

Estabilización Robusta de Vídeo basada en Diferencia de Nivel de Gris

Wilbert G. Aguilar

Departamento de Ingeniería de Sistemas, Automática e
Informática Industrial, ESAII
Universitat Politècnica de Catalunya, UPC-BarcelonaTech
Barcelona, España
wilbert.aguilar@upc.edu

Cecilio Angulo

Departamento de Ingeniería de Sistemas, Automática e
Informática Industrial, ESAII
Universitat Politècnica de Catalunya, UPC-BarcelonaTech
Barcelona, España
cecilio.angulo@upc.edu

Resumen— La estabilización de vídeo se está convirtiendo en una importante técnica de post-procesado para secuencias de fotogramas (frames) adquiridas con cámaras digitales, especialmente debido al uso generalizado de cámaras de mano (hand-held) así como la utilización de estos dispositivos como elementos de entrada en sistemas robotizados complejos, robots humanoides o vehículos aéreos no tripulados.

El presente artículo propone una combinación del método iterativo RANSAC (RANDOM SAMPLE CONSENSUS), para otorgar robustez a la estimación del movimiento como parte del proceso de estabilización de vídeo, en conjunto con una función coste basada en la diferencia del nivel de gris entre imágenes.

Palabras Clave: Estabilización de Vídeo, RANSAC, Transformación Afín, Transformación Proyectiva, SIFT, SURF.

I. INTRODUCCIÓN

Tanto en dispositivos de mano de adquisición de secuencias de imágenes, cámaras digitales o smartphones, como en sistemas de dinámica compleja, robots cuadrúpedos, humanoides o aéreos que utilizan cámaras embarcadas, se registran movimientos indeseados de rotación y traslación de las imágenes, generados como producto de las vibraciones de la mano o por la naturaleza de locomoción de los sistemas que transportan los dispositivos de captura de vídeo.

En la literatura [1] [2] [3] se pueden encontrar múltiples técnicas para la compensación de estos efectos indeseados de movimiento, en cuyo proceso, por lo general, se pueden diferenciar 3 fases:

- Estimación del movimiento.
- Estimación robusta del movimiento acumulado.
- Compensación del movimiento.

Estimación del movimiento: Esta fase del proceso consiste en la estimación de los parámetros de movimiento que relacionan el frame a ser compensado con el frame definido como referencia o consigna. Trabajos desarrollados sobre este tema proponen principalmente 2 enfoques: uno basado en *flujo óptico* (optical flow) [4] y otro basado en el modelo de

transformación geométrica [5] [6] [7]. En el presente artículo se utiliza la segunda propuesta.

Independientemente del enfoque utilizado, es necesaria la detección y descripción previa de puntos de interés sobre el frame analizado. En la literatura se puede encontrar una variedad de algoritmos para llevar a cabo estas tareas previas [8] [9]. De entre ellas, SIFT [10] (Scale Invariant Feature Transform) y SURF (Speed Up Robust Feature) [11] son los algoritmos más utilizados tanto en la problemática que nos ocupa, de estimación de movimiento, como en la mayor parte de problemas relativos a visión por computador [12].

Estimación robusta del movimiento acumulado: Como parte del proceso de estimación del movimiento se realiza una búsqueda de correspondencias entre puntos de interés, tarea que se vuelve crucial si se toma en cuenta que la estimación de los parámetros de movimiento depende directamente de la fiabilidad de las correspondencias computadas. Por tanto, otorgar robustez a la estabilización del vídeo significa obtener parejas correctas de puntos de interés en correspondencia.

RANSAC es la técnica mayoritariamente empleada en la literatura para la estimación de parámetros de un modelo matemático a partir de un conjunto de puntos que contengan falsas correspondencias [13] [14] [15].

Dado que la secuencia completa de movimiento debe tener concordancia, es importante que los parámetros estimados sean válidos para el movimiento global de la imagen y no sólo para aquel relativo entre frames consecutivos.

Compensación de movimiento: Finalmente, en el proceso de compensación del movimiento se deforma el frame actual a partir de los parámetros obtenidos mediante estimación robusta, lo que debe producir una secuencia estable de movimiento.

El presente artículo se encuentra estructurado de la siguiente forma: En la sección 2, se describe la metodología para la estimación de los parámetros de movimiento a partir de los puntos de interés y el modelo de transformación geométrica correspondiente. En la sección 3 primeramente se desestima las falsas correspondencias de puntos de interés, luego se recalcula

los parámetros de movimiento con las nuevas correspondencias, para finalmente determinar el movimiento global con base en la acumulación de la transformación geométrica obtenida. La sección 4 está destinada al estudio del comportamiento, a lo largo de todos los fotogramas en la secuencia de video, de los parámetros de movimiento extraídos del modelo. Finalmente los resultados, conclusiones y trabajos futuros se presentan en las secciones 5 y 6, respectivamente.

II. ESTIMACIÓN DEL MOVIMIENTO BASADO EN LA TRANSFORMACIÓN AFÍN

El proceso de estimación de movimiento tiene como objetivo determinar los parámetros de movimiento que permiten relacionar el frame actual y el frame fijado como consigna.

Este proceso puede ser estructurado a su vez en 2 partes:

- Detección, descripción y comparación de puntos de interés.
- Estimación de los parámetros de movimiento a partir de las correspondencias de puntos.

A. Detección, descripción y comparación de puntos de interés

Pese a que los 2 algoritmos que con mayor frecuencia se utilizan en la literatura son SIFT y SURF, los resultados que se presentan en [12] confirman que el coste computacional de SURF es considerablemente menor al de SIFT sin que implique pérdida de robustez en el algoritmo. Esto se debe a que, utilizando la matriz hessiana y la función espacio-escala [16], SURF localiza los puntos de interés para luego describirlos mediante un vector 64-dimensional que contiene las características diferenciables de cada uno de ellos.

Una vez conseguidos los vectores descriptores, el matching o búsqueda de correspondencias entre puntos de interés consiste en el proceso simple de selección de aquellos pares de puntos de interés que contengan la menor diferencia vectorial entre sus descriptores 64-dimensionales.

B. Estimación de los parámetros de movimiento

Una vez obtenido el conjunto de puntos de interés, ya pueden estimarse los parámetros de movimiento entre el frame actual y el frame consigna.

El movimiento existente entre dos frames específicos se puede expresar matemáticamente mediante la transformación geométrica que relaciona los puntos de un frame con sus correspondencias en el segundo frame [17] [18] [19],

$$X_{sp} = H_t \cdot X_t \quad (1)$$

donde X_{sp} es el conjunto de puntos de interés que corresponden a la imagen consigna, H_t es la matriz de transformación y X_t es el conjunto de puntos de interés de la imagen a compensar.

Esta transformación geométrica posee un modelo paramétrico de movimiento, del cual dependerá H_t , que es distinto según qué tipo de transformación se utilice. Los 3 modelos comunes son:

Modelo de traslación

Es el más simple de los 3 modelos descritos. Hace referencia a los movimientos de la imagen cuando únicamente existe traslación del dispositivo de captura en un plano paralelo a plano de la imagen (modelo geométrico de cámara pinhole [18]),

$$H_t = \begin{bmatrix} 1 & 0 & t_y \\ 0 & 1 & t_x \\ 0 & 0 & 1 \end{bmatrix} \quad (2)$$

donde t_y y t_x son las traslaciones en y y x respectivamente.

Modelo afín

En el modelo o transformación afín existen 4 parámetros a estimar: dos desplazamientos en el plano paralelo al de la imagen, que se describen en el modelo de traslación, una rotación en la dirección del eje roll y la escala, que es proporcional al movimiento en la dirección del eje roll,

$$H_t = \begin{bmatrix} s \cos(\theta) & -s \sin(\theta) & t_y \\ s \sin(\theta) & s \cos(\theta) & t_x \\ 0 & 0 & 1 \end{bmatrix} \quad (3)$$

donde t_y y t_x son las traslaciones en y y x , s es el factor de escala y θ es la rotación alrededor del eje perpendicular al plano xy .

Modelo proyectivo

El modelo proyectivo es el modelo completo de movimiento, en el cual se encuentran expresadas matemáticamente las 3 rotaciones y traslaciones posibles. Su matriz de transformación H posee 8 parámetros linealmente independientes.

$$H_t = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{32} \\ h_{31} & h_{32} & 1 \end{bmatrix} \quad (4)$$

Modelo seleccionado

Tanto en el caso de cámaras digitales hand-held como dispositivos monoculares de visión a bordo de robots de dinámica compleja, el movimiento indeseado y vibraciones parásitas en la imagen únicamente se consideran significativas alrededor del eje roll. Por lo tanto, el modelo seleccionado para la estimación de los parámetros de movimiento es el modelo Afín.

Cabe señalar dos ventajas adicionales que se obtienen utilizando este modelo:

- La primera es referente al tiempo de cómputo, significativamente menor debido a que, a diferencia del modelo proyectivo que contiene 8 parámetros independientes, este únicamente depende de 4 parámetros.
- Una segunda ventaja es respecto a la facilidad en la extracción directa de los correspondientes parámetros

de movimiento: escala, rotación roll, y traslaciones a través del plano xy .

III. ESTIMACIÓN ROBUSTA DEL MOVIMIENTO ACUMULADO

Con el objetivo de plantear una visión clara del enfoque usado para dotar de robustez y estabilidad al vídeo, esta sección se dividirá en dos partes:

- Estimación robusta del movimiento en la secuencia de fotogramas completa basada en RANSAC.
- Estimación del movimiento global basado en la Transformación afín acumulada.

A. Estimación robusta del movimiento

Partiendo de la perspectiva que establece como robustez la asignación correcta de pares de puntos de interés en correspondencia, y luego de realizar la revisión de la literatura relativa al tema, se deduce que RANSAC (Algoritmo 1) es una técnica iterativa fiable para la desestimación de pares de puntos de interés no ajustados al modelo matemático que los relaciona, en este caso, al modelo afín.

Sin embargo, es importante definir la función coste en la cual estará basado RANSAC. En el presente artículo se propone, como función coste, la diferencia absoluta de intensidad, pixel a pixel, entre el fotograma compensado y el de referencia.

$$f_{cost} = |H_i I_{i-1} - I_i| \quad (5)$$

Por lo tanto el modelo afín H_{opt} que minimiza la función coste es:

$$H_{opt} = \arg \min_{(\theta,s,t)} \sum_i |H_i I_{i-1} - I_i| \quad (6)$$

Algoritmo 1 RANSAC:

1. Seleccionar aleatoriamente el mínimo número de puntos requeridos para determinar los parámetros del modelo.
 2. Calcular los parámetros del modelo.
 3. Determinar cuántos puntos del conjunto de todos los puntos se ajustan al modelo con una tolerancia predefinida.
 4. Si la fracción de números de inliers del total de números de puntos en el conjunto excede un umbral predefinido. Reestimar los parámetros del modelo usando todos los inliers identificados y terminar.
 5. **Caso contrario**, repetir desde el paso 1 al 4, un número limitado de ocasiones.
-

En el algoritmo 2, f_{cost} y H_{opt} se incluyen como parte de RANSAC.

Algoritmo 2 RANSAC con Función coste H_{opt}

for $i=1$ to N **do**

Estimación de la Transformación Afín i -ésima: H_i

Calculo de la función coste i -ésima:

$$f_{cost_i} = |H_i I_{i-1} - I_i|$$

end for

Selección del H_i que minimice la función coste: $H_{opt} = \arg \min_{(\theta,s,t)} \sum_i |H_i I_{i-1} - I_i|$

B. Estimación del movimiento global basado en la transformación afín acumulada

Se ha establecido que el modelo Afín contiene los parámetros de movimiento entre 2 fotogramas, uno a ser compensado y otro de referencia. Sin embargo, no se ha planteado qué fotogramas corresponden a cada respectivo caso.

Recordando que el objetivo final que se persigue es la estabilización del video, se establece que el fotograma a ser compensado es el actual, mientras que en el caso de la consigna se tiene 3 opciones:

- El fotograma inicial (I_0).
- El fotograma anterior (I_{i-1}).
- El fotograma anterior compensado ($H_i I_{i-1}$).

Opción 1: El fotograma inicial

Una primer candidato a imagen consigna es el fotograma inicial (Figura 1). En este punto cabe señalar que el inicio de la secuencia de vídeo puede ser fijada en cualquier instante de tiempo, por tal motivo, es conveniente que el fotograma seleccionado corresponda al instante de tiempo en que el dispositivo se encuentra paralelo a la superficie. Si el fotograma se escoge de forma incorrecta, toda la secuencia de vídeo se ve comprometida.

La ventaja de utilizar al fotograma inicial continuamente como consigna es que, al realizar el recálculo constante de los parámetros de movimiento de H_i , no existe error acumulado. Sin embargo, en $H_{(\theta,s,t)}$ existe una considerable desventaja respecto a la estabilidad del conjunto de frames compensados que se obtiene. Dado que cada frame es compensado de forma independiente respecto al fotograma inicial, se genera un efecto vibratorio consecuencia de la aleatoriedad del error.

Otra desventaja que es importante señalar es que, a medida que el frame actual se aleja del frame inicial, la deformación que se genera producto del movimiento se vuelve considerable, gran parte de la información se pierde y el error en la estimación de los parámetros de movimiento se puede volver crítico.

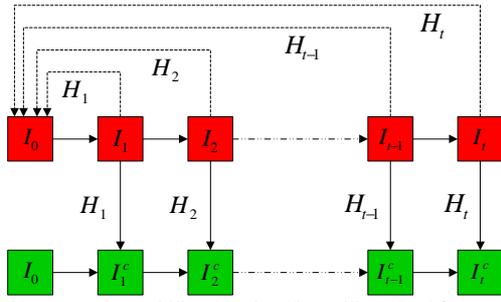


Figura 1. Esquema de estabilización de vídeo utilizando al fotograma inicial como consigna

Opción 2: El fotograma anterior

Un segundo candidato a consigna es el fotograma inmediato anterior al fotograma actual (Figura 2). La idea de utilizar esta opción se basa en que la diferencia entre los instantes de tiempo, en que frames adyacente son capturados, es la mínima detectable por la cámara de vídeo. La transformación afín que se aplique al fotograma actual no es la que se compone únicamente de los parámetros de movimiento estimados entre dos frames consecutivos, sino de la transformación afín acumulada de frame a frame.

Si bien en esta segunda opción, los parámetros de movimiento contenidos en H_i son estimados a partir del frame en el instante i e $i - 1$, cada H_i es acumulado como una transformación del fotograma inicial. Por lo tanto, al igual que en el primer caso, la selección de este fotograma inicial es fundamental. La única desventaja que presenta es el error acumulado de frame a frame, problema que puede solventarse mediante técnicas de suavizado [22] o segmentación [23].

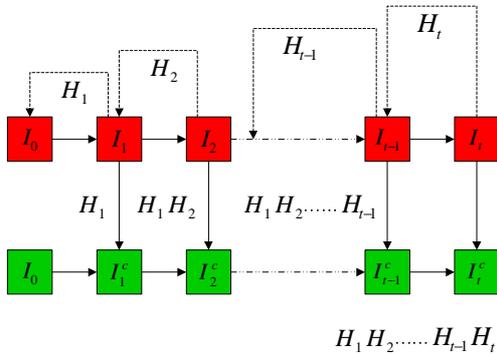


Figura 2. Esquema de estabilización de vídeo utilizando al fotograma anterior como consigna

Opción 3: El fotograma anterior compensado

Una última alternativa, que es una variante del segundo caso, es la utilización del fotograma anterior compensado como consigna (Figura 3). Sin embargo, esta alternativa no solo es la más costosa a nivel computacional, sino que además, utilizando como función coste base la diferencia matricial promedio por unidad de pixel entre la imagen compensada y la imagen consigna, experimentalmente es menos exacta que el enfoque basado en el fotograma anterior.

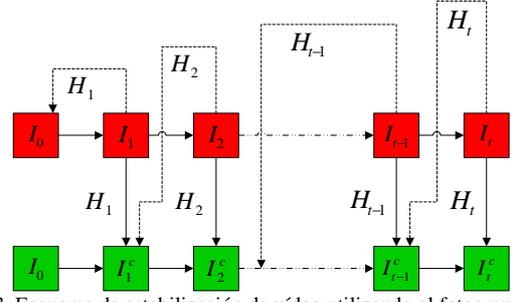


Figura 3. Esquema de estabilización de vídeo utilizando al fotograma anterior compensado como consigna

Opción seleccionada

Luego de llevar a cabo experimentalmente un análisis comparativo de los 3 enfoques propuestos basados en distinto candidatos a fotograma de referencia, se ha seleccionado el fotograma anterior como consigna. Los resultados experimentales obtenidos, respecto a la función coste y al tiempo de computo, se presentan en la Tabla I y II.

Cabe señalar que una razón adicional por la cual se ha optado por el enfoque del frame anterior es que, a lo largo de la secuencia de vídeo, cada nuevo fotograma presenta una mayor deformación respecto a la consigna original, alcanzando un frame límite en el que la deformación sea tan alta que se imposibilite la estimación del movimiento. Incluso en el caso en que se utiliza el frame anterior o el frame anterior compensado como referencia, es importante que la consigna se refresque cada cierto número de frames.

Como función coste se utilizó la diferencia promedio de nivel de gris normalizado de todos los fotogramas por unidad de pixel. A continuación se presentan las 3 funciones coste para cada una de las opciones:

Opción 1: El fotograma inicial

$$cost_0 = \frac{(\sum_i \sum_m \sum_n |H_i I_i - I_0|)}{i * m * n} \quad (7)$$

Opción 2: El fotograma anterior

$$cost_{i-1} = \frac{(\sum_i \sum_m \sum_n |H_i I_i - I_{i-1}|)}{i * m * n} \quad (8)$$

Opción 3: El fotograma anterior compensado

$$cost_{(i-1)^c} = \frac{(\sum_i \sum_m \sum_n |H_i I_i - H_{i-1} I_{i-1}|)}{i * m * n} \quad (9)$$

TABLE I. FUNCIÓN COSTE

Nombre del Vídeo	Fotograma		
	Inicial	Anterior	Anterior Compensado
Vídeo 1	0.0528	0.0119	0.0125
Vídeo 2	0.0867	0.0178	0.0188
Vídeo 3	0.1392	0.0210	0.0219
Vídeo 4	0.1483	0.0156	0.0169
Vídeo 5	0.1050	0.0158	0.0160
Vídeo 6	0.0452	0.0093	0.0094
Vídeo 7	0.1071	0.0166	0.0169

TABLE II. TIEMPO DE COMPUTO POR FRAME (SEGUNDOS)

Nombre del Vídeo	Fotograma		
	Inicial	Anterior	Anterior Compensado
Vídeo 1	0.796	0.728	0.874
Vídeo 2	0.786	0.729	0.857
Vídeo 3	0.796	0.718	0.874
Vídeo 4	0.796	0.717	0.841
Vídeo 5	0.796	0.718	0.863
Vídeo 6	0.786	0.718	0.874
Vídeo 7	0.786	0.718	0.896

Las características del computador y software, utilizados para el procesamiento de datos, se detallan a continuación como información complementaria:

- Fabricante: Acer
- Modelo: Aspire 5951G
- Procesador: Intel® Core™ i7-2670QM 2.20GHz with Turbo Boost up to 3.1GHz
- Memoria Instalada (RAM): 16,0 GB (15,9 GB Utilizable)
- Tipo de Sistema: Sistema Operativo de 64 bits
- Sistema Operativo Windows 7 Home Premium
- MATLAB 7.12.0.

IV. RESULTADOS

Finalmente, se puede obtener una perspectiva visual de los resultados obtenidos para cada uno de los escenarios sobre los cuales se ha llevado a cabo la experimentación.

En los vídeos que se muestran en las figuras 4-9 se puede constatar la compensación del movimiento, basado en un modelo de transformación afín, en el cual, existen parámetros de: traslación a lo largo de los 2 ejes paralelos al plano de la imagen en el modelo de cámara pin-hole, rotación única alrededor del eje perpendicular a los 2 ejes anteriores, y escalado.

La línea roja que se ha graficado permite evidenciar visualmente la estabilidad de la escena para cada fotograma compensado (izquierda) respecto a los fotogramas originales (derecha). Esta línea ha sido ubicada en la región sobre la cual existe un mayor número de puntos de interés, es decir que contiene información significativa sobre la escena.

Dado que la estimación del movimiento se basa en los puntos de interés, la estabilización de vídeo se realiza respecto a las regiones sobre las cuales existe una mayor aglomeración de estos puntos.

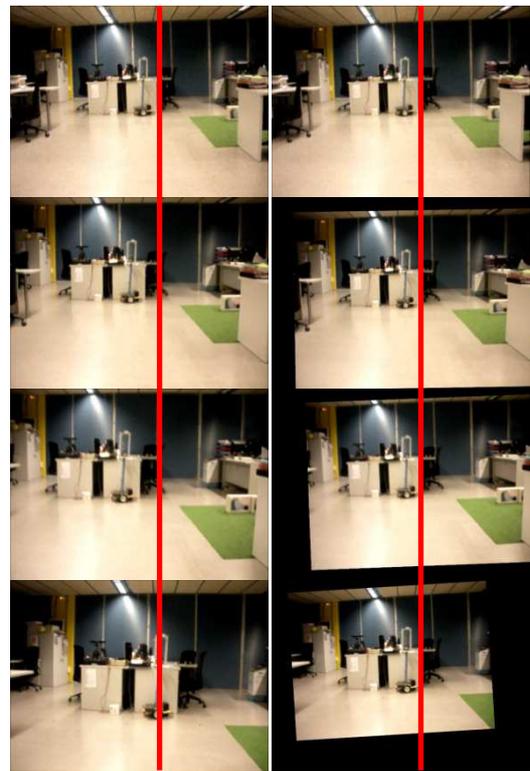


Figura 4. Vídeo 1 Estabilizado: Fotograma 1, 30, 60 y 90

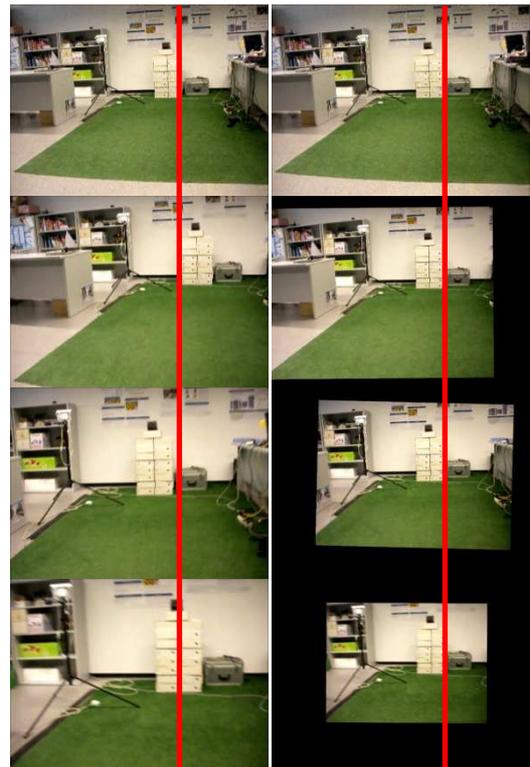


Figura 5. Vídeo 2 Estabilizado: Fotograma 1, 30, 60 y 90

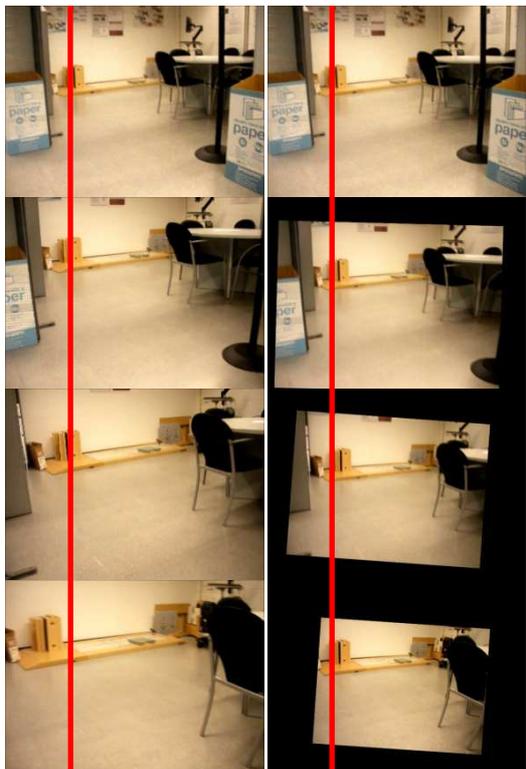


Figura 6. Vídeo 3 Estabilizado: Fotograma 1, 30, 60 y 90

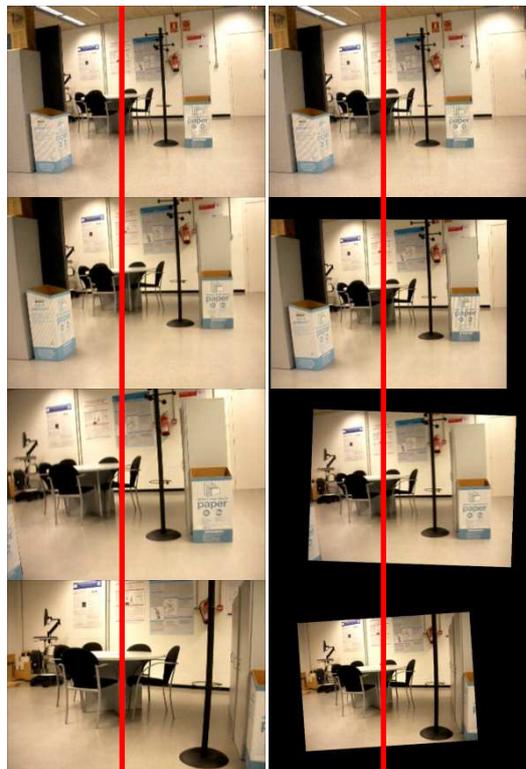


Figura 8. Vídeo 5 Estabilizado: Fotograma 1, 30, 60 y 90

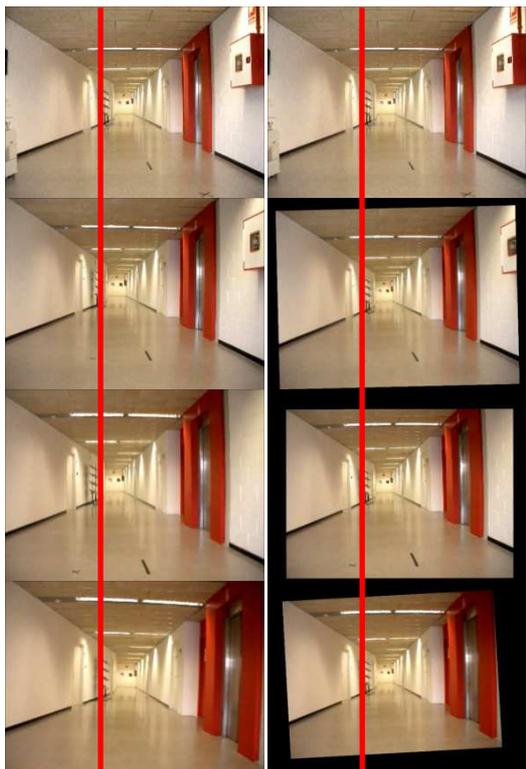


Figura 7. Vídeo 4 Estabilizado: Fotograma 1, 30, 60 y 90

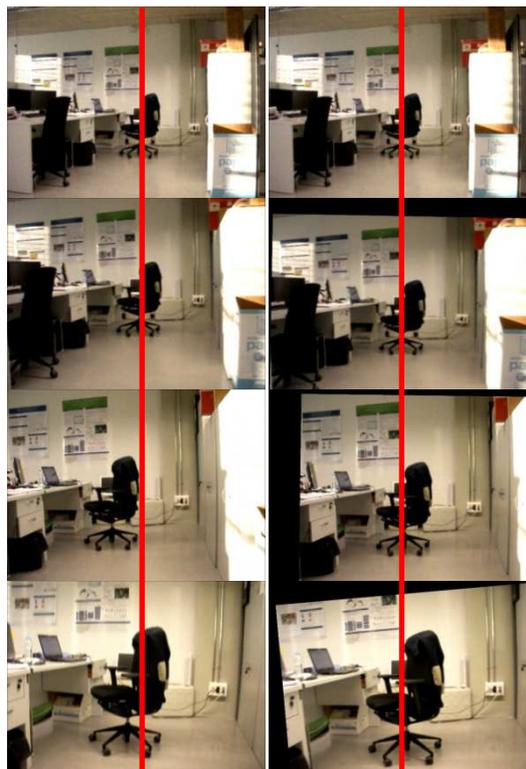


Figura 9. Vídeo 7 Estabilizado: Fotograma 1, 30, 60 y 90

Consecuentemente, las regiones de la escena, especificadas por la línea roja en el vídeo resultante, permanecen virtualmente estacionarias pese a la inestable dinámica de movimiento del dispositivo de captura. En las figuras 4-8 se ha compensado la escala con motivos ilustrativos, de tal forma que se pueda apreciar con facilidad la relación entre la imagen en un instante de tiempo específico y la consigna.

Sin embargo, con el objetivo de evitar una secuencia de vídeo que resulte estática, se puede obviar la compensación del parámetro de escala como se aprecia en la figura 9. Esto no afecta a la región señalada por la línea roja, que mantiene el comportamiento estacionario respecto a la traslación y rotación en el plano de la imagen.

Es importante mencionar algunas de las limitantes, del enfoque descrito en el presente artículo, que eventualmente pueden presentarse:

- Presencia de objetos cercanos
- Escenas con objetos en movimiento
- Escenas cuadro a cuadro
- Desplazamientos significativos
- Vídeos de baja frecuencia
- Desplazamientos de alta velocidad

Asimismo, cabe mencionar que una secuencia de vídeo estable, respecto a regiones específicas de la imagen y seleccionadas por el usuario, puede ser obtenida. Para ello, el algoritmo debe modificarse, reemplazando la detección de puntos de interés por una selección manual de dichos puntos.

V. CONCLUSIONES Y LÍNEAS FUTURAS

Luego de realizar un estudio inicial del desempeño de RANSAC, en la estabilización robusta del vídeo, se ha constatado experimentalmente su óptimo desempeño como algoritmo de desestimación de falsas correspondencias.

El modelo matemático que mejor se relaciona con movimientos indeseados típicos en dispositivos hand-held es el modelo afín; esto se debe a que las rotaciones pitch y yaw son mínimas, y su impacto sobre la deformación final de la imagen es casi imperceptible.

Como parte del algoritmo RANSAC, la función coste basada en la diferencia absoluta de intensidad o gray level, pixel a pixel, entre el fotograma compensado y el de referencia, permite la penalización de modelos de transformación cuyo efecto de deformación sobre el fotograma actual difieran, en mayor medida, respecto al fotograma consigna

El mejor candidato a consigna es el fotograma anterior, tanto desde el punto de vista de la función coste planteada, como de tiempo de cómputo. A pesar que el uso de este candidato como consigna minimiza el efecto vibratorio de frame a frame, se puede generar un error acumulado. Esta última problemática puede ser solventada utilizando técnicas de suavizado y segmentación.

Los trabajos futuros se los plantean en dos líneas:

- Estabilización de vídeo en dispositivos de mano (cámaras hand-held, smartphones, tablets).
- Estabilización de vídeo en cámaras a bordo de dispositivos de dinámica compleja (cuadrúpedos, humanoides, Rotor-UAVs).

En la primera línea, se enfrentarán problemas derivados del error acumulado, presencia de objetos cercanos, escenas con objetos en movimiento, desplazamientos significativos de frame a frame, vídeos de baja frecuencia, desplazamientos de alta velocidad, utilizando técnicas de motion smoothing [22] y motion segmentation [23].

En la segunda línea, el modelo de compensación se utilizará para determinar, mediante un sistema de entrenamiento, un modelo sensorimotor de robots de dinámica compleja. Este modelo permitirá al sistema realizar la compensación de los efectos generados en la imagen, y consecuentemente la estabilización del vídeo, en tiempo real [20] [21].

AGRADECIMIENTOS

Este trabajo ha sido financiado parcialmente por el Ministerio de Ciencia e Innovación del Gobierno de España, a través del proyecto SOFIA (TIN 2011-28854-C03-01).

El Investigador en Formación Wilbert G. Aguilar agradece la financiación a través de una beca del Programa "Convocatoria Abierta 2011" concedida por la Secretaría de Educación Superior, Ciencia, Tecnología e Innovación SENESCYT de la República del Ecuador.

REFERENCIAS

- [1] S. Battiato, G. Gallo, G. Puglisi, and S. Scellato, "SIFT features tracking for video stabilization," in *Proceeding of ICIAP*, Modena, Italy, pp. 825-830, 2007.
- [2] Y. F. Hsu, C. C. Chou, and M. Y. Shih, "Moving camera video stabilization using homography consistency". 19th IEEE International Conference on Image Processing ICIP, pp. 2761 - 2764, 2012.
- [3] C. Song, H. Zhao, W. Jing, and H. Zhu, "Robust video stabilization based on particle filtering with weighted feature points," *IEEE Transactions on Consumer Electronics*, vol. 58, pp. 2:570 - 577, 2012
- [4] H. C. Chang, S. H. Lai, and K. R. Lu, "A robust and efficient video stabilization algorithm," In *Proceedings of IEEE International Conference on Multimedia and Expo*, pp. 1:29-32, 2004.
- [5] K. Y. Lee, Y. Y. Chuang, B. Y. Chen, and M. Ouhyoung, "Video Stabilization using Robust Feature Trajectories." National Taiwan University, 2009.
- [6] J. Yang, D. Schonfeld, and M. Mohamed, "Robust video stabilization based on particle filter tracking of projected camera motion," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 19, pp. 7: 945-954, 2009.
- [7] C. Wang, J. H. Kim, K. Y. Byun, J. Ni, and S. J. Ko, "Robust digital image stabilization using the Kalman filter," *IEEE Transactions on Consumer Electronics*, vol. 55, pp. 1:6-14, 2009.
- [8] C. Harris, and M. Stephens, "A Combined Corner and Edge Detector," *Proceedings of the 4th Alvey Vision Conference*, pp. 147-151, 1988.
- [9] J. Canny, "A Computational Approach to Edge Detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 8:679-714, 1986.
- [10] D. Lowe, "Object Recognition from Local Scale-Invariant Features," *International Conference of Computer Vision*, 1999.
- [11] H. Bay, T. Tuytelaars, and L. Van Gool, "Surf: Speeded Up Robust Features," *European Conference on Computer Vision*, pp. 1:404-417, 2006.

- [12] J. Luo, and G. Oubong, "A comparison of SIFT, PCA-SIFT and SURF", *International Journal of Image Processing*, pp. 143-152, 2009.
- [13] M. Fischler, and R. Bolles, "Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography", *Communications of the ACM*, vol. 24, pp. 6:381-395, 1981.
- [14] B. Tordoff, and D. W. Murray, "Guided sampling and consensus for motion estimation," *European Conference on Computer Vision*, 2002.
- [15] K. G. Derpanis, "Overview of the ransac algorithm", Technical report, *Computer Science, York University*, 2010. http://www.cse.yorku.ca/~kosta/CompVis_Notes/ransac.pdf
- [16] K. Mikolajczyk, and C. Schmid. "Scale and Anfine Invariant Interest Point Detectors," *International Journal of Computer Vision*, vol. 60, pp. 1:63-86, 2004.
- [17] O. Faugeras, Q. Luong, and T. Papadopoulo, *The Geometry of Multiple Images*, MIT Press, 2001.
- [18] R. Hartley, and A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, 2003.
- [19] D. Forsyth, and J. Ponce, *Computer Vision, a Modern Approach*, Prentice Hall, 2003.
- [20] W. G. Aguilar, and C. Angulo, "Compensación y Aprendizaje de Efectos Generados en la Imagen durante el Desplazamiento de un Robot," *Proceedings of X Simposio CEA, Barcelona, España*, ISBN 978-84-615-7398-1, pp.165-170, 2012.
- [21] W. G. Aguilar, and C. Angulo, "Compensación de los Efectos Generados en la Imagen por el Control de Navegación del Robot Aibo ERS 7," *VII Congreso de Ciencia y Tecnología ESPE, Quito, Ecuador*, ISSN: 1390-4663, 2012.
- [22] J. Xu, H. W. Chang, S. Yang, and M. Wang, "Fast Feature-Based Video Stabilization without Accumulative Global Motion Estimation," *IEEE Transactions on Consumer Electronics*, vol. 58, pp. 3:993- 999,2012.
- [23] S. J. Kang T. S. Wang, D. H. Kim, A. Morales, and S. J. Ko, "Video stabilization based on motion segmentation," *IEEE International Conference on Consumer Electronics ICCE*, pp. 416-417, 2012.