

Fusión Sensorial para una Plataforma de Seguimiento

Gabriela Arri¹, María Curetti¹, Santiago García Bravo¹, Ladislao Mathé¹

¹ Departamento de Guiado y Control, Centro de Investigaciones Aplicadas,
Instituto Universitario Aeronáutico, Av. Fuerza Aérea 6500,
5000 Córdoba, Argentina

garri@iua.edu.ar, sgbravo@iua.edu.ar,
mcuretti@iua.edu.ar, mathe@ieee.org

Abstract. Este paper presenta una estrategia de seguimiento de objetivos aéreos basada en fusión sensorial para una plataforma de seguimiento óptico. Se describe la incorporación de datos de un sistema de posicionamiento global (GPS) a la información provista por un sensor óptico montado en la plataforma. Se especifica el método de recepción y procesamiento de la información y se utiliza un estimador de filtros de Kalman para la fusión de datos. El método propuesto es implementado experimentalmente y los resultados obtenidos evidencian un seguimiento robusto del objetivo en tiempo real.

Keywords: Fusión sensorial, plataforma de seguimiento, Global Positioning System (GPS), *visual servoing*, tiempo real, procesamiento de imágenes.

1 Introducción

La detección y el seguimiento de objetos es un área de investigación en el campo de la visión artificial. El seguimiento por imágenes es crucial para muchas aplicaciones que van desde sistemas de vigilancia [1], [2], [3], robótica [4], [5], interfaces máquina-humano [6], [7], [8] y reconstrucción por imágenes [9], [10], [11] hasta automatización agrícola [12], [13], imágenes médicas [14], [15], sistemas de detección de incendios, reconocimiento de rostros y gestos y muchas otras más [16], [17], [18].

Básicamente, el seguimiento de objetivos, conocido como “*visual servoing*” o “*controlado por imágenes*”, es el proceso de localización de un objeto en movimiento, observado a través de una videocámara, a lo largo de una secuencia de cuadros o capturas. Los sensores ópticos proveen la información necesaria para determinar la posición de los objetivos seleccionados dentro del cuadro de la imagen, mediante una máscara o patrón específicos que modelan el aspecto.

El estado de este arte ha avanzado significativamente [19 --26]. En los últimos 30 años se han desarrollado diferentes tipos de sistemas controlados o guiados por imágenes [27--30]. En concreto, estos sistemas utilizan un algoritmo (implementado en una rutina de software) que reconoce el objeto seleccionado en la secuencia de imágenes de entrada calculando sus coordenadas con respecto al centro de cada uno de

los cuadros de la secuencia adquirida. En los algoritmos de seguimiento de objetos de naturaleza variable, generalmente el objeto a seguir es seleccionado por el usuario, quien lo identifica en el primer cuadro. Luego, el sistema de control guía o dirige la cámara con el fin de lograr que la imagen del objetivo esté en el centro del campo visual (centro del cuadro de la imagen) y lo mantenga allí. Restricciones de movimiento restringen el espacio de los lugares o las regiones en las que se va a buscar el objetivo.

Sin embargo, lo que dificulta el seguimiento es la variabilidad potencial de las imágenes de un objeto a lo largo del tiempo. Esta variabilidad se debe a las siguientes razones principales: variación en la actitud y tamaño (rotación + traslación) del objetivo, variación en la iluminación y oclusiones parciales o totales del objetivo. Es decir que, los sistemas deben ser adecuados para seguir objetos ante repentinos cambios de iluminación, contrastes de colores insuficientes, niveles de ruido en la imagen, aparición de obstáculos como montañas, edificios, nubes, etc. Ignorar alguna de estas razones es suficiente para ocasionar que el algoritmo de seguimiento por imágenes pierda su objetivo. Para poder superar estas dificultades se ha recurrido a distintos enfoques con diferente grado de carga computacional asociado. [31], [32], [33], [34]

La actualización de patrón cada cierto número de capturas es un método elemental que, aplicado con suficiente regularidad, constituirá una solución válida para algunos de los problemas que se presentan. Su principal falla es que a lo largo del tiempo el objeto puede moverse dentro de la máscara y terminar escapando de la misma. Para mantener el elemento centrado en el patrón se propone en [35] conservar el primer patrón y utilizarlo para el centrado del patrón actual. También se pueden producir oclusiones pasajeras, deteriorando la máscara rápidamente en algoritmos poco robustos. En [36] se utilizan filtros de Kalman para estimar la intensidad del nuevo patrón y de esta manera se robustece el algoritmo.

Muchos de estos métodos han demostrado ser efectivos ante los cambios de aspecto y las oclusiones parciales o totales de corta duración. Sin embargo, el inconveniente que no puede ser solucionado utilizando la información visual únicamente es el de las oclusiones totales prolongadas. En ese tipo de situaciones es casi imposible predecir la región donde el objetivo volverá a hacerse visible para la cámara, y la mayoría de las veces ocasiona la discontinuidad en el seguimiento.

Para este tipo de situaciones es necesario recurrir a información sensorial complementaria a la visual. Por consiguiente, son necesarios métodos de *fusión sensorial*, que en este contexto se refiere a procesos en los cuales se integran los datos provenientes de diferentes sensores para detectar objetos y para estimar parámetros y estados necesarios para el seguimiento, la ubicación geográfica, la ejecución de movimiento y la toma de decisiones manteniendo al mínimo la complejidad computacional para un funcionamiento eficiente en tiempo real y generando así un sistema más robusto y resistente a factores aleatorios. [37], [38], [39], [40].

2 Descripción del Sistema

El proyecto para el desarrollo de un sistema de seguimiento de precisión por procesamiento de imágenes, llamado TV Tracker, surge en el Instituto Universitario Aeronáutico a partir de la necesidad de lograr la visualización directa y registro de imágenes de ensayos de móviles aéreos (blanco) y de los eventos que sucedan durante los mismos. El proyecto cuenta con financiamiento del Ministerio de Defensa (PIDDEF).

El sistema TV Tracker consta de una plataforma servo controlada en azimut y en elevación sobre la cual se colocó una cámara filmadora cuya señal de video es utilizada en la computadora de control, donde se procesa la imagen para extraer la información de la posición del objetivo blanco respecto al centro del cuadro de la imagen. El desarrollo del algoritmo de procesamiento de imágenes para obtener dicha información no es trivial. Tomando como punto de partida las soluciones planteadas en los trabajos [32], [35], [36], se elaboró un software de seguimiento propio [41] que se ajusta a las necesidades específicas del proyecto.

La posición en pixeles obtenida por la etapa de procesamiento de imágenes, es traducida a ángulos de comando para los motores que centran en pantalla al blanco. La Fig. 1 muestra el diagrama en bloques correspondiente a la plataforma servocontrolada. Dentro del recuadro gris se observa la etapa de control C, el driver de potencia D que comanda a los motores E y la realimentación de posición de los encoder representada por F1.

Al extremo de la plataforma se asocia la cámara de filmación de video V, cuyas imágenes son digitalizadas en Z y procesadas para generar la realimentación F2, y de este modo generar el próximo comando de control para lograr el seguimiento.

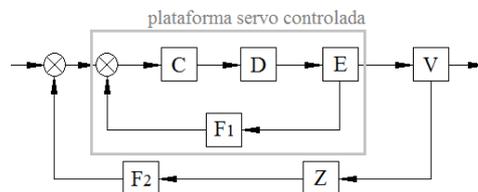


Fig. 1. Diagrama en bloques del sistema TV Tracker.

En este trabajo se plantea entonces incorporar un sistema GPS (Global Positioning System) inercial, montado en el objetivo a seguir, que proporcione información de posicionamiento global compuesta por las coordenadas de latitud, longitud, y altitud. Esto permitirá robustecer el seguimiento de los blancos posibilitando la continuidad en el seguimiento, ante la ocurrencia de oclusiones totales prolongadas, y proveerá además datos iniciales útiles para el proceso de pre-apuntado en el cual se localiza el blanco por primera vez.

3 Entrada de la información proveniente del GPS

El enlace de datos de telemetría proveniente del GPS se realiza a través de una configuración peer-to-peer de dos módulos radio-modem LAIRDTECH CL4790 que se comunican mediante un protocolo serie a 5600 baudios.

El módulo remoto cuenta con una unidad de GPS inercial Xsense MTI-G que transmite la trama telemétrica que contiene la información de latitud, longitud y altitud (WGS-84) codificada en el estándar IEEE-754 de 32 bits. La información de coordenadas geodésicas es actualizada a una tasa de 25Hz y es contenida en una trama que incluye doble verificación de protocolo con el fin de evitar el procesamiento de información errónea en etapas posteriores de operación.

El programa incorpora la información a través de una lectura serial del puerto de entrada, verifica el valor de checksum de la trama y de ser válido lo habilita para las etapas posteriores de procesamiento.

4 Transformación de coordenadas

Como el rango de vuelo en el cual se opera se encuentra limitado a una área pequeña debido al rango visual, es conveniente modelar a la Tierra como un sub-espacio plano en lugar de una esfera achatada. Para esto, se utiliza un sistema de coordenadas llamado Tierra-Plana (x, y, z), el cual es un sistema cartesiano con origen en la posición del TV Tracker sobre la superficie cuyo eje z apunta verticalmente hacia abajo, el eje x apunta al Norte y el eje y, al Este.

Una vez obtenidas las coordenadas de posición provenientes del GPS, a saber: latitud geodésica, longitud y altitud, se utiliza una función del Toolbox Aerospace de MATLAB que estima la posición Tierra-Plana de un objeto a partir de aquellas coordenadas; y lo hace con respecto a una locación de referencia definida (por su latitud, longitud y altura), en nuestro caso: la plataforma de seguimiento. Para esta aplicación, se debe tener en cuenta el modelo elipsoidal deseado del planeta especificando el índice de achatamiento y el radio ecuatorial.

La estimación comienza encontrando las pequeñas variaciones en latitud y longitud entre la latitud y longitud recibidas del GPS menos la latitud y longitud inicial de referencia.

$$d\mu = \mu - \mu_0 \quad (1)$$

$$dl = l - l_0 \quad (2)$$

Para convertir latitud geodésica y longitud a coordenadas Norte-Este, se utiliza el radio de curvatura en el vertical primario (R_N) y el radio de curvatura en el meridiano (R_M). R_N y R_M están definidos por las siguientes relaciones:

$$R_N = \frac{R}{\sqrt{1 - (2f - f^2) \sin^2 \mu_0}} \quad (3)$$

$$R_M = R_N \frac{1-(2f-f^2)}{1-(2f-f^2) \sin^2 \mu_0} \quad (4)$$

donde (R) es el radio ecuatorial del planeta y (f) es el achatamiento/aplastamiento del planeta.

Pequeñas variaciones en las posiciones Norte (dN) y Este (dE) son aproximadas mediante pequeños cambios en la latitud y longitud:

$$dN = \frac{d\mu}{\operatorname{atan}\left(\frac{1}{R_M}\right)} \quad (5)$$

$$dE = \frac{dl}{\operatorname{atan}\left(\frac{1}{R_N \cos \mu_0}\right)} \quad (6)$$

Mediante la conversión de coordenadas Norte-Este a coordenadas Tierra-Plana x e y, la transformación tiene la forma:

$$\begin{bmatrix} p_x \\ p_y \end{bmatrix} = \begin{bmatrix} \cos \Psi & \sin \Psi \\ -\sin \Psi & \cos \Psi \end{bmatrix} \begin{bmatrix} N \\ E \end{bmatrix} \quad (7)$$

donde (Ψ) es el ángulo medido en grados horarios entre el eje x y el Norte.

Por otro lado, el valor de la coordenada z Tierra-Plana es igual a la altitud negativa proveniente de GPS menos la altura de referencia (h_{ref}).

$$p_z = -h - h_{ref} \quad (8)$$

Una vez calculadas las coordenadas Tierra-Plana x, y, z del objeto con respecto a la plataforma de seguimiento, se utilizan funciones trigonométricas para calcular el ángulo respecto al Norte, es decir el ángulo de azimut, y el ángulo respecto al plano horizontal o ángulo de elevación. El ángulo de azimut se encuentra en el rango 0-360 y está definido positivo en sentido horario.

5 Proceso de geo-referenciamiento

Para que la plataforma de seguimiento sea capaz de interpretar correctamente las coordenadas enviadas por el GPS es necesario llevar a cabo con antelación el proceso de geo-referenciamiento basado en establecer la posición del norte geográfico y del plano horizontal (paralelo al nivel del mar).

En primer lugar, para establecer la posición del norte es necesario conocer al ángulo de la plataforma respecto a éste para lo cual se debe contar con las posiciones o coordenadas de dos (2) puntos a saber: la plataforma de seguimiento y cualquier otro punto de coordenadas conocidas.

Luego, para encontrar el plano horizontal, debe determinarse el plano de rotación sobre el cual ha sido montado el sistema, que en la mayoría de los casos no se encuentra paralelo al nivel del mar debido a la irregularidad del terreno. Para esto, deberá agregarse un tercer punto de ubicación conocida y de este modo quedará definido el plano centrado en la plataforma que contiene los tres (3) puntos (posición propia, el

anterior y el más reciente). La intersección de este plano con un cilindro centrado en las coordenadas del pedestal definirá una curva centrada en la plataforma de seguimiento que permitirá parametrizar la corrección en elevación necesaria con respecto al plano horizontal según el valor del ángulo en azimut.

La Fig. 2 muestra un ejemplo de la curva de corrección obtenida, centrada en la plataforma respecto al horizonte real.

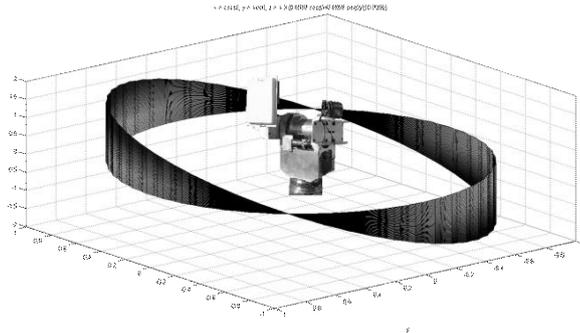


Fig. 2. Curva de corrección en el eje z en el rango completo de azimut.

6 Etapa de fusión sensorial

En un sistema de seguimiento con múltiples sensores, se puede obtener una estimación óptima del estado del objetivo por medio de un seguidor/fusionador centralizado que incorpora las mediciones de cada uno de los sensores. El fin esencial es aprovechar las virtudes y suplir las falencias de cada tipo de sensor. En este trabajo se busca realizar el seguimiento aprovechando un sistema de posicionamiento global (GPS) y un sistema óptico, con el fin de lograr el seguimiento continuo de objetivos a lo largo de un ensayo determinado.

6.1 Consideración sobre sistema GPS

Como es sabido, la precisión en la estimación de la posición global depende de varios factores entre los que se destaca principalmente el número de satélites disponibles en el momento. De los tres parámetros que definen la posición de un punto del espacio, el menos preciso resulta ser la altitud. La figura 8 muestra la variación de éste parámetro a lo largo de un ensayo. Esta incertidumbre también se cumple, con menor dispersión, en los parámetros de latitud y longitud.

Esta dispersión en los valores de las posiciones debe ser tenida en cuenta con el fin de alcanzar la precisión necesaria para lograr el registro de imágenes de los blancos, donde el ángulo de visión es del orden del grado de arco.

6.2 Consideraciones sobre imágenes

En contraparte, el sistema de seguimiento por imágenes es altamente preciso ya que el ángulo de visión se encuentra discretizado a la porción angular que corresponde a un pixel (valor variable según las características de la cámara). Sin embargo esta discriminación del espacio carece de una referencia absoluta, es decir que el seguimiento por imágenes sólo tiene sentido dentro del rango de visión, y si el blanco se encuentra fuera del campo de visión (FOV) se debe re-apuntar para continuar el seguimiento. Asimismo, también puede suceder que la imagen del blanco se vea obstruida temporalmente por algún obstáculo, como por ejemplo: nubes, árboles, terreno, o simplemente se aleje más allá del alcance visual de la cámara. En estos casos, la información proporcionada por el sensor óptico se ve interrumpida y para poder continuar el seguimiento es imprescindible contar con datos aportados por otro tipo de sensor.

6.3 Fusión sensorial

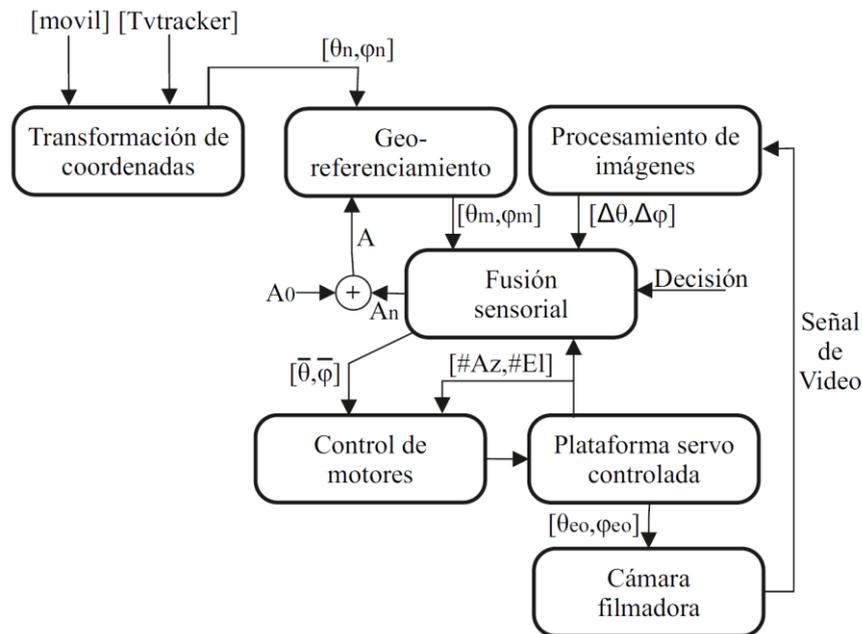


Fig. 3. Diagrama en bloques de la fusión sensorial del sistema

En el diagrama se muestran los bloques principales del sistema. El bloque transformación de coordenadas tiene como entradas $[movil]$ y $[Tvtracker]$ que corresponden a las coordenadas geodésicas del objetivo y el sistema respectivamente, el resultado de este bloque son las coordenadas θ_n y φ_n que son los ángulos de apuntado respecto al Norte de azimut y elevación respectivamente. Éstos son interpretados por el

bloque de geo-referenciamiento que corrige según el horizonte artificial y traduce los valores de azimut y elevación relativos al marco de referencia de TV Tracker (θ_m , φ_m).

La posición angular del objetivo proveniente del módulo procesamiento de imágenes es de naturaleza incremental, referido al eje óptico de la cámara ($\Delta\theta$, $\Delta\varphi$).

Fusión sensorial es el bloque que a partir de las entradas anteriormente mencionadas, y de las cuentas de los encoders de los motores (#Az, #El) ajusta un modelo de dinámica sencillo (aceleración constante) mediante un filtro de Kalman clásico, generando las estimaciones sobre el estado del objetivo. Esta estimación por el momento no contempla correlación entre azimut y elevación.

Estas estimaciones son la referencia para el control de la plataforma servo controlada, y son utilizadas para la conformación del conjunto de puntos A_n que ajustan el horizonte artificial junto con los datos de apuntado inicial A_0 .

La variable booleana “Decisión”, es actualmente controlada por el operario del sistema. Su función es definir la validez de los datos $\Delta\theta$ y $\Delta\varphi$. Al no poseer información válida de la etapa de imágenes, se altera el modelo dinámico, y se continúa estimado los estados con la información disponible.

A partir de lo expuesto se planea el siguiente ensayo que permitirá validar la estrategia de fusión, y además obtener una primera base de datos para desarrollos futuros.

7 Resultados experimentales

Para comprobar el funcionamiento de la plataforma de seguimiento con la incorporación de los datos proporcionados por el sistema GPS, se realizó un ensayo cuyo objetivo fue evaluar la velocidad de pre apuntado o apuntado inicial y la capacidad de seguimiento del sistema ante la ocurrencia de oclusiones de distinta duración. El ensayo contó con dos etapas: en la primera se realizaron maniobras de detección de objetivo, con y sin información de GPS, y en la segunda se hicieron tres ejercicios de seguimiento continuo con oclusiones totales del objetivo de duración variable. Para todos estos ensayos se utilizó como blanco u objetivo a seguir un módulo de transmisión construido para tal efecto con el sistema GPS-inercial Xsense MTI-G (Fig. 4).



Fig. 4. Módulo de transmisión – GPS inercial.

En la primer atapa se llevaron a cabo quince maniobras de pre-apuntado con GPS y quince maniobras con el sistema original para comparar la performance de ambos métodos. Para esto se colocó el blanco en una posición fija y la plataforma en una determinada posición inicial. Los tiempos obtenidos en esta serie de ensayos resultaron considerablemente menores para el caso con información de GPS, obteniéndose un valor medio 16 veces menor. Durante esta etapa del ensayo se evidenció además la dispersión de los datos en la altitud estimada del objetivo.

En la segunda etapa del ensayo, el módulo transmisor fue trasladado realizando distintas trayectorias con períodos visibles y ocluidos de distinta duración (Fig. 5, 6). A lo largo de las trayectorias fue posible mantener el objetivo dentro del campo visual de la cámara (FOV) incluso durante las oclusiones (Fig. 7). Nuevamente el mayor error durante las etapas de seguimiento sin contacto visual, se encontraba en la información de altitud. Durante la oclusión de mayor duración de 24 minutos aprox. fue necesario disminuir el zoom (aumentar el FOV) para mantener el objetivo dentro del campo visual. En todas las oclusiones la altitud estimada aumentaba en forma monótona mientras el objetivo se mantenía a la misma altura, alcanzando un error máximo de 17.36 metros (Fig. 8).

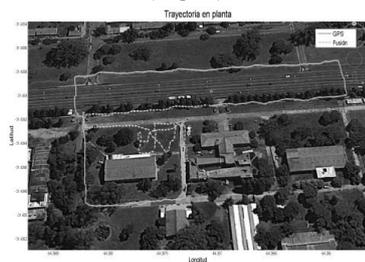


Fig. 5. Trayectoria con fusión de datos y trayectoria con información GPS únicamente



Fig. 6. Módulo centrado durante una de las trayectorias.

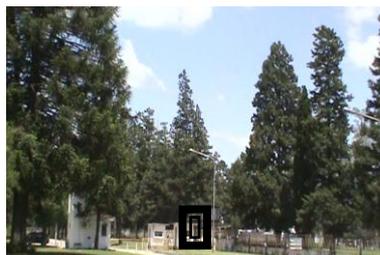


Fig. 7. El blanco permanece en FOV durante la oclusión.

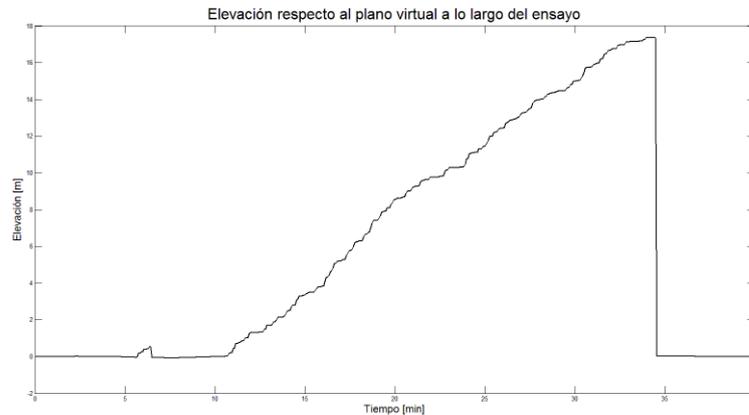


Fig. 8. Acumulación de error en la altitud durante las oclusiones.

8 Conclusión

Con la incorporación de información de posicionamiento global al sistema TV Tracker, se logró continuar el seguimiento del objetivo sin dificultades a pesar de la pérdida del contacto visual con el mismo por periodos prolongados de tiempo. Además se consiguió automatizar el pre apuntado del sistema, disminuyendo el tiempo necesario para dicha maniobra en 16 veces. También se advirtió un error acumulativo en la altitud estimada por el sensor GPS-inercial que podría afectar el correcto funcionamiento del sistema ante la ocurrencia de oclusiones de duración mayor a 24 minutos, especialmente cuando el objetivo esta posicionado en las cercanías de TV Tracker.

En etapas futuras del proyecto se espera repetir este ensayo con un blanco aéreo para profundizar sobre el modelo de fusión y la teoría de decisión a implementar en el futuro. Además se prevé estudiar el comportamiento del sistema ante la incorporación de otro tipo de sensores (IMU, radar, etc).

9 Referencias

1. J. Zhu, Y. Lao, and Y. Zheng, "Object tracking in structured environments for video surveillance applications", *IEEE Trans. On Circuits and Systems for Video Technology*, vol. 20, no. 2, pp. 529-551, 2010.
2. R. Howarth and H. Buxton, "Visual Surveillance Monitoring and Watching," *Proc. European Conf. Computer Vision*, vol. 2, pp. 321-334, 1996.
3. T. Frank, M. Haag, H. Kollnig, and H.-H. Nagel, "Tracking of Occluded Vehicles in Traffic Scenes," *Proc. European Conf. Computer Vision*, vol. 2, pp. 485-494, 1996.
4. P. Allen, B. Yoshimi, and A. Timcenko, "Hand-Eye Coordination for Robotics Tracking and Grasping," K. Hashimoto, ed., *Visual Servicing*, pp. 33-70. World Scientific, 1994.

5. N. Papanikolopoulos, P. Khosla, and T. Kanade, "Visual Tracking of a Moving Target by a Camera Mounted on a Robot: A Combination of Control and Vision", *IEEE Trans. Robotics and Automation*, vol. 9, no. 1, pp. 14-35, 1993.
6. A.F. Bobick and A.D. Wilson, "A State-Based Technique for the Summarization of Recognition of Gesture", *Proc. Int'l Conf. Computer Vision*, pp. 382-388, 1995.
7. T. Darrell, B. Moghaddam, and A. Pentland, "Active Face Tracking and Pose Estimation in an Interactive Room", *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 67-72, 1996.
8. D. Gavrilu and L. Davis, "Tracking Humans in Action: A 3D Model-Based Approach," *Proc. Image Understanding Workshop*, pp. 737-746, 1996.
9. E. Boyer, "Object Models From Contour Sequences", *Proc. European Conf. Computer Vision*, vol. 2, pp. 109-118, 1996.
10. L. Shapiro, "*Affine Analysis of Image Sequences*". Cambridge, England: Cambridge Univ. Press, 1995.
11. C. Tomasi and T. Kanade, "Shape and Motion From Image Streams Under Orthography: A Factorization Method", *Int'l J. Computer Vision*, vol. 9, no. 2, pp. 137-154, 1992.
12. R.C. Harrell, D.C. Slaughter, and P.D. Adsit, "A Fruit-Tracking System for Robotic Harvesting", *Machine Vision and Applications*, vol. 2, pp. 69-80, 1989.
13. D. Reynard, A. Wildenberg, A. Blake, and J. Marchant, "Learning Dynamics of Complex Motions From Image Sequences," *Proc. European Conf. Computer Vision*, vol. 1, pp. 357-368, 1996.
14. E. Bardinet, L. Cohen, and N. Ayache, "Tracking Medical 3D Data With a Deformable Parametric Model," *Proc. European Conf. Computer Vision*, vol. 1, pp. 317-328, 1996.
15. P. Shi, G. Robinson, T. Constable, A. Sinusas, and J. Duncan, "A Model-Based Integrated Approach to Track Myocardial Deformation Using Displacement and Velocity Constraints," *Proc. Int'l Conf. Computer Vision*, pp. 687-692, 1995.
16. Courbon, Y. Mezouar, N. Guenard, and P. Martinet, "Vision-based navigation of unmanned aerial vehicles ", *Control Engineering Particle*, vol. 18, no. 7, pp. 789-799, 2010.
17. H. Hile, A. Liu, G. Borriello, "Visual navigation for mobile devices", *IEEE Multimedia*, vol. 17, no. 2, pp. 16-25, 2010.
18. T. Sasaki, D. Brscic, and H. Hashimoto, "Human-observation-based extraction of path patterns for mobile robot navigation," *IEEE Trans. On Industrial Electronics*, vol. 57, no. 4, pp. 1401-1410, 2010.
19. Lucas, B., Kanade, T. "An iterative image registration technique with an application to stereo vision", *IJCAL*, pp. 674-679, 1981.
20. Shi, J., Tomasi, C., "Good features to track", *IEEE CVPR*, pp. 593-600, 1994.
21. Isard, M., Blake, A., "A smoothing filter for condensation", *ECCV*, pp. 767-781, 1998.
22. Comaniciu, D., Ramesh, V., Meer, P., "Real-time tracking of non-rigid objects using mean shift", *IEEE CVPR*. 2142,2000.
23. Adam, A., Rivlin, E., Shimshoni, I., "Robust fragments-based tracking using the integral histogram", *IEEE CVPR*, pp. 798-805, 2006.
24. Avidan, S., "Ensemble tracking", *IEEE PAMI* 29 , pp. 261-271, 2007.
25. Li, Y., Ai, H., Yamashita, T., Lao, S., Kawade, M.: Tracking in low frame rate video: A cascade particle filter with discriminative observers of different life spans. *IEEE PAMI* 30, pp. 1728-1740, 2008.
26. Ozuysal, M., Calonder, M., Lepetit, V., Fua, P., "Fast keypoint recognition using random ferns", *IEEE Trans. Pattern Anal. Mach. Intell.* 32 , pp. 448-461, 2010.

27. P. K. Allen, A. Timcenko, B. Yoshimi, and P. Michelman, "Automated tracking and grasping of a moving object with a robotic hand-eye system," *IEEE Trans. on Robotics and Automation*, vol. 9, no. 2, pp. 152-165, 1993.
28. N. Hollinghurst, R. Cipolla, "Uncalibrated stereo hand eye coordination" *Image and Vision Computing*, vol. 12, no. 3, pp. 187-192, 1994.
29. G. O. Hager, W. e. Chang, A.S. Morse, "Robot hand-eye coordination based on stereo vision," *IEEE Control Systems Magazine* vol. 15, no. 1, pp. 30-39, 1995.
30. E. Malis, F. Chaumette, S. Boudet, "Visual Servoing", *IEEE Trans. On Robotics and Automation*, vol. 15, no. 2, pp. 238-250, 1999.
31. J. Xiao, T. Kanade, and J. Cohn, "Robust Full-Motion Recovery of Head by Dynamic Templates and Re-Registration Techniques," *Proc. IEEE Int'l Conf. Automatic Face and Gesture Recognition*, pp. 163-169, 2002.
32. Yu Zhong, Anil K. Jain, Fellow, and M.-P. Dubuisson-Jolly, "Object Tracking Using Deformable Templates" *IEEE Transactions On Pattern Analysis And Machine Intelligence*, vol. 22, no. 5, may 2000.
33. T. Vetter, M. Jones, and T. Poggio, "A Bootstrapping Algorithm for Learning Linear Models of Object Classes," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 40-46, 1997.
34. I. Matthews and S. Baker, "Active Appearance Models Revisited," *Int'l J. Computer Vision*, 2004, also appeared as Technical Report CMU-RI-TR-0302, Robotics Inst., Carnegie Mellon Univ., Pittsburgh, Penn.
35. Iain Matthews, Takahiro Ishikawa, and Simon Baker, "The Template Update Problem" *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE*, VOL. 26, NO. 6, JUNE 2004.
36. Hieu T. Nguyen, Marcel Worring, and Rein van den Boomgaard "Occlusion robust adaptive template tracking" *Intelligent Sensory Information Systems University of Amsterdam, Faculty of Science Kruislaan 403, NL- 1098 SJ, Amsterdam, The Netherlands*.
37. Jurado, Juan, Kenneth Fisher, and Michael Veth. "Inertial and imaging sensor fusion for image-aided navigation with affine distortion prediction." *Position Location and Navigation Symposium (PLANS), 2012 IEEE/ION. IEEE*, 2012.
38. Yoshitomi, Y.; Sung-Ill Kim; Kawano, T.; Kilazoe, T., "Effect of sensor fusion for recognition of emotional states using voice, face image and thermal image of face," *Robot and Human Interactive Communication, 2000. RO-MAN 2000. Proceedings. 9th IEEE International Workshop on*, vol., no., pp.178,183, 2000
39. Van Baar, Jeroen, et al. *Sensor Fusion for Depth Estimation, including TOF and Thermal Sensors.* *3D Imaging, Modeling, Processing, Visualization and Transmission (3DIMPVT), 2012 Second International Conference on. IEEE*, 2012.
40. Wang, Xiaogang, Naigang Cui, and Jifeng Guo. "INS/VisNav/GPS relative navigation system for UAV". *Aerospace Science and Technology* (2012).
41. Curetti, Maria; Garcia Bravo, Santiago; Arri, Gabriela S.; Mathe, Ladislao, "Adjustable Tracking Algorithm with Adaptive Template Matching", *Latin America Transactions, IEEE (Revista IEEE America Latina)* , vol.11, no.1, pp.34,39, Feb. 2013 doi: 10.1109/TLA.2013.6502774.