena dentro del espectro visible y un sensor infrarrojo representa en una imagen la cantidad de calor que emiten los objetos en la escena.

Diversos trabajos se han realizado para la detección y seguimiento de objetos con estos sensores de forma individual. Sin embargo, los algoritmos de detección y guimiento de objetos con sensores infrarrojos deben ser más robustos al ruido comparados con los sensores visibles, no obstante, los sensores infrarrojos pueden trabajar sobre escenarios donde no existe iluminación, lo cual no es posible con un sensor visible.

La fusión de estos sensores también se ha explorado a nivel características, donde estas son seleccionadas dependiendo del contexto de trabajo y los objetos interés, como las basadas en textura [6], contraste [7], color y forma. Sin embargo, presentan diversos inconvenientes, entre otros la cantidad de características a fusionar o en la selección de la función adaptiva de ponderación, la cual determina valor de cada característica en el proceso de fusión.

El propósito de este trabajo es fusionar la información que proviene de los dos sensores seleccionados, para generar una imagen fusionada que contenga únicamente regiones que satisfagan las características del objeto de interés. La detección basa en la fusión de tres características locales (diferencia de medias, gradiente promedio, varianza local) de las imágenes que provienen de ambos sensores, usando una ventana que corresponde al tamaño de los objetos de interés, el tamaño de esta ventana se determina usando un algoritmo robusto de extracción adaptiva del fondo. Después del proceso de detección donde se obtienen los posibles blancos en la imagen, es necesario efectuar la correlación existente entre los posibles blancos detectados y un patrón de entrada para mantener el seguimiento del objeto en movimiento.

2 Fusión multisensor

Existen tres niveles de fusión de datos de imágenes [4]: *Nivel pixel, nivel características* y *nivel decisión.* En la Figura 1 se observa como los niveles dependen de forma en la que se encuentran los datos del sensor en el momento de realizar fusión, es decir, el nivel es usado para describir el punto durante el procesamiento

de imágenes en el cual ocurre la combinación de los datos que provienen de diferentes sensores.

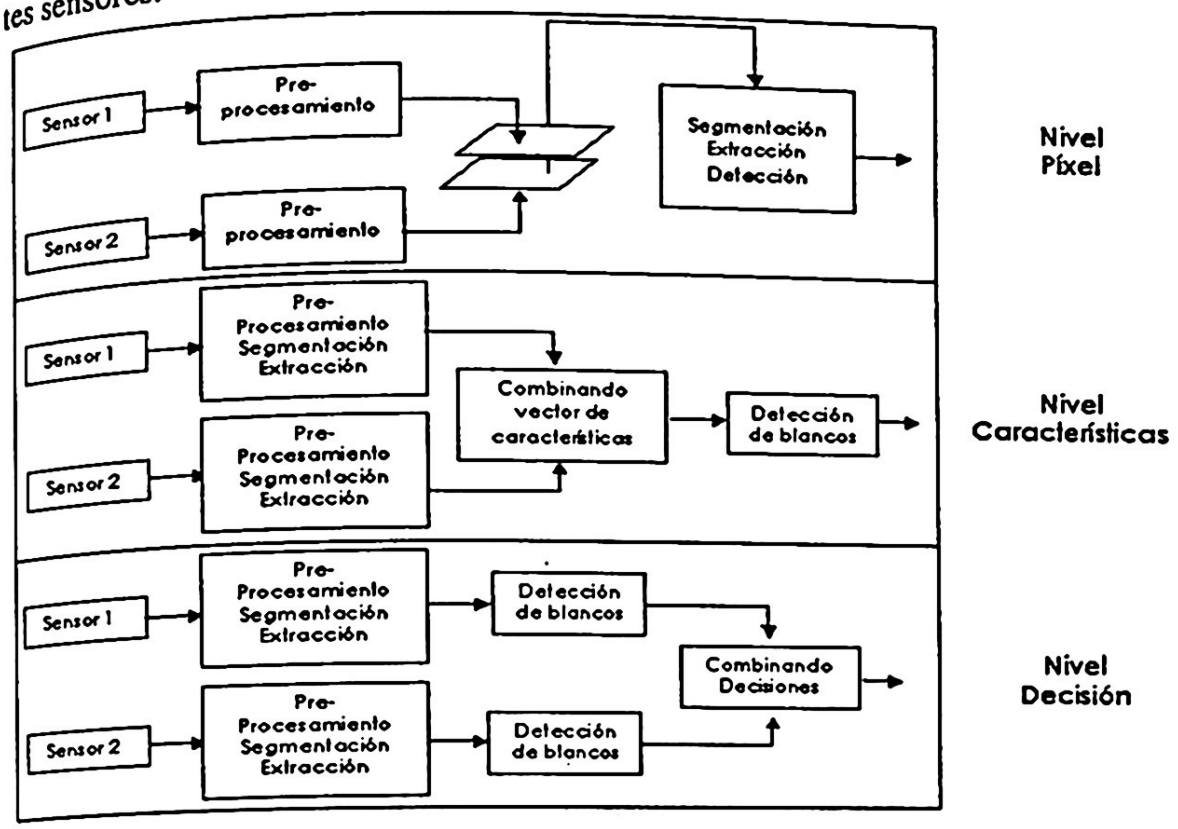

Fig. 1 Niveles de fusión, en la figura se observan los módulos del procesamiento de imágenes y sobre cuales se aplican los tres diferentes niveles.

El nivel mas bajo de fusión es a *nivel pixel* que consiste en la combinación de los pixeles registrados de todo el conjunto de sensores para después ejecutar diferentes funciones de detección y discriminación de objetos. El segundo nivel, *fusión a nivel características* combina las características de los objetos (dimensiones, área, textura, contraste, etc.) que se extraen de los datos que provienen del sensor antes de ser combinados.

El nivel mas alto, *fusión a nivel decisión* consiste en fusionar las decisiones propuestas en cada dato de entrada obteniendo como resultado una decisión. En este trabajo se muestran los resultados obtenidos de la fusión a nivel características para la detección de los objetos en movimiento.

3 Fusión adaptiva distribuida

El propósito de fusionar la información de diferentes sensores es complementar la información que proporciona un solo sensor, sin embargo, una problemática a resolver cuando se integran varios sensores es evaluar la información relevante que proporciona cada sensor. Aunque actualmente ya se tienen trabajos sobre la fusión de sensores, pocos son los que se basan en fusión adaptiva, lo cual permite tener un campo de investigación a explorar y proponer técnicas adaptivas de fusión usando

los sensores seleccionados. Además, parte importante para la comunicación de sensores es la arquitectura seleccionada, donde la mayoría de los sistemas de fusión sensorial usan una arquitectura centralizada que ocasiona la dependencia de información que generan los sensores, por lo que la tendencia en el diseño de estos sistemas es crear arquitecturas descentralizadas que permitan la funcionalidad sistema aun cuando un sensor presente problemas en su funcionamiento.

El modelo de los elementos básicos de un sensor en la detección y seguimiento de múltiples objetos requiere un proceso de asociación de datos el cual se define como asociación local y se muestra en la Figura 2, pero además cuando se incrementa el número de sensores se requiere de una asociación global para determinar que información proporcionada por diferentes sensores pertenece al mismo objeto físico.

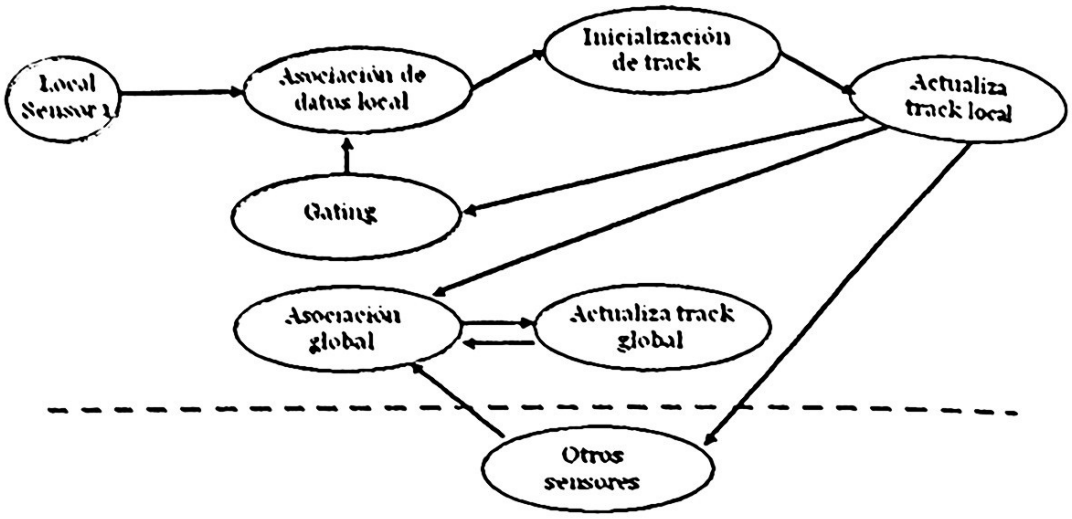

Fig. 2 Fusión distribuida, esta Figura muestra el modelo básico de un sistema de asociación global junto con la etapa de fusión de diferentes sensores.

En base al estudio del estado del arte se propone un algoritmo de fusión adaptiva que permita tener un seguimiento robusto a los cambios que sufra el objeto en forma y tamaño a través del tiempo, así mismo la importancia de realizar una fusión distribuida ha sido poco explorada por su complejidad pero proporciona ventajas el manejo de recursos del sistema. El objetivo es tener un algoritmo de seguimiento de múltiples objetos para cada sensor y después combinar de manera distribuida información de los objetos detectados.

4 Detección y seguimiento de objetos

La primer pregunta que se debe contestar en la selección de las características para la fusión es "¿Qué es importante en la imagen y como puede ser aislado?" [6], esto significa que es necesario determinar el tipo de objetos a detectar y como se representan en los sensores seleccionados. Por ejemplo, en aplicaciones militares, gene-

primero los objetos que son de interés deberán ser mas calientes o fríos que el fondo, cuando se usan sensores infrarrojos.

El sensor infrarrojo usado en este trabajo representa el nivel de calor de un objeto con niveles de gris, a medida que aumenta el nivel de gris significa que el cuerpo esta emitiendo mayor cantidad de calor, es decir, un objeto caliente se representará con un nivel de gris cercano al 255, sin embargo, un objeto frío tendrá un nivel de gris cercano al 0.

En este trabajo como ya se mencionó, se han digitalizado secuencias con dos sensores (infrarrojo y visible), donde cada una de las características se obtienen localmente usando dos ventanas, una interna y una externa. El tamaño de la ventana más pequeña es igual al objeto de interés.

4.1 Selección de características y tamaño de la ventana

Existen diferentes características que nos permiten identificar a un objeto. En [5] usan cuatro características locales: máximo local, diferencia de medias, gradiente promedio y variación local; por medio de estas características detectan automóviles fácilmente con secuencias obtenidas por dos sensores, infrarrojo y visible.

Estas características basadas en la diferencia de contraste también serán usadas para la detección de objetos en este trabajo. Sin embargo, solo se usarán 3 características: diferencia de media, gradiente promedio y variación local. Esto se debe a que el máximo local no proporciona información relevante que discrimine al objeto del fondo, en las secuencias obtenidas.

El tamaño de la ventana debe contener al objeto de interés ya que es de suma importancia obtener resultados confiables en la fusión de sensores y seguimiento [8]. Una técnica de segmentación es usada para determinar la ventana asociada al objeto de interés; la cual realiza un procesamiento temporal, basado en la representación multi-normal a nivel del pixel. Este método usado en [7] identifica a los pixeles del fondo en cada cuadro nuevo mientras actualiza el modelo de cada pixel.

Los pixeles etiquetados como fondo pueden integrarse como parte de los objetos usando un algoritmo de componentes conectados. Este algoritmo proporciona el tamaño de la ventana interior. El tamaño de la ventana exterior se determina de acuerdo al tamaño aproximado del objeto de interés [5].

4.1.1 Diferencia de Medias

La diferencia entre medias es una medida del nivel de gris promedio de una región del tamaño del objeto de interés (ventana interior), comparado con el nivel de gris de los píxeles vecinos que forman parte del fondo (ventana exterior).

Esta característica es ampliamente usada para detectar píxeles que son más calientes que el fondo y que no están contenidos en la ventana interior. La diferencia de medias detecta las regiones locales en donde los cambios de intensidad son abruptos, ésta se calcula por medio de las ecuaciones (1), (2) y (3) para cada pixel de la imagen:

$$C_{i,j}^{1} = \mu_1(i,j) - \mu_2(i,j)$$

$$\mu_1(i,j) = \frac{1}{n_1} \sum_{(k,l)\in N_1(i,j)} f(k,l)$$

$$\mu_2(i,j) = \frac{1}{n_2} \sum_{(k,l)\in N_2(i,j)} f(k,l)$$

Donde n_2 es el número de píxeles en $N_2(i,j)$, n_1 es el número de píxeles en la ventana interior que contiene a los objetos de interés. $N_2(i,j)$ contiene a todos los píxeles vecinos de la ventana centrada en el píxel (i,j), excepto los píxeles que son parte la ventana interior, como se muestra en la Figura 3.

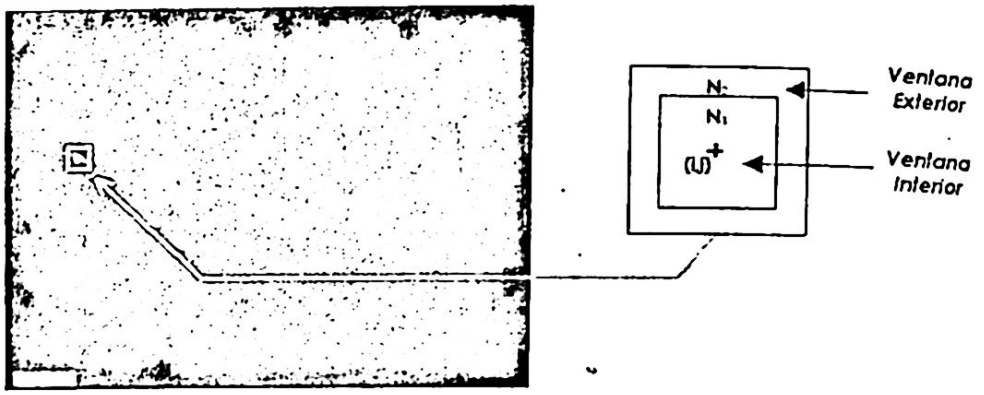

Fig. 3 Ventana interior y exterior, sobre las cuales se obtienen las tres características a fusionar.

4.1.2 Gradiente promedio

El gradiente promedio es una característica que detecta las variaciones de intensidad internas del objeto. El gradiente promedio se calcula con la siguiente ecuación:

$$C_{i,j}^{2} = \frac{1}{n_1} \sum_{(k,l)\in N_1(i,j)} G_1(k,l) - \frac{1}{n_2} \sum_{(k,l)\in N_2(i,j)} G_2(k,l) \tag{4}$$

La ecuación 4 es la diferencia entre el gradiente de la ventana interior y el de la ventana exterior. Los gradientes para cada ventana G_i se obtienen mediante la ecuación 5 que representa la diferencia entre los gradientes horizontales G_i^h y gradientes verticales G_i^v de la ventana.

$$G_i(k,l) = G_i^h(k,l) - G_i^v(k,l) \qquad i=1,2 \tag{5}$$

$$G_i^h(k,l) = \left| f(k,l) - f(k,l+1) \right| \tag{6}$$

$$G_i^v(k,l) = \left| f(k,l) - f(k+1,l) \right| \tag{7}$$

donde i=1 corresponde a la ventana interior y i=2, corresponde a la ventana exterior.

4.1.3 Variación local

La variación local es usada para detectar pequeños cambios en la variación de intensidad sobre las regiones locales. La variación local se obtiene de la siguiente manera:

$$C_{i,j}^3 = \frac{1}{n_1} \sum_{(k,l)\in N_1(i,j)} V_1(k,l) - \frac{1}{n_2} \sum_{(k,l)\in N_2(i,j)} V_2(k,l) \tag{8}$$

donde

$$V_1 = \left| f(k,l) - \mu_1(i,j) \right| \tag{9}$$

$$V_2 = \left| f(k,l) - \mu_2(i,j) \right| \tag{10}$$

4.2 Normalización de características

Después de obtener las tres características para cada pixel se deben normalizar de acuerdo a la imagen correspondiente, generando tres imágenes características. De esta manera el valor del pixel representa el número de desviaciones estándar que el pixel tiene de acuerdo a la imagen.

$$C_{i,j}^{m,N} = \frac{C_{i,j}^m - \mu_m}{\sigma_m} \quad \text{m=1,2,3} \tag{11}$$

Asimismo, se realiza la normalización (N) para cada característica (m), la media y la desviación estándar de cada característica se calculan sobre todos los píxeles de la imagen.

4.3 Fusión multisensor

Una vez normalizadas las características, el siguiente paso es combinarlas. Las características se fusionan de tal forma que cada una debe ser valorada o pesada de acuerdo a la información que proporciona para la detección del objeto, es decir, la característica de mayor peso es la más relevante para el tipo de objetos a detectar.

El peso de la característica se mide como el promedio de los N_t valores más altos de la característica, donde N_t representa el número de pixeles en la región del tamaño del objeto de interés. Este proceso enfatiza las regiones de interés mientras que resto se suprimen. La ecuación 12 obtiene la imagen multisensor combinando las características para ambos sensores.

$$F_{i,j} = \sum_{m=1}^{6} P_m C_{i,j}^{m,N} \tag{12}$$

donde

$$P_m = \frac{V_m}{\sum_{i=1}^{3} V_i} \qquad \text{m=1,2,3} \tag{13}$$

$$V_m = \frac{\sum_{S=1}^{N_t} C_S^{m,N}}{N_t} \tag{14}$$

4.4 Seguimiento de objetos

Después de la fusión de las características, se segmenta la imagen fusionada para detectar los posibles objetos de interés en la escena. Un patrón de búsqueda debe definirse para llevar a cabo el seguimiento. La correlación es procesada únicamente sobre los posibles objetos de interés y no sobre toda la imagen usando el patrón dado.

5 Resultados

Las secuencias de imágenes para este trabajo se adquirieron con un sensor infrarrojo PUMA con resolución de 640 x 480 y con un lente de 100 mm. Un sensor visible color DFK4003 con la misma resolución con un zoom de 8-108 mm. Ambos sensores se calibraron manualmente y se mantienen estáticos sobre la escena de interés.

El algoritmo fue evaluado con tres secuencias de imágenes adquiridas con ambos sensores sobre la misma escena con variaciones de iluminación. La Figura 4 muestra las imágenes adquiridas en diferentes instantes de tiempo t=0, t=20; las imágenes la izquierda fueron obtenidas con un cámara infrarroja y las imágenes de la derecha son en el rango visible, en las imágenes se aprecia un objeto en movimiento a largo del campo de visión de ambas cámaras. Este objeto es una podadora desplazándose sobre un campo. En ambos casos, existe diferencia de contraste del objeto de interés con respecto al fondo.

En la Figura 5 se muestran las imágenes resultantes de la fusión de las tres características para cada uno de los sensores, en estas imágenes se observa como la fusión realiza un filtrado de los objetos que satisfacen las características seleccionadas con respecto al fondo, aunque en ambos casos, aún se tienen varias regiones que serán eliminadas en el siguiente proceso de fusión multisensor.

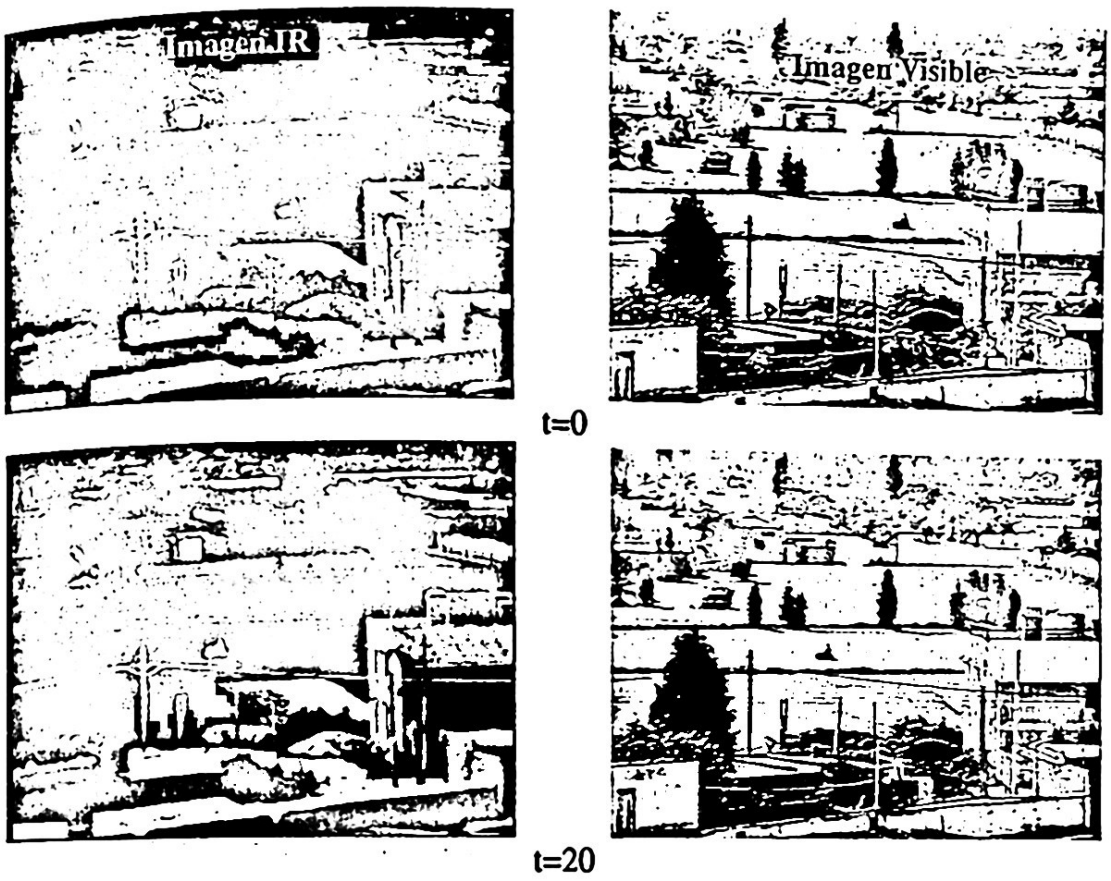

Fig. 4 Imágenes adquiridas usando los sensores infrarrojo y visible a diferentes tiempos. Nótese el objeto de interés (podadora).

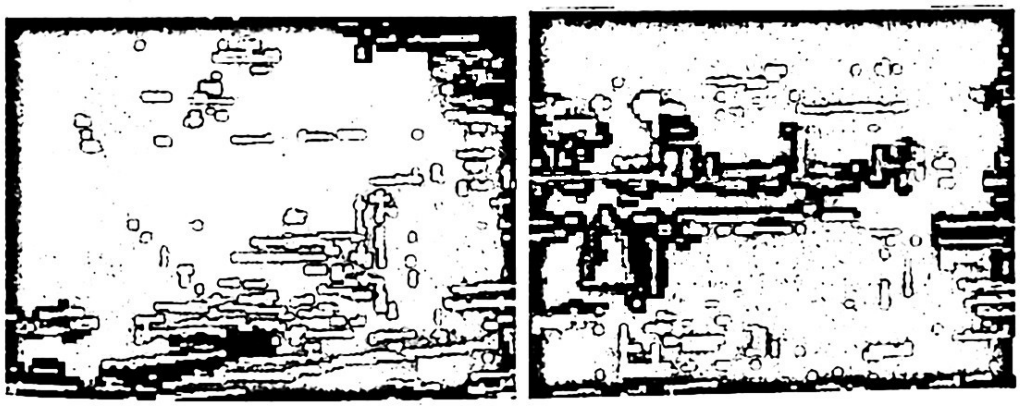

Fig. 5 Imágenes fusionadas

La figura 6 presenta la imagen multisensor resultante de combinar las seis características definidas en la sección cuatro, así como las regiones resultantes después de

realizar el proceso de segmentación. Finalmente la Figura 7 muestra el resultado del proceso de seguimiento. Los resultados en la detección y seguimiento del objeto para tres secuencias digitalizadas son satisfactorios ya que se discriminan en lo posible los objetos que no son de interés en la imagen con la fusión de las características seleccionadas para ambos sensores lo cual permite mantener el seguimiento del objeto en la escena.

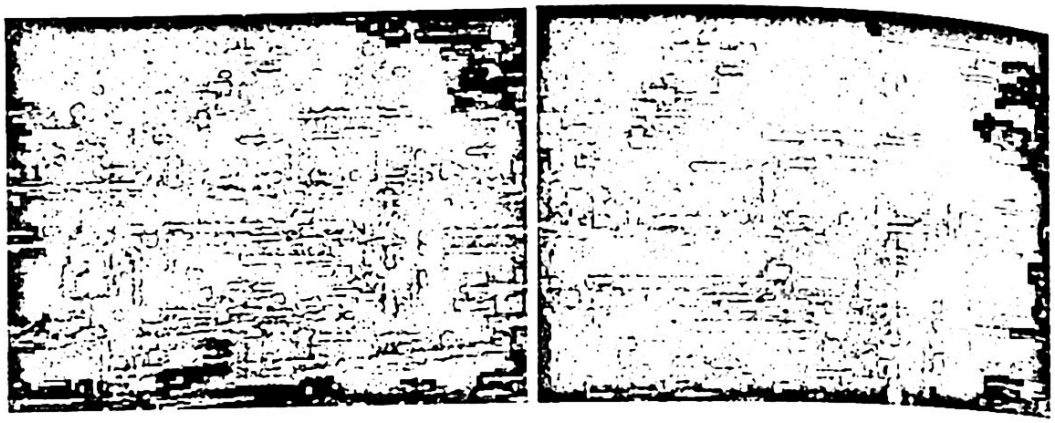

Fig. 6 Resultados de la fusión multisensor

Fig. 7 Resultados de la detección y seguimiento del objeto.

Al igual que en la figura 4 se muestran imágenes capturadas en diferentes instantes de tiempo para observar el seguimiento.

6 Conclusiones

En este trabajo se ha presentado un algoritmo de detección y seguimiento de objetos en secuencias de imágenes multisensor, el cual consiste de tres módulos principales: el primero, es determinar el tamaño de la ventana de los objetos de interés mediante un algoritmo de segmentación basado en la representación Multi-Normal a nivel pixel; el segundo es la fusión de características para la detección de objetos y último módulo es el seguimiento de objetos mediante correlación sobre los posibles objetos de interés detectados. La detección de la ventana de los objetos de interés

hace automáticamente. Su tamaño determina la eficiencia en el funcionamiento de los siguientes módulos, mostrando con nuestra propuesta resultados satisfactorios. Con la fusión de las características seleccionadas se logra eliminar la mayor cantidad de objetos que no son relevantes. El sistema completo se probó con 3 secuencias de aproximadamente 100 imágenes cada una.

Referencias

1. Beauvais M., Lakshmanan S., CLARK: A heterogeneous sensor fusion method for finding lanes and obstacles, Image and Vision Computing, Vol. 18, No. 1, 2000, 397-413.
2. Collins T. R., Lipton J. A., Fujiyoshi H., Kanade T., Algorithms for Cooperative Multisensor Surveillance, Proceedings of the IEEE, Vol. 89, No. 10, 2001, 1456-1477.
3. Fujiyoshi H., Lipton J. A., Real-time human motion analysis by image skeletonization, Proc. of the Workshop on Application on Computer Vision, October 1998.
4. Hall L. D., Llinas J., Handbook of Multisensor Data Fusion, CRC Press, Florida, 2001.
5. Kwon H., Der Z. S., Nasrabadi M. N., Adaptive multisensor target detection using feature-based fusion, Optical Engineering, Vol. 41, No. 1, January 2002, 69-80.
6. Lallier E., Farooq M., A Real Time Pixel Level Based Image Fusion Via Adaptive Weight Averaging, 3rd. International Conference on Information Fusion, FUSION 2000, Vol. 2, July 2000, WeC3-3 to WeC3-10.
7. Pavlidis I., Morellas V., Tsiamyrtzis P., Harp S., Urban surveillance systems: from the laboratory to the commercial world, Proceedings of the IEEE, Vol. 89, No.10, October 2001, 1478-1497.
8. Son G. J., Lim W. C., Choi I. Kim C. N., Adaptive Sizing of Tracking Window for Correlation-Based Video Tracking, IEICE Trans. Inf. & Syst., Vol. E85-D, No. 6, June 2002, 1015-1021.
9. Valet L., Mauris G., Bolon, A Statistical Overview of Recent Literature in Information Fusion, IEEE AESS Systems Magazine, March 2001, 7-14