

Segmentación y posicionamiento de sillas de ruedas en espacios inteligentes mediante minimización de funciones de energía

Losada, Cristina; Mazo, Manuel; Palazuelos, Sira; Blanco, Edward

Departamento de Electrónica. Universidad de Alcalá.
Alcalá de Henares, Madrid, España.
Tlf. +34918856592 Fax: +34918856591
losada@depeca.uah.es;

Resumen

En este trabajo se aborda la segmentación y posicionamiento de objetos móviles (y en especial sillas de ruedas) en espacios inteligentes a partir de las imágenes captadas por una o varias cámaras fijas. Se propone el uso de GPCA (*Generalized Principal Components Analysis*) para la inicialización de un conjunto de curvas que definen los contornos de los objetos detectados y la posterior minimización de una función de energía para la segmentación y posicionamiento de las sillas de ruedas. Esta propuesta ha sido implementada en un “espacio inteligente” y se han realizado diversas pruebas que han demostrado la validez de la misma así como la notable mejora que supone en cuanto a la precisión de la segmentación y el tiempo de procesamiento frente a otras propuestas en las que se utiliza únicamente la minimización de energía para la segmentación de movimiento.

1. Introducción

La segmentación de movimiento es una tarea fundamental para el análisis de secuencias de imágenes digitales siendo de gran interés en una amplia gama de aplicaciones de la visión artificial, tales como vigilancia, robótica móvil, etc. En la literatura se presentan diferentes técnicas que tratan de analizar y extraer información acerca del movimiento de los objetos en una escena. Las primeras aproximaciones son las basadas en discontinuidades del flujo óptico (Black and Anandan 1991). Estas técnicas presentan inconvenientes derivados del problema de apertura, así como de la presencia de ruido en las estimaciones del flujo óptico. Tradicionalmente, la segmentación de movimiento ha estado ligada con la detección de movimiento de forma que, cada región de una imagen corresponde a un modelo de movimiento que explica los cambios temporales en dicha región (Weiss 1996). Otras alternativas para segmentación de movimiento son las

basadas en *clustering* (Costeira and Kanade, 1998), (Kanatani 2001) En otra línea, en (Sekkti and Mitiche 2006a) se propone obtener la segmentación de movimiento rígido 3D mediante la minimización de una función de energía. Esta técnica permite recuperar la posición 3D de los objetos en movimiento, pero presenta algunas limitaciones importantes ya que tanto los resultados obtenidos como el número de iteraciones necesarias hasta la convergencia dependen fuertemente de la inicialización. Además, se trata de un algoritmo lento, ya que requiere entre 700 y 10.000 iteraciones antes de alcanzar la convergencia (Sekkti and Mitiche 2006b).

En este trabajo se propone un algoritmo robusto de segmentación de movimiento rígido 3D a partir de las imágenes captadas por una o varias cámaras fijas en el entorno basado en la minimización de una función de energía.

A continuación, en la sección 2 se describe el sistema propuesto cuyas etapas se explican con mayor detalle en los apartados 3 y 4 para, finalmente, en la sección 5 presentar los resultados obtenidos con el sistema.

2. Sistema propuesto

En la Fig. 1 se presenta un diagrama general del sistema propuesto en el que se muestran las diferentes etapas de dicho sistema. En este diagrama se puede observar que, el algoritmo de segmentación de movimiento y obtención de la posición 3D se aplica de forma independiente para las imágenes captadas por cada una de las cámaras fijas en el entorno. Incluyéndose la posibilidad de realizar un proceso de fusión de datos de las n cámaras que permita mejorar la precisión de la segmentación.

La segmentación de movimiento de sillas de ruedas y otros objetos móviles se obtiene minimizando una función de energía (1) definida a partir de la restricción de brillo 3D (Sekkti and Mitiche 2006a) en función de los contornos que

definen la segmentación $\{\gamma_k\}_{k=1}^{N-1}$ las componentes de velocidad 3D, lineal $\{\mathbf{T}_k\}_{k=1}^N$ y angular $\{\omega_k\}_{k=1}^N$ y la distancia del objeto a la cámara (Z).

$$E[\{\gamma_k\}_{k=1}^{N-1}, \{\mathbf{T}_k\}_{k=1}^N, \{\omega_k\}_{k=1}^N, Z] = \sum_{k=1}^N \left[\int_{\Omega_k} \psi_k^2(\mathbf{x}) d\mathbf{x} + \mu \int_{\Omega_k} g(\|\nabla Z\|) d\mathbf{x} \right] + \sum_{k=1}^{N-1} \lambda \oint_{\gamma_k} ds \quad (1)$$

En (1) λ y μ son constantes reales y positivas que ponderan la contribución de cada uno de los términos de la ecuación y ∇ el operador gradiente espacial.

Para la minimización de (1) se emplea un algoritmo *greedy*, iterativo de tres etapas que requiere de valores iniciales tanto para la distancia del objeto a la cámara (Z) que se supondrá igual para todos los puntos de la imagen (y se inicializa como un plano a una distancia D de la cámara) como para los contornos que definen la segmentación. Debido a que la inicialización de las curvas influye de forma notable en los resultados se propone utilizar la técnica GPCA para la inicialización de las curvas, según se explicará más adelante.

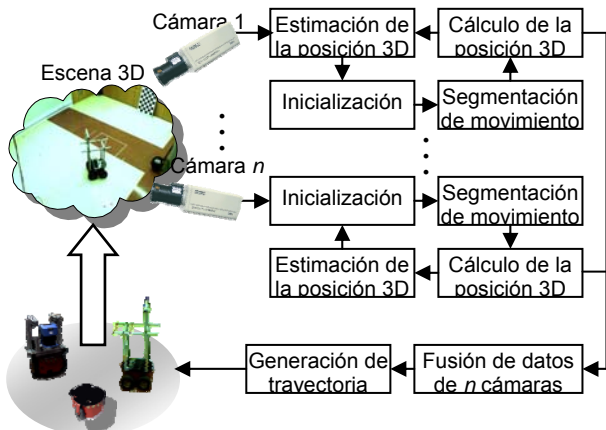


Fig. 1. Diagrama de bloques general del sistema

Además de la segmentación de movimiento, la minimización de (1) permite obtener las componentes de velocidad 3D y la distancia del objeto a la cámara (Z) de forma que, conociendo la calibración de las cámaras, es sencillo obtener la posición 3D de los objetos móviles.

Por otro lado, se ha implementado un filtro de *Kalman* por cada uno de los objetos móviles de forma que, en cada iteración se estima la posición que ocupará el objeto móvil en la siguiente imagen.

3. Inicialización de las curvas

En (Sekkati and Mitiche 2006a) se propone utilizar circunferencias para la inicialización de las curvas, sin embargo, en este trabajo, en el que la

segmentación se realiza a partir de las imágenes tomadas por cámaras fijas en un entorno interior, se propone modelar el fondo de la escena a partir de un conjunto de imágenes del mismo y, posteriormente comparar cada par de imágenes de entrada con dicho modelo. Tanto para la obtención del modelo de fondo como para la comparación de las imágenes de entrada con ese modelo se emplea GPCA (Jieping et al. 2004).

Dado un conjunto de imágenes en las que únicamente aparece el fondo de la escena $\{\mathbf{I}_j\}_{j=1}^N$, GPCA permite obtener dos matrices de transformación \mathbf{L} y \mathbf{R} las cuales permiten proyectar cualquier imagen de fondo a un espacio transformado de dimensión inferior a la del espacio de entrada y, posteriormente recuperar la imagen original conservando las características principales de la misma. El cálculo del modelo de fondo conlleva un alto coste computacional pero sólo es necesario realizarlo una vez, pudiendo además hacerlo *off-line*.

El modelo de fondo se compara con cada par de imágenes a segmentar para determinar qué píxeles son candidatos a pertenecer a objetos que han entrado en la escena después de haber tomado las imágenes del fondo. Estos elementos pueden ser tanto sillas de ruedas, como posibles obstáculos en movimiento. Para la comparación de las nuevas imágenes con el modelo de fondo en primer lugar se proyecta la imagen al espacio transformado definido por las matrices \mathbf{L} y \mathbf{R} obtenidas anteriormente (2) y a continuación se procede a recuperar la imagen (3):

$$\mathbf{I}_T = \mathbf{L}^T (\mathbf{I} - \mathbf{M}) \mathbf{R} \quad (2)$$

$$\mathbf{I}_R = \mathbf{L} \mathbf{I}_T \mathbf{R}^T + \mathbf{M} \quad (3)$$

El error de recuperación, que se define como la diferencia entre la imagen original y la recuperada, proporciona una medida de la diferencia entre el modelo de fondo y la imagen de entrada en cada píxel de la imagen. Consideramos píxeles candidatos a pertenecer a objetos en movimiento aquellos en los que el valor del error de recuperación obtenido supere un umbral.



Fig. 2. Curvas iniciales (a) circunferencia (b) curva obtenida usando GPCA

En la Fig. 2 se muestran las curvas iniciales obtenidas al inicializar con circunferencias (Sekkati and Mitiche 2006a) y la curva obtenida usando el algoritmo basado en GPCA. Se puede observar cómo en la figura de la derecha, la curva se

aproxima al contorno del robot permitiendo obtener la segmentación en un número de iteraciones menor. Sin embargo, la curva obtenida con GPCA incluye la sombra proyectada por el robot sobre el suelo. Esta sombra se desplaza a la misma velocidad que el robot, por lo que no es sencillo eliminarla en la etapa de segmentación. Para minimizar el efecto de las sombras se ha incluido en la inicialización un proceso de eliminación de sombras (Finlayson et al. 2002). Mediante (4) es posible proyectar una imagen de entrada χ a una imagen invariante en escala de grises. El ángulo θ de proyección se obtiene como aquel que minimiza la entropía del histograma de la imagen proyectada (Finlayson et al. 2004).

$$I_{inv} = \chi_1 \cos \theta + \chi_2 \sin \theta \quad (4)$$

Se consideran bordes de sombras los píxeles en los que la norma del gradiente de la imagen original supera un umbral dado, mientras que la norma del gradiente de la imagen invariante es inferior a un segundo umbral. De esta forma se define la función de umbral (5) que nos permite eliminar la zona de sombra de la curva inicial.

$$S = \begin{cases} 1 & \|\nabla I_g(x, y)\| > U_g, \|\nabla I_{inv}(x, y)\| < U_{inv} \\ 0 & \text{resto} \end{cases} \quad (5)$$

En la Fig. 3 se muestra la curva resultante tras aplicar el proceso de eliminación de sombras a la curva inicial obtenida con GPCA (Fig. 2(b))

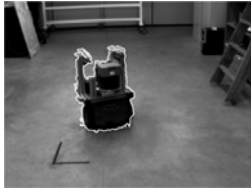


Fig. 3. Curva inicial tras la eliminación de la sombra

4. Minimización de la función de energía

Tras la inicialización de las curvas y la profundidad, la minimización de la función de energía (1) se realiza en tres pasos que se repiten hasta alcanzar la convergencia. Estas etapas se describen brevemente a continuación:

En la primera de las tres etapas, se suponen conocidas las curvas y la profundidad, de forma que la función a minimizar se reduce a (6):

$$E\left(\{\mathbf{T}_k\}_{k=1}^N, \{\boldsymbol{\omega}_k\}_{k=1}^N\right) = \sum_{k=1}^N \int_{\Omega_k} \psi_k^2(\mathbf{x}) d\mathbf{x} \quad (6)$$

donde $\boldsymbol{\Psi}_k = \mathbf{I}_t + \mathbf{s} \cdot \frac{\mathbf{T}}{Z} + \mathbf{q} \cdot \boldsymbol{\omega}$ y \mathbf{s} y \mathbf{q} son los siguientes vectores que dependen de la distancia focal de la cámara f y las derivadas parciales ($\mathbf{I}_x, \mathbf{I}_y$)

$$\mathbf{s} = \begin{pmatrix} f \mathbf{I}_x \\ f \mathbf{I}_y \\ -x \mathbf{I}_x - y \mathbf{I}_y \end{pmatrix} \quad \mathbf{q} = \begin{pmatrix} -f \mathbf{I}_y - \frac{y}{f} (x \mathbf{I}_x + y \mathbf{I}_y) \\ -f \mathbf{I}_x - \frac{x}{f} (x \mathbf{I}_x + y \mathbf{I}_y) \\ -y \mathbf{I}_x + x \mathbf{I}_y \end{pmatrix}$$

Dado que $\boldsymbol{\Psi}_k$ depende linealmente de \mathbf{T}_k y $\boldsymbol{\omega}_k$, la minimización se realiza resolviendo un sistema de ecuaciones lineales mediante mínimos cuadrados. En la segunda etapa la función a minimizar para recuperar la profundidad es la mostrada en (7) donde χ_k es la función característica de la región Ω_k . Esta ecuación nos permite actualizar el valor de la profundidad.

$$E(Z) = \int_{\Omega} \sum_{k=1}^N [\chi_k(\mathbf{x}) (\psi_k^2(\mathbf{x}) + \mu g(\|\nabla Z\|))] d\mathbf{x} \quad (7)$$

En la tercera y última etapa, dada la profundidad y las componentes de velocidad, la función de energía a minimizar con respecto a las curvas es (8). A partir de esta ecuación se actualizan las curvas mediante descenso por curvas de nivel.

$$E\left[\{\gamma_k\}_{k=1}^{N-1}\right] = \sum_{k=1}^N \int_{\Omega_k} \xi_k(\mathbf{x}) d\mathbf{x} + \lambda \sum_{k=1}^{N-1} \oint_{\gamma_k} ds \quad (8)$$

El proceso de minimización finaliza cuando se alcanza la convergencia, obteniéndose tanto los parámetros de movimiento 3D, como las curvas que definen la segmentación, y la distancia de cada punto a la cámara.

5. Resultados

Para comprobar las prestaciones del algoritmo propuesto se han realizado diversos experimentos. Todas las pruebas han sido realizadas utilizando un equipo Intel Core 2 CPU 6600 a 2.40GHz con 3.50 Gb de memoria RAM.

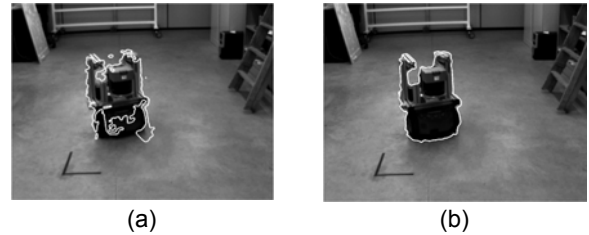


Fig. 4. Resultado de la segmentación (a) 100 iteraciones, inicialización con circunferencias (b) 3 iteraciones, inicialización usando GPCA

En las Fig. 2(a) y Fig. 3 se han presentado dos curvas para inicializar el algoritmo. A continuación se muestra el resultado de la segmentación en cada uno de los casos. En la Fig. 4(a) se puede observar cómo, en el caso de usar circunferencias en la inicialización, tras 100 iteraciones no se ha segmentado completamente el robot mientras que, en la Fig. 4 (b) se muestra cómo tres iteraciones son suficientes para obtener la segmentación al inicializar usando la técnica propuesta

En la Fig. 5 se muestra el tiempo de ejecución para cada una de las dos alternativas de inicialización al segmentar 10 pares de imágenes de entrada. En esta figura, en la que el eje X corresponde con la posición de la primera de las dos imágenes utilizadas para la segmentación, mientras que en el eje Y se muestra el valor del tiempo de ejecución en segundos, se puede observar cómo el tiempo de cómputo necesario para la segmentación se reduce de forma significativa al realizar la inicialización utilizando la técnica basada en GPCA que se ha propuesto en este trabajo, frente a las circunferencias propuestas en (Sekkati and Mitiche 2006a).

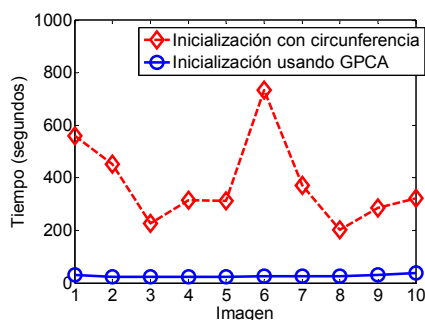


Fig. 5. Tiempo de procesamiento para la segmentación de 10 pares de imágenes.

El algoritmo propuesto también permite recuperar la posición 3D de los objetos en movimiento. En la Fig. 6 se presenta la trayectoria real del robot (en línea continua negra) y la obtenida usando el algoritmo propuesto (línea discontinua roja) en una secuencia formada por 50 imágenes. Tanto la coordenada X como la Y se muestran en milímetros.

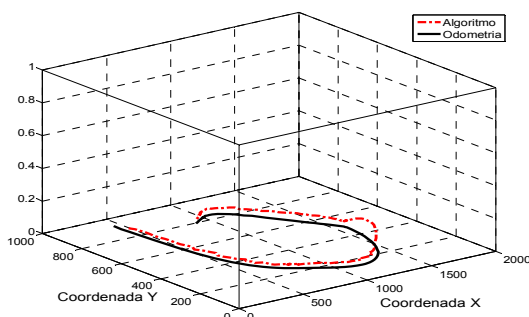


Fig. 6. Trayectoria seguida por el robot

6. Conclusiones

En este trabajo se ha presentado un sistema que permite la segmentación y posicionamiento de sillas de ruedas, así como otros objetos móviles, presentes en una secuencia de imágenes captada por una o varias cámaras fijas en el entorno. Tanto la segmentación, como la posición de las sillas de ruedas se obtienen tras la minimización de una función de energía. Dicha minimización se realiza

empleando un algoritmo *greedy*, iterativo, de tres etapas.

En las pruebas experimentales se ha observado que tanto el tiempo de cómputo como los resultados finales dependen fuertemente de los parámetros con que se inicializa la función a minimizar. En este trabajo se ha propuesto una técnica que permite inicializar las curvas usando GPCA. Por otro lado, también se ha propuesto incluir un filtro de *Kalman* por cada uno de los objetos móviles presentes en la imagen de forma que sea posible obtener una estimación de la posición que ocuparán dichos objetos en la siguiente imagen. Además se ha comprobado que la correcta inicialización de las curvas mejora notablemente el tiempo de cómputo, así como los resultados finales de la segmentación.

7. Agradecimientos

Este trabajo ha sido posible gracias a la financiación del Ministerio de Educación y Ciencia (MEC) a través del proyecto RESELA (REF-TIN2006-14896-C02-01).

8. Referencias

- Black, M., Anandan, P. (1991). *Robust dynamic motion estimation over time*. Proc. of the IEEE International Conference on Computer Vision and Pattern Recognition. 296-302.
- Costeira, J., Kanade, T. (1998) *A multibody factorization method for independently moving objects*. International Journal of Computer Vision. Vol.29, N°3. 159-179.
- Finlayson, G.D., Hordley, S.D., Drew, M.S. (2002) *Removing shadows from images*. Proc. of the 7th European Conf. on Computer Vision-Part IV, 823-836.
- Finlayson, G.D., Drew M.S., Lu, C. (2004) *Intrinsic Images by Entropy Minimization*. Proc. 8th European Conf. on Computer Vision. 582-595.
- Jieping, Y; Janardan, R; Li Q. (2004) *GPCA: an efficient dimension reduction scheme for image compression and retrieval*. Proc. 10th ACM SIGKDD international conference on Knowledge discovery and data mining. 354-363
- Kanatani, K. (2001) *Motion Segmentation by Subspace Separation and Model Selection*. Proc. of the 8th IEEE International Conf. on Computer Vision. Vol.2. 586-591.
- Sekkati, H, Mitiche, A (2006a) *Concurrent 3D Motion Segmentation and 3D Interpretation of Temporal Sequences of Monocular Images*. IEEE Trans. on Image Processing, Vol.15, N°3, 641-653.
- Sekkati, H, Mitiche, A (2006b) *Joint Optical Flow Estimation, Segmentation, and Interpretation with Level Sets*. Computer Vision and Image Understanding, Vol.103, No.2, 89-100.
- Weiss, Y. (1996) *A unified mixture framework for motion segmentation: incorporating spatial coherence and estimating the number of models*. Proc. of the IEEE International Conf. on Computer Vision and Pattern Recognition. 321-326