Una Estrategia para la Selección Dinámica de Características Aplicada a la Estabilización de Secuencias de Imágenes

An Strategy for the Dynamic Selection of Features Applied to the Stabilization of Image Sequences

Hugo Jiménez Hernández[1] y Joaquín Salas Rodríguez[2]
Centro de Investigación en Ciencia Aplicada y Tecnología Avanzada del IPN
[1]hugojh@gmail.com; [2]salas@ieee.org

Artículo recibido en Marzo 10, 2010; aceptado en Junio 11, 2010

Resumen. En este trabajo se presenta una estrategia para la discriminación entre las características pertenecientes a objetos fijos y móviles de una escena observada desde una cámara sujeta a vibración. Nuestra estrategia selecciona como características fijas aquellas que minimizan el error de la proyección de la homográfica entre las imágenes, siendo tolerante a oclusiones de regiones y cambios lumínicos en la escena. Una posible aplicación de este resultado es la estabilización de secuencias de imágenes. En una etapa experimental, utilizando distintos escenarios en exteriores, mostramos los resultados y evidencia que los niveles de precisión obtenidos son mejores, que los obtenidos por propuestas eficientes basadas en la selección aleatoria de características, tal como la de RANSAC.
Palabras clave: Selección dinámica de características, estimación de la homografía, método no supervisado.

Abstract. This work introduces an algorithm to discriminate between either moving or static features of a given scene as observed from a fixed camera, which is under the effects of vibration,. In our strategy, we select the features minimizing the registration error from one image to the next one. The process discards the features corresponding to moving objects and untrackable regions. The algorithm is applied to the task of stabilizing an image sequence. In our experiments, we benchmark our approach with several images sequences and match the results with a randomized strategy known as RANSAC.
Key words: Dynamic selection of features, estimation of the copying, unsupervised method.

1 Introducción

En este documento, estudiamos el problema del mantenimiento de un sistema referencia cuando las características que son utilizadas para su determinación están sujetas a cambios de posición, son ocluidas, o cambian de forma. Esto es, dado un escenario que contiene objetos en movimiento y una cámara sujeta a pequeños movimientos debido a vibraciones, estudiamos el problema de la distinción entre los elementos de la escena que permanecen fijos y aquellos que tienen movimiento. El problema es complicado pues con una cámara en movimiento todo pareciera estar cambiando de posición. Varios investigadores, por ejemplo Tomasi [Tomasi, 1993], han desarrollado el concepto mediante el cual un conjunto de regiones altamente distinguibles puede servir para modelar propiedades de la escena en su conjunto, particularmente su estructura tridimensional. Todavía más, Jianbo y Tomasi [Jianbo y Tomasi, 1994], estudiando el método de seguimiento desarrollado por Lucas y Kanade [Lucas y Kanade, 1981], mostrando que efectivamente las características útiles son aquellas que sirven mejor para la solución del problema propuesto. El problema de identificar la calidad de las características ha sido también estudiado por Tan *et al*. [Tan *et al*; 1996]. Con todo ello, en este estudio desarrollamos la idea de que cuando la mayoría del conjunto de características seguibles es parte de la porción de la escena que permanece estática, entonces es posible construir un algoritmo mediante el cual discriminarlas de aquellas que pertenecen a objetos en movimiento, son ocluidas, o cambian de geometría.

Así pues, la selección dinámica de características permite que algoritmos tales como los de reconstrucción tridimensional, discriminación de objetos en movimiento o que la estabilización de imágenes sea más robustos bajo diferentes condiciones. Collins y Liu [Collins y Liu; 2003] estudian el problema de la determinación de objetos en movimiento mediante la selección dinámica de características, aun cuando en su propuesta la cámara no está sujeta a vibración. Luego, en [Hwann *et al.,* 2004], se propone un esquema basado en la selección dinámica de

características de color aplicando un filtro de partículas, asumiendo que el desplazamiento de las características es causado sólo por el desplazamiento de los objetos. Las características de color son usadas para distinguir los objetos en movimiento en la escena. Por otra parte en [Tan y Tao, 2005] se propone un esquema de selección de características basado en la selección de características invariantes a escala (SIFT) [Mikolajczyk y Schmid, 2005], donde a cada objeto en movimiento se le asocia un grafo de relación entre sus características. En conjunto, la topología del grafo representa la estructura espacial de los objetos en movimiento en la escena. Sin embargo, ante movimientos repentinos de la cámara, ocasionan una ineficiente detección de los objetos en movimiento. En [Jurisica *et al*., 2006], se presenta un método de selección de características para el seguimiento, basado en el flujo óptico. Su contribución consiste en detectar las zonas que hacen que las ecuaciones de movimiento estén bien comportadas, asumiendo que la cámara permanece fija. Finalmente, en [ZuWhan, 2008], se muestra una selección dinámica de características basada en la detección de esquinas. Las características son calculadas a partir de las regiones con movimiento mediante un modelo de sustracción de fondo y son agrupadas en cúmulos de acuerdo a la dinámica del movimiento. Para diferenciar a los objetos en movimiento de los falsos positivos, asumen que los descriptores del fondo siempre permanecen estáticos. En general, estos trabajos muestran que la selección dinámica de características permite elegir mejor los descriptores de los objetos. Sin embargo, se asume que la cámara está libre de movimiento y el movimiento de las características sólo es causado por el desplazamiento de objetos en movimiento.

La selección dinámica de características ha sido utilizada para desarrollar métodos de detección de alineación de imágenes. Una de las aproximaciones comúnmente utilizadas es *Random Sample Consensus* (RANSAC) [Fischler y Bolles, 1981]. Mediante esta aproximación se construye un conjunto de transformaciones considerando un muestreo aleatorio sobre los datos y utilizando una función de distancia se eliminan las aberraciones en la transformación. Más recientemente, en [Lacey *et al*., 2000] y en [Chum y Matas, 2008], se intenta mejorar la precisión y complejidad del método RANSAC usando un criterio de selección, basado en el radio de probabilidad secuencial de Walds (SPRT) [Wald, 1945], al reducir el número de aberraciones en los datos. Pero, para que estas aproximaciones resulten eficientes, se asume que las imágenes sólo contienen elementos estáticos. También, el uso de características dinámicas se ha tratado como un problema de modelado y filtrado. En un trabajo seminal [Gil, 1994 *et al*], Gil determina las características relacionadas con la forma, posición y textura de los vehículos en movimiento. Estas características son utilizadas para estimar la posición mediante un filtro de Kalman [Kalman, 1960]. Las características son adaptadas de acuerdo al cambio de la perspectiva de los vehículos observados. Posteriormente, en [Cham y Rehg, 1999] se propone una estrategia para ordenar y seleccionar un conjunto de características de un patrón de búsqueda a una imagen, por medio de una transformación lineal a partir de un filtro de Kalman. Pero, estas aproximaciones requieren de una retroalimentación constante y se asume conocimiento sobre el modelo de movimiento de los objetos, lo cual puede no ser el caso en algún escenario.

En este trabajo se presenta una estrategia para la selección dinámica de características en las imágenes de una secuencia. Las características corresponden a las regiones que permanecen estáticas en la escena, a pesar de que la cámara que capta las imágenes esté sujeta a pequeñas vibraciones o que algunas de las características presenten oclusiones. El proceso de selección consiste en utilizar aquellas características que estimen, con un error mínimo, una proyección homográfica entre cada par consecutivo de imágenes en una secuencia. El proceso analiza la función de densidad del error de modelado de la homografía estimada a partir de un conjunto de características. Utilizando la estrategia propuesta, se desarrolla un algoritmo para estabilizar secuencias de imágenes tomadas en exteriores, en donde las condiciones lumínicas y de movimiento cambian de forma no controlada. El resto del documento está dividido de la siguiente manera: El proceso de estimación de la transformación se detalla en §2. Luego, en §3, se describe el proceso de selección de las características y las condiciones para que el proceso de selección resulte eficiente. Posteriormente, en §4, se presenta el algoritmo de estabilización de

imágenes y algunas consideraciones sobre la estabilidad numérica. En §5, se describe el proceso de prueba y validación del algoritmo, utilizando secuencias generadas artificiales y varias secuencias de exteriores. Finalmente, se discuten los resultados obtenidos.

2 Modelado de la Transformación Homográfica

Un conjunto de características F seleccionadas en la imagen I_k son utilizadas para detectar el desplazamiento de la escena, hacia la siguiente imagen I_{k+1}. El conjunto de características F se compone de las posiciones de los máximos de la superficie discreta que representa el valor del segundo valor singular del tensor estructural en una posición dada en la imágen [Jianbo y Tomasi, 1994]. El desplazamiento de cada característica puede ser calculado utilizando el método de Lucas y Kanade [Lucas y Kanade, 1981]. En este caso, la deformación de la característica se modela como un desplazamiento y una transformación *afín.*
Las nuevas posiciones de las características F, desplazadas en la siguiente imagen, se representan por F'. El movimiento global de estas características se modela por una transformación homográfica entre cada par de imágenes I_k e I_{k+1}, donde, cada elemento $\mathbf{x}_k \in F$, está relacionado como

$$\begin{pmatrix}\mathbf{x}'\\1\end{pmatrix} = H\begin{pmatrix}\mathbf{x}\\1\end{pmatrix}, \tag{1}$$

y la matriz de transformación H tiene dimensiones de 3×3 .
Los parámetros de la proyección homografía H se estiman utilizando el conjunto de características F y sus desplazamientos correspondientes F', de manera que se construye una matriz A de dimensiones $n \times 8$ como sigue

$$A = \begin{pmatrix} x_1 & y_1 & 1 & 0 & 0 & 0 & -x_1x'_1 & -x_1x'_1 \\ 0 & 0 & 0 & x_1 & y_1 & 1 & -x_1y'_1 & -x_1y'_1 \\ x_2 & y_2 & 1 & 0 & 0 & 0 & -x_2x'_2 & -y_2x'_2 \\ 0 & 0 & 0 & x_2 & y_2 & 1 & -x_2y'_2 & -y_2y'_2 \\ : & : & : & : & : & : & : & : \\ x_n & y_n & 1 & 0 & 0 & 0 & -x_ny_n & -x_ny_n \\ 0 & 0 & 1 & x_n & y_n & 1 & -x_ny_n & -x_ny_n \end{pmatrix}, \tag{2}$$

donde los primeros 8 parámetros de la matriz H se representan en forma de vector $\mathbf{h}^T = (h_{11}, h_{12}, h_{13}, h_{21}, h_{22}, h_{23}, h_{31}, h_{32})$ y se asume que $h_{33} = 1$. Entonces, utilizando mínimos cuadrados, se obtiene la solución óptima $\mathbf{h}^* = (AA^T)^{-1}A^T\mathbf{p}$ mediante la pseudoinversa usando las nuevas posiciones de las características $\mathbf{p}^T = (x'_1, y'_1, \dots, x'_n, y'_n)$. Los valores estimados en $\mathbf{h}^*$ representan los parámetros de la matriz homográfica H entre el par de imágenes I_k e I_{k+1}. Adicionalmente, el error de modelado es la diferencia entre las posiciones verdaderas y las predichas en la imagen I_{k+1}. Así pues, podemos definirlo como

$$\xi = x_i' - x_i'^*, \tag{3}$$

donde x_i' es la posición calculada y $x_i'^*$ la posición predicha de la característica que se desplaza de I_k a I_{k+1}mediante H.

3 Discriminación entre Características Fijas y Móviles

Cuando la cámara se mueve, todo en la imagen parece moverse. Nuestra propuesta para la detección de características pertenecientes a objetos estáticos consiste en analizar la función de densidad $f(\xi)$ del error de modelado de la proyección homográfica para cada par de imágenes. Así, dados el conjunto de características en la escena F, su versión desplazada F', y la homografía estimada H, se tiene que cuando el desplazamiento de F' se modela correctamente por H, y suponiendo que la función de distribución del error $f(\xi)$ para cada componente tiene una distribución normal $f(\xi_i) \sim G(\xi_i; 0, \sigma)$, (la distribución normal del error es consecuencia de la estimación por mínimos cuadrados [Kariya y Kurata; 2004] y puede ser verificado experimentalmente mediante la prueba Kolgomorov-Smirnov [Eadie *et al.,* 1971]), la discriminación de las características que pertenecen a objetos estáticos puede ser realizada mediante el siguiente procedimiento.
Considérese un conjunto de características $G \subset F$, que tienen un desplazamiento modelado por una homografía distinta H'; al estimar los parámetros de la homografía $\mathbf{h}^*$, la sumatoria de la función del error ξ se ve afectada para cada componente por

$$\Sigma_{p_j \in (F-G)}(p_j - p_j^*) + \Sigma_{p_j \in G}(p_j - p_j^*) \approx 0, \tag{4}$$

donde el primer término representa a las características que se modelan por H, y el segundo término por H'. Por las suposiciones anteriores,

ambos términos tienen una función de distribución normal $G(p_j; x, \sigma)$. El desplazamiento de la media p_j es consecuencia del error existente entre la nueva homografía estimada, y las homografías H y H', respectivamente. Cuando existe más de un desplazamiento para los datos en F, cada movimiento H'_k, representa una Gaussiana en $f(\xi)$. La probabilidad de observación depende del número de características que contiene. En general, la función de densidad se puede modelar como la mezcla de Gaussianas

$$f(\xi) \sim \sum_{i=1}^{n} \alpha_i g(\mu_i, \sigma_i), \quad (5)$$

donde el factor α_i está en función de la cantidad de elementos que conforma a cada Gaussiana. Los parámetros dependen del número y proporción del total de las características en F. Los parámetros de las Gaussianas se pueden estimar, por varias aproximaciones. Debido a que se tienen datos discretos, se considera un problema de estimación con datos incompletos. Una forma eficiente de estimar estos parámetros es mediante el algoritmo de *Expectation-Maximization* (EM) [Dempster *et al*., 1997]. Entonces, las características que se modelan por H, son distinguibles cuando la probabilidad de observar la Gaussiana $f(\xi)$ es la mayor. Estas características representan a las regiones fijas en la escena. Esto es

$$C[f(\xi)] = \mathrm{E}[f(\xi)], \quad (6)$$

donde $C[f(\xi)]$ representa la función de clasificación y $\mathrm{E}[f(\xi)]$ es la Gaussiana más probable. Entonces, para seleccionar las características que pertenecen al fondo, se construye un nuevo conjunto de características F^*, con las características que pertenecen a las Gaussianas más probables de cada componente. Adicionalmente, una característica x_i pertenece a la Gaussiana G_i con una certidumbre del 0.95 si se encuentra a lo más a $n\sigma_i$ de la media μ_i, para un $n = 2$. Asignando $F \leftarrow F^*$, se calculan nuevamente los parámetros $\mathbf{h}^*$. El proceso de selección se repite iterativamente, hasta que sólo existe una sola Gaussiana. Esto garantiza que las Gaussianas cercanas que se hayan fusionado, sean detectadas en iteraciones posteriores. Cuando se tiene una sola Gaussiana, se eligen iterativamente sólo las características con una probabilidad de 0.95 de pertenencia, permitiendo desechar aquellas que incrementan el error en la estimación de la transformación. El proceso de selección termina cuando no se descartan características en F o se ha alcanzado la precisión requerida en el rango de la función $f(\xi)$. Regularmente, las primeras iteraciones eliminan a los elementos que no siguen el movimiento dominante, luego las siguientes iteraciones eliminan a los elementos que incrementan el error de proyección en la construcción de la homografía. El pseudocódigo de este proceso se muestra a continuación.

Algoritmo 1. Selección Dinámica de Características(I_1, I_2)

Entrada: Imágenes consecutivas I_1 y I_2.
Salida: Homografía $\mathbf{h}_1$, en forma de columna, de la transformación de I_1 a I_2 y lista de características fijas F .

1. Utilizando I_1, se calcula el conjunto de posiciones $F_1 = \{(x_k^1, y_k^1) | k = 1, \ldots, n\}$ para las cuales el tensor estructural T, definido sobre un pequeño vecindario W de tamaño $(2m+1) \times (2m+1)$, con centro en (x_k, y_k), representa un máximo local sobre una superficie discreta M, resultado del segundo valor singular σ_2.
1. Las características F_1 son seguidas desde I_1 a I_2 con las nuevas posiciones desplazadas en F_2.
2. **Repetir**
   - 2.1. Una homografía $\mathbf{h'}_1$ se calcula resolviendo el sistema $\mathbf{h'}_1 = (AA^T)^{-1}A^T\mathbf{p}$, para A y $\mathbf{p}$. Un error $\xi = x_i - x_i^*$, que representa las diferencias entre la posición actual x_i de las características I_2 y las predichas por la homografía x_i^*, es calculado.
   - 2.2. Asumiendo que ξ es de la forma de la ec. (5) para cada componente, las características $G_2 \subset F_2$ que no corresponden a la Gaussiana con mayor probabilidad son descartadas de F_1 y F_2formando un nuevo conjunto de características F'_1 y F'_2 respectivamente.
   - 2.3. Se actualiza $F_1 \leftarrow F'_1$ y $F_2 \leftarrow F'_2$.

Hasta que exista una única moda en $f(\xi)$ y el rango sea menor un umbral λ.
3. Hacer $\mathbf{h}_1 \leftarrow \mathbf{h'}_1$.

La selección dinámica requiere la estimación de una transformación homográfica. La condición para que se pueda estimar la homografía es que el producto AA^T sea invertible. Esta condición se cumple si el $Rank(A) = 8$. Esto se consigue si al menos se tienen 4 puntos. Además, para obtener

una solución satisfactoria sobre el modelado de la homografía, las características deben estar distribuidas uniformemente sobre el campo de visión de la cámara, es decir no ser colineares. En nuestro caso, la colinealidad se mide mediante el ajuste por SVD de una recta [Tomasi, 2004]. La magnitud del residuo del error determina si los datos son colineares o no. Entonces, para el conjunto de características fijas F en la imagen I_k , la recta que mejor ajusta a los datos se logra mediante $Q = P - \mathbf{1}\mathbf{p}^T$, donde P es una matriz de $2 \times n$ que contiene la posición (x_i, y_i) de cada característica, $\mathbf{p} = \frac{1}{n}P\mathbf{1}$ es el centroide, $\mathbf{1}$ es un vector de $1 \times n$ elementos. Factorizando Q, se tiene que $Q = S\mathbf{\Sigma}V^T$, donde el segundo valor singular σ_2 de $\mathbf{\Sigma}$ representa el residuo de la ecuación homogénea $Q\mathbf{v} = 0$ con $\mathbf{v^T} = (a, b)$. Cuando el valor de σ_2 es demasiado pequeño, indica geométricamente que los puntos F representan una recta, mientras que cuando adopta un valor grande, indica el eje de dispersión de los datos en la segunda componente ortogonal. Entonces, la dispersión de ambas componentes ortogonales debe ser similar a las dimensiones de la imagen para lograr una distribución uniforme. Finalmente, una distribución adecuada se tiene cuando al comparar la proporción de los valores singulares con respecto a las dimensiones de la imagen son similares. Esto es, para las dimensiones $(n \times m)$ de la imagen, la proporción $\frac{\sigma_2}{\sigma_1} \approx \frac{n}{m}$ se cumple, en caso contrario, las características no están distribuidas y la homografía estimada no es fiable.

4 Una Aplicación: Algoritmo de Estabilización

Aquí desarrollamos la aplicación del algoritmo de selección dinámica de características anteriormente descrito al problema de estabilizar una secuencia de imágenes tomada desde una cámara fija sujeta a vibración. Un algoritmo eficiente de estabilización debe de eliminar los movimientos causados por el desplazamiento abrupto en la cámara, de manera que la escena conserve la consistencia espacial. Empleando un proceso de selección dinámica de características fijas en la escena, se estiman las proyecciones en el tiempo del desplazamiento de la cámara. En conjunto, las proyecciones estimadas modelan la deformación de la imagen referenciándola con su antecesora, proporcionando consistencia espacial. Consecuentemente, en una secuencia de imágenes al tomarse como referencia la primera imagen I_0, es posible obtener las transformaciones que proyectan a cualquier imagen I_n a la imagen de referencia multiplicando cada una de las homografías estimadas para cada par de imágenes.

4.1 Definición del Algoritmo

El algoritmo de estabilización trabaja sobre una secuencia de imágenes. Para cada imagen subsecuente, se calcula la homografía de I_j a I_{j+1}, de manera que la imagen I_{j+1} es mapeada a la imagen de referencia I_0, empleando las homografías calculadas $H_j^* = H_0 H_1 \dots H_j$. El proceso de estabilización es exitoso siempre y cuando se calcule una proyección fiable para cada par de imágenes. El pseudocódigo del proceso se muestra a continuación.

Algoritmo 2. Estabilización de una Secuencia de Imágenes(I_1, I_2)

Entrada: Imágenes consecutivas I_j y I_{j+1} y la homografía $\mathbf{h}$ que mapea de la primera imagen a la actual.
Salida: Homografía $\mathbf{h}_{j+1}$, que describe la transformación de la imagen j a la $j + 1$.

1. Mapear las imágenes I_j e I_{j+1} usando H_j, que corresponde a $\mathbf{h}_j$ en forma de matriz para obtener I'_j e I'_{j+1}; ambas ahora estan expresadas en función de la primera imagen de referencia en el sistema.
2. Con I'_j, I'_{j+1} y $\mathbf{h}_j$, usar el algoritmo $\mathbf{h}_j \leftarrow$ Selección Dinámica de Características(I_1, I_2).

La complejidad de la selección dinámica de características depende del cálculo de los tensores, el cual, tiene una complejidad $O(n_1^2)O(n_2^2)$, al recorrer y calcular los tensores con una ventana de dimensión $n_2 \times n_2$, para $n_2 = 2m + 1$. Además, el cálculo del desplazamiento de cada característica tiene una complejidad $k_1 O(n_3^2)$. La constante k_1 está en función de la condición de paro donde $k_1 = \min(\{maxIter, minError\})$. Cada iteración del algoritmo evalúa al menos $|F|^*$ características, donde $|F|^*$ es el número estimado de características en la imagen. Entonces, la complejidad para una iteración es $k_1 O(n_3^2)|F|^*$. El proceso de la selección de características fijas itera a lo más $|F|^* - 4$ veces (condición mínima para que el sistema tenga solución). Por cada iteración,

se evalúan $|F|^*$ características y se estima una nueva homografía con una complejidad del orden $O(|F|^{*2})$. La complejidad del proceso de estabilización es la suma de la complejidad del proceso de selección dinámica de características y el proceso de proyección de imágenes. El proceso de proyección añade una complejidad $k_2(O(n^2))$ donde $k_2 = 9$. Cuando se interpolan los datos, la complejidad es $k_2(O(n^2))(O(m^2))$ dependiendo del tamaño del vecindario m. Sumarizando, la complejidad de los procesos de la selección dinámica y la estabilización de imágenes, está determinada por

$$\begin{aligned} &C_{Selección}(n_1, n_2, n_3) \\ &= \underbrace{O(n_1^2)O(n_2^2)}_{\text{Tensor}} + \underbrace{k_1 O(n_3^2)|F|^*}_{\text{Seguimiento}} \\ &+ \underbrace{(O(|F|^*) + O((|F|^*)^2)(|F|^* - 4)}_{\text{Selección de Características Fijas}} \end{aligned} \tag{7}$$

$$\begin{aligned} C_{estabilización}(n_1, n_2, n_3, n_4) \\ = C_{Selección}(n_1, n_2, n_3) \\ + \underbrace{k_2 O(n_1^2)O(n_4^2)}_{\text{Proyección}}. \end{aligned}$$

4.2 Degradación Numérica

Para detectar el grado de error del sistema se compara el estado actual contra estados en su pasado lejano. Si los estados presentan diferencias considerables, el sistema se ha degradado. Para detectar la degradación numérica del sistema, se emplea la homografía H_j^* asociada a la imagen I_j, la homografía distante H_i^*, asociada a la imagen I_i, las imágenes I_i' y I_j' que son la proyección homográfica de I_i y I_j respectivamente. Con las imágenes I_i' e I_j' se calcula la transformación homográfica H_{ij} mediante el Algoritmo 1. Entonces, la homografía H_{ij} debe de satisfacer que el $det(H_{ij}) = 1$ porque H_{ij} debe de ser la matriz identidad. Esta aseveración es válida, dado que la proyección I_i' y I_j' son proyecciones en el mismo espacio. Cuando el determinante no es igual a la unidad el sistema está en un estado inconsistente. Entonces, cuando el valor absoluto de la diferencia de 1 y el $det(H_{ij})$ sea mayor a un umbral preestablecido, se sustituye la homografía actual H_j^* con la homografía $H_{j'}^*$. Esta homografía mapea entre la imagen I_j a la imagen I_i'.

5 Modelo Experimental y Discusión de Resultados

Se ha desarrollado un modelo experimental para evaluar el funcionamiento de la estrategia para la selección dinámica de características. El modelo consta de tres etapas: la etapa de validación del funcionamiento de la estrategia de selección dinámica de características, y la etapa de prueba y verificación del funcionamiento en ambientes en exteriores.

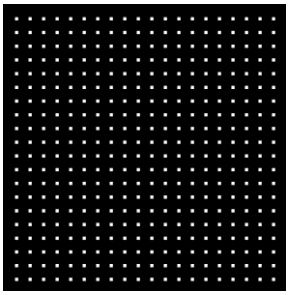


Fig. 1. Patrón artificial utilizado para evaluar el desempeño del algoritmo. Los cuadrados blancos representan a los objetos en la escena

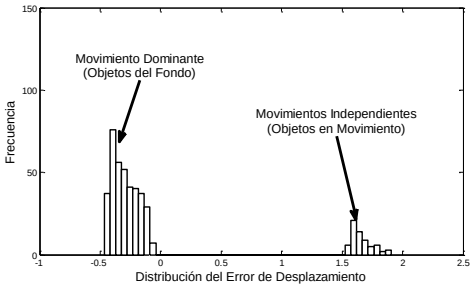


Fig. 2. Distribución del error en imágenes artificiales. El máximo global corresponde a los objetos en el fondo. Los máximos locales a movimientos de objetos en movimiento

5.1 Validación del Algoritmo

El proceso de validación cuantifica el error de la matriz homográfica estimada para cada par de imágenes. El nivel de eficiencia se evalúa por el Error Cuadrático Medio (RMES) [Anderson y Woessner, 1992] de las características fijas encontradas en una imagen al proyectarlas a una segunda. El proceso de validación utiliza un patrón

artificial (ver Fig. 1) y un conjunto de imágenes de exteriores donde se les induce desplazamiento aleatorio, simulando la vibración.

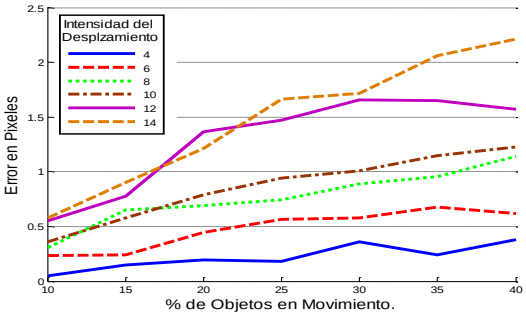


Fig. 3. Gráfica de error RMSE en las imágenes generadas artificialmente y con ruido blanco inducido. La intensidad del ruido y el porcentaje de objetos en movimiento se han variado

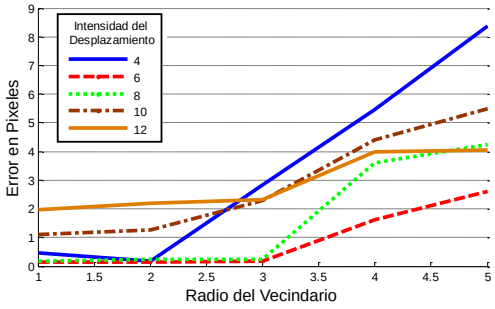
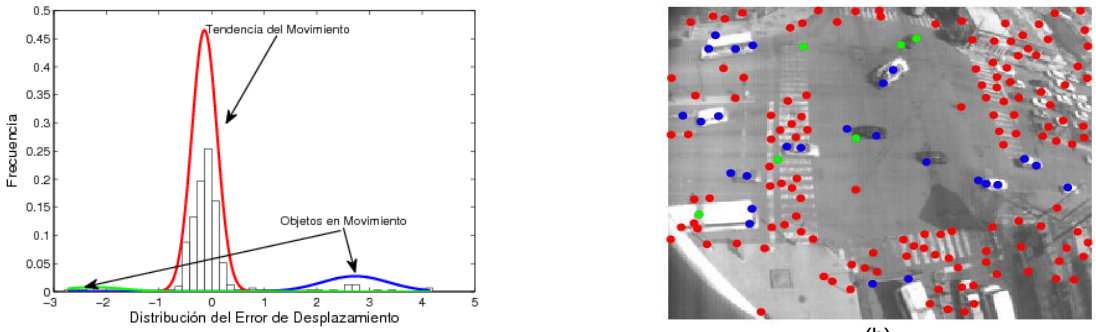


Fig. 4. Gráfica del RMSE de la estimación del desplazamiento en una secuencia corta para validar el algoritmo. El ruido ha sido inducido como un desplazamiento en el plano. Se ha variado la intensidad del desplazamiento y el radio del vecindario de las características candidatas. Los radios y las intensidades están expresados en pixeles

El patrón artificial consiste en un conjunto de marcas de 5×5 pixeles distribuido uniformemente. Para simular el efecto de la vibración, se desplazan las marcas aleatoriamente. Las marcas se dividen en dos grupos: los objetos fijos en la escena, y los objetos con movimiento. Ambos grupos se desplazan con un movimiento independiente uno del otro. El experimento se repite 100 veces por cada combinación de parámetros, variando la proporción de objetos en movimiento (de 10 % a 40% con intervalos de 5%), la intensidad del desplazamiento (de 4 a 14 pixeles con incrementos de 2) y utilizando un vecindario de 4×4 pixeles. El error es considerablemente pequeño a pesar de la intensidad del desplazamiento (Fig. 3). Pero, resulta más sensible a incrementos en el desplazamiento que a la proporción de objetos fijos y en movimiento. El error de la homografía estimada por la aproximación es pequeño a pesar que se incrementa el número de características en movimiento. El proceso de selección de características $C[f(\xi)]$ resulta eficiente, porque la distribución de los objetos es identificable en la función de densidad de ξ (ver Fig. 2).

En la segunda etapa, con una muestra de 100 imágenes libres de vibraciones tomadas de un crucero, se simula un medio con vibraciones induciendo desplazamientos aleatorios a cada imagen. La intensidad del desplazamiento varía de 2 hasta 10 pixeles en incrementos de 2. El radio r del vecindario alrededor de cada $\mathbf{x}'_i \in F$, se varía de 1 a 5 con incrementos de 1. En los resultados (Fig. 4) se observa un error ligeramente mayor al obtenido con las imágenes artificiales. El tamaño del vecindario incide en el grado del error. En vecindarios pequeños no existe información de textura suficiente para estimar el desplazamiento, y en vecindarios grandes se tiene el riesgo de utilizar zonas que no corresponden al objeto. Los resultados muestran que la estrategia propuesta estima en forma adecuada la transformación entre las imágenes, en situaciones controladas. El proceso de selección de características se ilustra en la Fig. 5 (a)-(b). En la Fig. 5 (a) se muestra la función de densidad $f(\xi)$. Cada Gaussiana estimada se representa con un color distinto. La función de densidad ha sido normalizada de acuerdo a la cantidad de elementos que contiene cada Gaussiana. En la Fig. 5 (b) se muestra la imagen y las regiones codificadas en color de acuerdo a la Gaussiana que corresponden. La más probable corresponde a regiones del fondo. Si el rango es grande se desechan los elementos que no contribuyen a minimizar el error. Estos resultados muestran que la propuesta es capaz de identificar eficientemente los objetos fijos en imágenes de escenarios reales. El funcionamiento en escenarios reales, con cámaras sujetas a vibración se discuten en el siguiente apartado.



(a) (b)

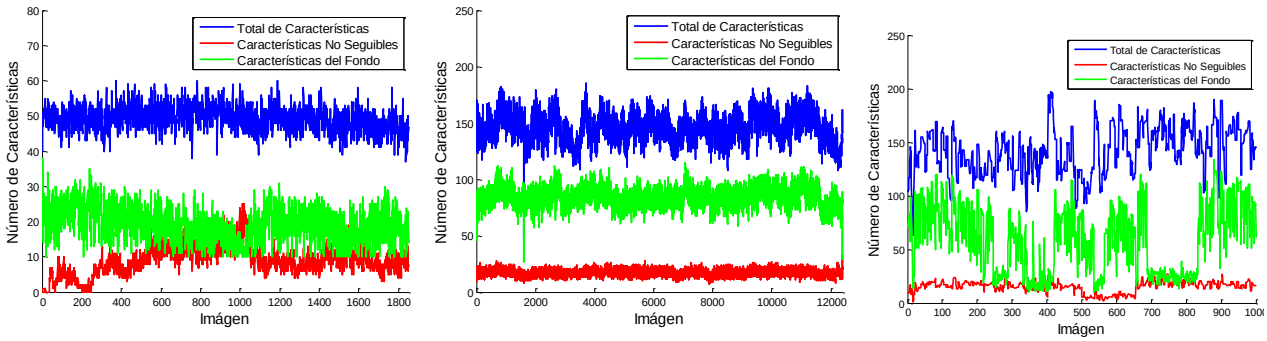
Fig. 5. Distribución del error de transformación: (a) Histograma de la distribución del error en el mapeo modelado como Gaussianas; (b) imagen con las regiones utilizadas codificadas en color

Tabla 1. Parámetros de las secuencias de imágenes usadas

Atributo	Edificio	Crucero	Avenida
Longitud de la	12,000	1,800	1,000
Resolución	320 × 240	320 × 240	360 × 242
Imágenes por	30	≈	≈
Nivel de	Ninguna	0.75	0.50
Transmisión de	Cable	Cable	Microondas

5.2 Ambientes Exteriores

Se utilizan tres escenarios diferentes (ver Tabla 1). La secuencia del edificio representa situaciones con excesivas rotaciones y vibraciones (intensidades de hasta 10 pixeles), y una distribución no uniforme de las regiones con textura. La secuencia del crucero muestra condiciones lumínicas y de movimiento cambiantes constantemente. Los objetos con movimiento incluyen vehículos, bicicletas, y peatones. En la escena existen reflejos ocasionados por vehículos, paso de nubes, y lluvia. La avenida rápida, representa situaciones de baja calidad de adquisición y un nivel de acercamiento grande (que incrementa las vibraciones), causando una pérdida de la información de la textura. En todas las secuencias la vibración es originada por el viento y el paso de vehículos.

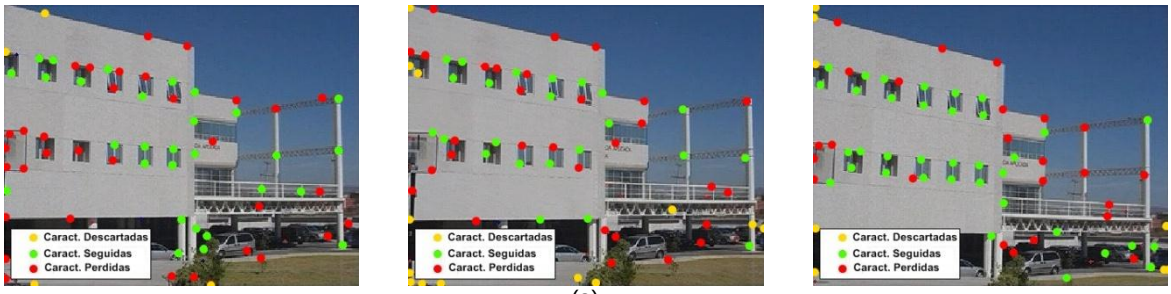


a) b) c)

Fig. 6. Gráfica de la evolución de las características seguibles en el tiempo para la secuencia (a) del edificio, (b) un crucero y (c) una avenida rápida. El total de características seguibles se mantiene en constante cambio (en color azul), donde un gran número de ellas son adecuadas para referenciar la imagen (en color verde) y sólo un porcentaje relativamente pequeño son características que no puede estimarse en forma adecuada el desplazamiento (en color rojo)

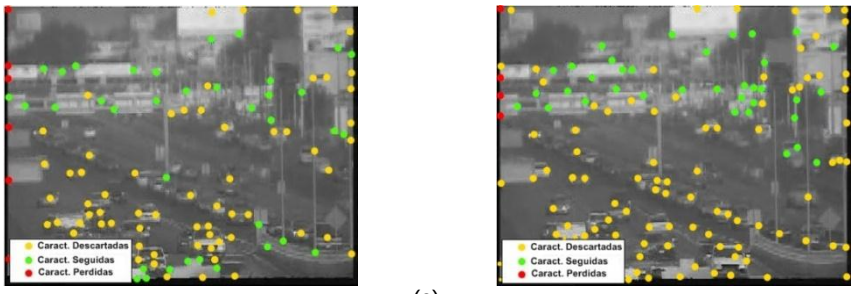
La Fig. 6 muestran el número de características encontradas y seleccionadas para cada una de las secuencias. Del total de características (líneas azules), se determinan aquellas que permanecen fijas (líneas verdes), descartando aquellas que permanecen fijas y aquellas que no pudieron seguirse (líneas rojas). Las oscilaciones existentes en el total de regiones de la escena y las seleccionadas, se deben a las variaciones lumínicas, a la cantidad de objetos en movimiento (reflejos, lluvia o nubes) y al nivel de compresión de las imágenes. En el caso de la avenida rápida, las condiciones del video, causa que contenga menos características fijas detectadas. En la Fig. 7 se muestran situaciones representativas de cada secuencia. A pesar que existen desplazamientos grandes y una distribución no uniforme de la textura (Fig. 7 (a)), o cambios lumínicos provocados por lluvia o reflejos (Fig. 7 (b)) o una baja calidad en la imagen (primera imagen de la Fig. 7 (c)), el algoritmo es capaz de determinar las características fijas, logrando estimar la homografía entre cada par de imágenes. Los cambios lumínicos repentinos y los reflejos son tolerados por el continuo cálculo de los tensores (Fig. 7 (b)). Cuando las características son insuficientes o no están correctamente distribuidas para estimar la homografía, la medida de fiabilidad determina la confiabilidad de la homografía estimada. Las dos imágenes de la Fig. 7 (c) muestran dos casos cuando la fiabilidad es adecuada y cuando no lo es. En la primera imagen, la proporción de los eigenvalores de la distribución de las características es 0.6113, que es cercano a la proporción entre las dimensiones de la imagen $\left(\frac{242}{360} = 0.6722\right)$, mientras el segundo caso, la proporción es de 0.2846, indicando que la homografía estimada no es fiable.

Cuando las regiones en movimiento son mayores proporcionalmente a las regiones fijas, los objetos que pertenecen al fondo, no son distinguibles de función de densidad $f(\xi)$ por una mezcla de Gaussianas. En la Fig. 8, se ilustra un escenario que no es posible determinar las características del fondo. Las flechas azules ilustran la dirección del movimiento de los vehículos. La función de densidad $f(\xi)$ (Fig. 8 (a)) presenta las Gaussianas estimadas. Para solucionar estas situaciones se tienen dos alternativas. La primera consiste en sustituir el criterio de selección $C[f(\xi)]$, utilizando información conocida de la escena. La segunda consiste en seleccionar regiones, más pequeñas, que tengan una proporción mayoritaria de objetos fijos.



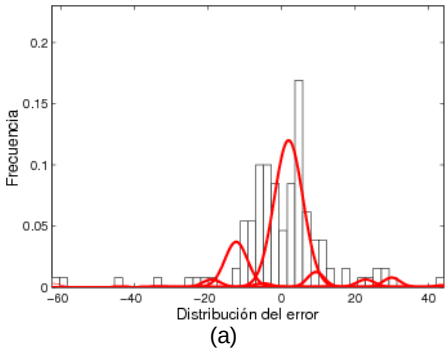
(a)

(b)

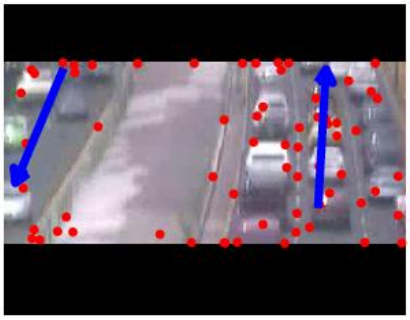


(c)

Fig. 7. Ejemplo de situaciones de interés en las secuencias analizadas: (a) Ejemplo de secuencia con grandes desplazamientos, las dos primeras imágenes tienen una diferencia de $\frac{1}{10}$ segs. y la tercera 30 segs. después.; (b) Ejemplo de cambios lumínicos generados por lluvia y reflejos causados por los vehículos, las dos primeras imágenes se han tomado con una diferencia de 1 seg.; (c) Ejemplos de imágenes con una distribución adecuada y no adecuada de características seguibles. En la primera imagen, se tiene un índice de fiabilidad de 0.6113 y la segunda de 0.2846



(a)



(b)

Fig. 8. Escenarios donde no es posible aplicar la propuesta. En (a) se muestra la distribución del error que no puede modelarse eficientemente como mezcla de Gaussianas; en (b) se muestra la imagen asociada con las características encontradas. La proporción de características fijas es menor a la que tienen movimiento

5.3 Comparativa contra otras aproximaciones

Finalmente, la estrategia de selección dinámica de características se compara contra RANSAC [Fischler y Bolles, 1981], como un método ampliamente aceptado y robusto para estimar la proyección en un par de imágenes. La eficiencia se compara utilizando el error cuadrático en la proyección en cada par de imágenes. Este error mide la precisión de la homografía estimada al proyectar cada par de imágenes en cada aproximación. El proceso se aplica a cada una de las secuencias de referencia. Los parámetros utilizados en la propuesta son: un rango mínimo de error de 0.05 pixels y un criterio de pertenencia de 2.5σ; y en RANSAC son de 10 iteraciones y un criterio de aceptación de 1 pixel.

El algoritmo de RANSAC es muy eficiente computacionalmente (Fig. 9 (a)-(c)) y el error es pequeño en las tres secuencias. Sin embargo, con la propuesta el error es menor (Fig. 9 (a)-(c)). En promedio, el error es menor a 0.05 pixeles en comparación de 0.4 con RANSAC. La diferencia entre los resultados se debe a que RANSAC selecciona aleatoriamente una muestra y asume que esta no contiene ruido y objetos en movimiento, pero a costa de una complejidad computacional menor. En cambio, la propuesta analiza y selecciona aquellas características que minimizan el error, ganando precisión, pero a costa de una mayor complejidad.

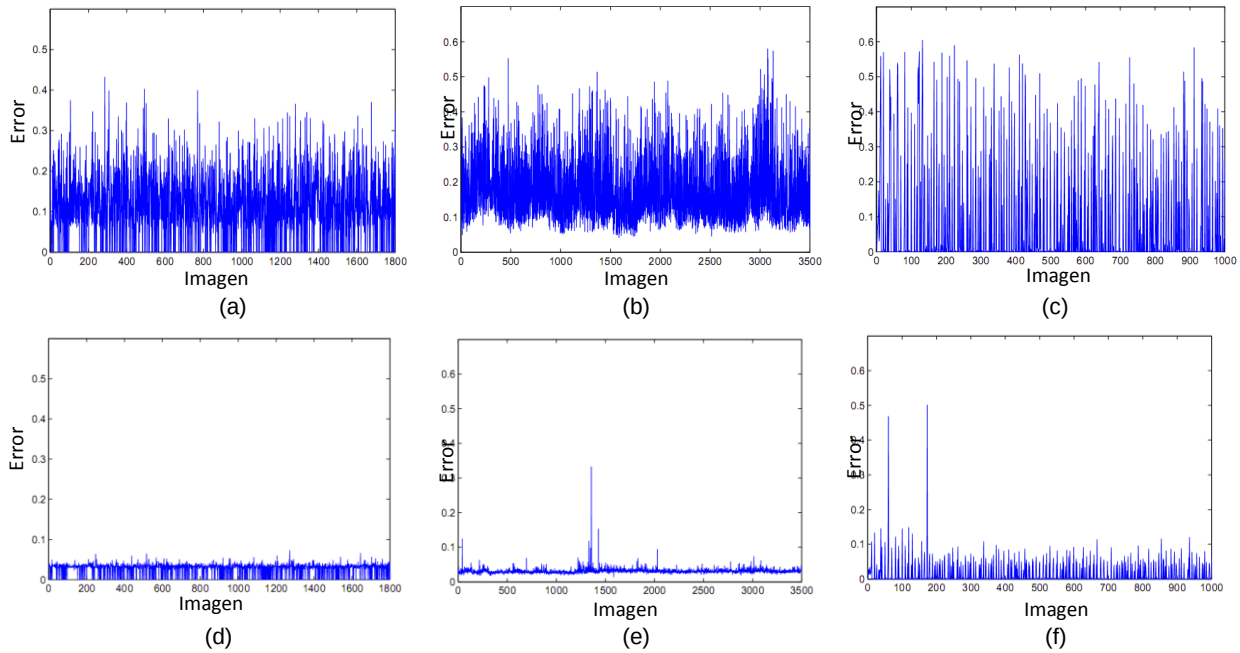


Fig. 9. Error cuadrático medio de las características entre el mapeo de un par de imágenes utilizando una transformación homográfica estimada por (a) RANSAC y por (b) nuestra propuesta. Las escalas de ambas gráficas son iguales con propósitos comparativos. El error utilizando RANSAC ofrece un desempeño adecuado con un error máximo de 0.6 pixeles. Pero, nuestra propuesta, en lo general ofrece un error menor a 0.1 pixeles

6 Conclusión

En este trabajo se ha presentado una estrategia para seleccionar dinámicamente un conjunto de características que permanece temporalmente fijas en la escena. El funcionamiento de la propuesta se ha evaluado en escenarios artificiales y en escenarios de exteriores obteniendo resultados satisfactorios. Así se ve como el uso dinámico de características y el criterio de selección de regiones fijas ha permitido que el algoritmo tenga un funcionamiento robusto y adaptable a condiciones cambiantes de la escena, tales como iluminación. También, al estimar una transformación homográfica entre cada par de imágenes, además de eliminar las vibraciones, se tiene una consistencia espacial de cada una de las imágenes de la secuencia. La calidad de la homografía estimada ha sido comparada contra el algoritmo de RANSAC. Nuestra propuesta, en los escenarios probados, obtiene mejores resultados al estimar la homografía entre cada par de imágenes.

Referencias

1. **Anderson, M. P. & Woessner, W. W. (1992).** *Applied Groundwater Modeling: Simulation of Flow and Advective Transport, San Diego, California,* Academic Press.
2. **Collins, R. T. & Lui, Y. (2003).** On-Line Selection of Discriminative Tracking Features, *Ninth IEEE International Conference on Computer Vision*, Nice, France, 346-352.
3. **Cham, T. & Rehg, J. (1999).** Dynamic Feature Ordering for Efficient Registration. *Seventh IEEE International Conference on Computer Vision*, 2, 1084-1091.
4. **Chum, O. & Matas**, **J. (2008).** Optimal Randomized RANSAC. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(8), 1472–1482.
5. **Dempster, A., Laird, N. & Rubin**, **D. (1997)**. Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society*, vol. 1, 1–38.
6. **Eadie, W., [et. all] (1971).** Statistical Methods in Experimental Physics. *Amsterdam: North-Holland.*
7. **Fischler, M. & Bolles, R. (1981).** Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. *Communications of the ACM*, 24(6), 381–395.

8. **Fleuret, F. (2004).** Fast Binary Feature Selection with Conditional Mutual Information. *Journal of Machine Learning Research*, 5, 1531-1555.
9. **Gil, S. Milanese, R. & Pun, T. (1994).** *Feature Selection for Object Tracking in Traffic Scenes*. (tr-94-060) International Computer Science Institute, California, USA.
10. **Hwann-Tzong, Ch., Tyng-Luh, L., & Chiou-Shann F. (2004).** Probabilistic Tracking with Adaptive Feature Selection. *17th International Conference on Pattern Recognition*, 2, 736-739.
11. **Jianbo, S. & Tomasi**, **C. (1994).** Good Features to Track*. IEEE Computer Society Conference on Computer Vision and Pattern Recognition,* Seattle, Washington, USA, 593–600*.*
12. **Julier, S. J. & Uhlmann, J. K. (1997).** A New Extension of the Kalman Filter to Nonlinear Systems*. 11th International Symposium on Aerospace/Defense Sensing, Simulation and Controls Orlando,Florida, USA,* 182-193.
13. **Junlan, Y.J., Schonfeld, D., Chong, Ch.. & Mohamed, M. (2006).** Online Video Stabilization Based on Particle Filters*. IEEE International Conference on Image Processing,* Atlanta, Georgia, USA, 1545-1548.
14. **Kalman, R. E. (1960)** A New Approach to Linear Filtering and Prediction Problems*. Transactions of the ASME Journal of Basic Engineering,* 82 (Series D), 35–45.
15. **Kariya, T. & Kurata**, **H. (2004).** Generalized Least Squares, Hoboken, NJ: Wiley.
16. **Knot J., Jurišica L. & Hubinsky P. (2006).** Feature Based Object Tracking for Oscillation Detection. *16th International Conference Radioelectronika 2006,* 316-319.
17. **Lacey, A.J., Pinitkarn, N. & Thacker, N.A (2000).** An Evaluation of the Performance of RANSAC Algorithms for Stereo Camera Calibrarion*. 11th BritishMachine Vision Conference, Bristol* U.K*.*
18. **Lucas, B.D. & Kanade**, **T. (1981).** An Iterative Image Registration Technique with an Application to Stereo Vision*. DARPA Image Understanding Workshop, San Francisco, CA, USA* 121-130.
19. **Matsushita, Y., Ofek, E., Weina, G., Xiaoou, T. & Heung-Yeung, S. (2006).** Full-Frame Video Stabilization with Motion Inpainting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28(7), 1150 – 1163.
20. **Mikolajczyk, K. & Schmid**, **C. (2005).** A performance evaluation of local descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(10),1615–1630.
21. **Salas, J., Jiménez, H., Gonzalez, J. & Hurtado**, **J. (2007).** Detecting Unusual Activities at Vehicular Intersections". *IEEE International Conference on Robotics and Automation*, Roma, Italy, 864–869.
22. **Tan, Y**., **Kulkarni, S. & Ramadge**, **P. (1996).** Extracting Good Features for Motion Stimation, *International Image Proccesing,* Lausanne, Switzerland, 117 – 120.
23. **Tang, F. & Tao**, **H. (2005).** Object Tracking with Dynamic Feature Graph, *2nd Joint IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance,* Beijin, China25-32.
24. **Thomas, G. B. & Finney, R. L. (1992).** *Calculus and Analytic Geometry*. (8th Ed.). Reading Mass: Addison Wesley.
25. **Tomasi C. (1993).** Input Redundancy and Output Observability in the Analysis of Visual Motion*. Sixth International Symposium on Robotics Research, Ithaca, NY, EUA,* 213-222*.*
26. **Tomasi, C. (2004).** *Cs 296.1Mathematical Modelling of Continuos Systems.* Durham, NC:Duke University*.*
27. **Tomasi, C. & Manduchi, R. (1998).** Stereo Matching as a Nearest-Neighbor Problem. *IEEE Transactions on Pattern Analysis and Machine Intelligence,*20(3), 333–340.
28. **Trucco, E. & Verri, A. (1998).** *Introductory Techniques for 3-D Computer Vision*. Upper Saddle River, NJ: Prentice Hall.
29. **Wald, A. (1945).** Sequential Tests of Statistical Hypotheses. *Annals of Mathematical Statistics*, 16(2), 117–186.
30. **ZuWhan, K. (2008).** Real Time Object Tracking Based on Dynamic Feature Grouping with Background Subtraction. *IEEE Conference on Computer Vision and Pattern Recognition*, Anchorage, Alaska, USA 1-8.

Hugo Jiménez Hernández

Ingeniero en Sistemas Computacionales en el Tecnológico Regional de Querétaro, Maestro en Ciencias de la Computación en el Centro de Investigación en Computación del IPN, México DF. y actualmente candidato a Doctor en el Centro de Investigación en Ciencia Aplicada y Tecnología Avanzada Unidad Querétaro, del IPN. Las líneas de investigación incluyen la detección automática de actividades, memorias asociativas y análisis de series de tiempo.

Joaquín Salas Rodríguez

Estudio la maestría en Ingeniería Eléctrica en el Centro de Investigación y de Estudios Avanzados del IPN, y cuenta con un Doctorado en Informática por el Instituto Tecnológico de Estudios Superiores de Monterrey. Su área de especialidad es Análisis de Imágenes. Ha sido profesor del CICATA desde 1997. Ha publicado 40 artículos en congresos y revistas internacionales en el área de Análisis de Imágenes. Es miembro del Sistema Nacional de Investigadores desde 1995.