

Estimación de la posición 3D de un quadrotor basado en visión

Gonzalo Rama-Sastre

Resumen: En este documento se explica el desarrollo de un sistema de estimación de la posición y orientación para un quadrotor utilizando técnicas de visión por computador. Los objetivos de este sistema son lograr realizar la estimación mediante la captura de imágenes que realiza el vehículo con la cámara que incorpora. Para la implementación de este sistema se ha utilizado la librería ARToolKitPlus y se ha evaluado su rendimiento tomando como referencia un dataset público de imágenes. Los resultados obtenidos por este sistema, muestran que este sistema es robusto en la detección de las referencias para realizar la estimación. Y que esta estimación depende de la distancia entre la cámara que captura las imágenes y los patrones de referencias.

Palabras claves: estimación posición, orientación, marca fiducial, quadrotor, UAV, ARToolKitPlus, realidad aumentada, visión computador.

Abstract: This paper explain the development of a system to estimate the pose and position of a quadrotor using computer vision techniques. The goals of this system are achieve estimation by mean of images captured with the camera installed on the quadrotor. For this system implementation was used the ARToolKitPlus library and was evaluated its performance compared to a public dataset. The results obtained for this system show that it is robust to detect the references to perform the estimation. And this estimation depends on the distance between the camera which capture the images and the reference pattern.

Index Terms: pose, position, estimation, fiducial marks, quadrotor, UAV, ARToolkitPlus, augmented reality, computer vision.

1 INTRODUCCIÓN

ESTE proyecto que se enmarca dentro del Trabajo de fin de grado esta relacionado con un tipo de vehículo que en la actualidad ha ganado mucha popularidad por la diversidad de aplicaciones en las que pueden ser utilizado. Estos vehículos llamados por su sigla en ingles UAV (*Unmanned Aerial Vehicle*) se presentan en diferentes modalidades, aunque las más desarrolladas son los simil avión y los que implementan hélices o rotores para su movimiento.

Los campos donde se utilizan son igual de variados que las implementaciones de estos vehículos, pero se pueden destacar sus usos en el área militar, que es donde estos vehículos empezaron a desarrollarse. Y en la actualidad tienen una gran utilidad en las áreas de la cartografía, agricultura, seguridad, hidrología y control fronterizo. [1]

Es por esta cantidad de aplicaciones y por las muchas que se podrán realizar en un futuro con este tipo de vehículos que es interesante llevar a cabo este trabajo.

El objetivo en este trabajo es desarrollar un sistema de estimación de la posición 3D para un quadrotor como el de la figura 1 basandose en las imágenes obtenidas con la cámara que incorpora. Resolver este problema es interesante ya que permite tener una referencia extra de la posición y orientación del vehículo con respecto a su entorno

en caso que los sensores de este no estuvieran disponibles.

Además la estimación de la posición de una cámara respecto a un punto de referencia es un problema de interés en otras áreas. Habitualmente suele resolverse para ser aplicado en diferentes áreas de la computación gráfica, teniendo últimamente una mayor importancia en el área de la realidad aumentada [3].

Este objetivo inicial de resolver el problema con las imágenes capturadas por el quadrotor se ha modificado



Fig.1. AR Drone 2.0. Quadrotor [2].

para que la estimación se realice a través de imágenes

obtenidas en un dataset. El motivo de este cambio se justifica por el hecho de que no era viable la obtención de datos de referencia con el quadrotor, ya que su precisión hubiese sido menor que la obtenida en el dataset que se utilizara como referencia.

Para realizar este proyecto se han revisado diferentes trabajos realizados en el campo donde se desarrolla este proyecto. La mayoría de ellos resuelven el problema de la estimación de la posición de dos formas diferentes. La primera forma de resolverlo es utilizar marcas fiduciales conocidas que son detectadas por el sistema. La segunda técnica es la asunción de planaridad de la región. [4][5][6]

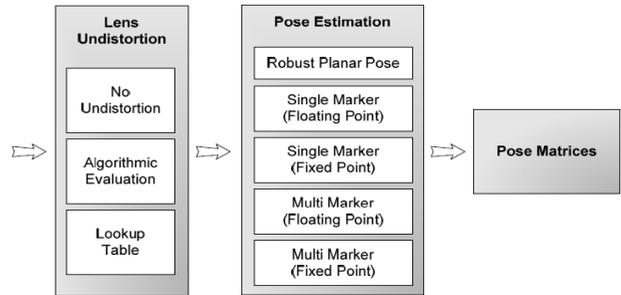
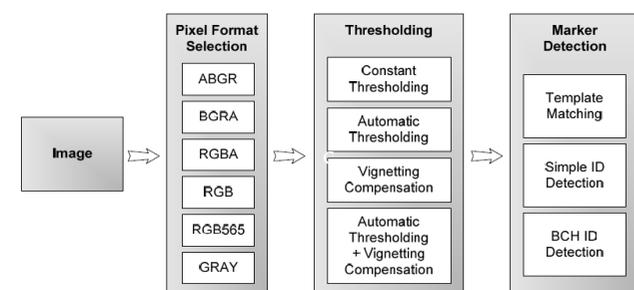
Por otra parte como se ha comentado anteriormente los datos de referencia para realizar la estimación se obtendrán de un dataset público. Por lo tanto otra tarea dentro de este proyecto ha sido la búsqueda de datos de referencia que fueran confiable y precisos.

En cuanto al software existente se ha realizado una búsqueda de cual se adaptaba mejor a la técnica de estimación elegida, en la que se ha decidido que para este proyecto ARToolKitPlus [7] era la mejor opción.

ARToolKitPlus fue presentado en 2007 como el sucesor de popular ARToolKit [8] por investigadores de Graz Technical University. Es una librería de funciones que facilitan la implementación de sistemas de estimación de la posición mediante marcas fiduciales. Se caracteriza por ser una optimización y extensión para la utilización dentro de dispositivos móviles tales como teléfonos, PDA, tablets, etc.

En el marco apartado anterior se ha explicado como se obtiene la posición de la cámara basándose en marcas fiduciales; todo este proceso es llevado a cabo por el flujo de trabajo de ARToolKitPlus.

Este flujo de trabajo se puede dividir en cinco módulos principales, en los que cada uno permite ejecutarse de una forma diferente.



Una vez la imagen ha sido capturada lo primero que se debe definir es el formato de pixel a utilizar mediante el módulo *Pixel Format Selection*.

Definido el formato de pixel para la imagen lo siguiente es ser tratada por el módulo *Thresholding* que será el encargado de obtener la imagen binaria aplicando el método que se le indique.

Como se ha visto anteriormente una vez se tiene la imagen binaria es posible realizar la detección de las marcas fiduciales. Esta tarea la lleva a cabo el módulo *Marker Detection*, el cual utiliza 3 métodos, uno mediante la comparación de imágenes y otros dos que se realizan analizando en busca de patrones que representen un identificador de la marca.

Una buena calibración de la cámara es crucial para la tarea de estimación de la posición, es por esto que el módulo *Lens Undistortion* da la posibilidad de realizar esta calibración a través de herramientas como GML MatLab Camera Calibration Toolbox.

Por último el módulo de *Pose Estimation* es el encargado de realizar la estimación de la posición y calcular matriz de transformación correspondiente.

2 MARCO TEORICO

2.1 Marcas fiduciales

Las marcas fiduciales son patrones de referencia que se ubican en la escena y que son automáticamente detectados mediante un algoritmo de reconocimiento.

Estas marcas fueron diseñadas para la rápida y fácil detección de las mismas, y el desempeño del diseño de las marcas se evaluar con métricas como las siguientes: 1) los falsos positivos en el reconocimiento. 2) la confusión con el patrón de otra marca. 3) los falsos negativos durante el reconocimiento.

- E-mail de contacte: gonzalo.rama@e-campus.uab.cat
- Menció realitzada: Enginyeria de Computació.
- Treball tutoritzat per: Daniel Ponsa (Ciències de la Computació)
- Curs 2013/14

Teniendo en cuenta las métricas y las posibilidades disponibles para utilizar en el proyecto se decidió que la mejor opción era utilizar marcas del tipo planteadas por el sistema ARTag [8]. Este sistema define marcas rectangulares con un fondo negro el cual se divide en una matriz de seis por seis en donde cada casilla de esta puede estar coloreada de color blanco o permanecer del color de fondo. Cada patrón representado en esta matriz es identificado con un id diferente.

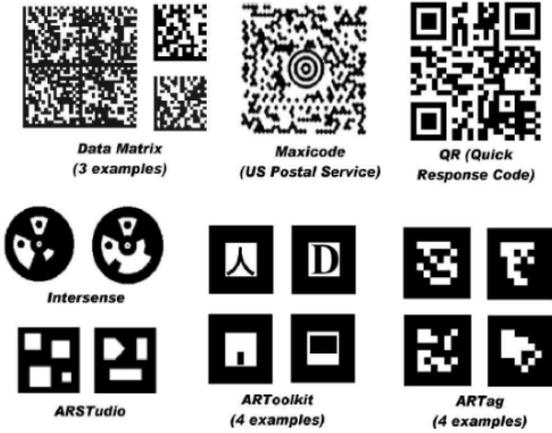


Fig. 3. Ejemplos de marcas fiduciales.

La ventaja de este sistema es la cantidad de parámetros diferentes que pueden representarse y además lo más importante es la rapidez con la que pueden ser decodificados ya que las 36 casillas de la matriz pueden representarse con un 0 ó 1 según el color de estas.

2.2 Estimación utilizando marcas fiduciales

El primer objetivo a la hora de realizar la estimación de la posición con esta técnica es realizar la detección de una marca fiducial en la imagen capturada.

En una imagen de entrada se realiza un procesamiento que consiste en transformar a escala de grises y aplicar un valor de umbral. Este proceso da como resultado una imagen binaria; la cual contiene una colección de objetos sobre un fondo negro.

Cada objeto de esta imagen es comparado con los patrones de marcas fiduciales. A las marcas detectadas se les aplica un algoritmo de detección de contornos y esquinas que serán utilizados para determinar la posición de la marca.

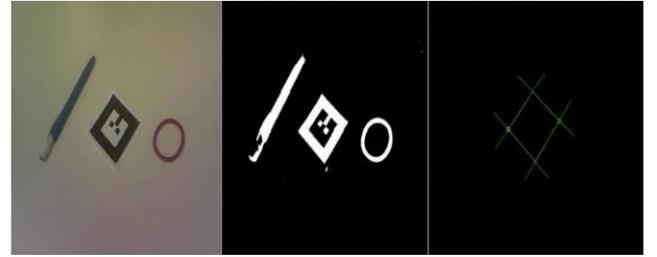


Fig. 4. A la izquierda la imagen de entrada, en el centro la imagen binaria y a la derecha la detección de contornos y esquinas.

En las marcas fiduciales se representa un sistema de coordenadas como el de la figura 4.

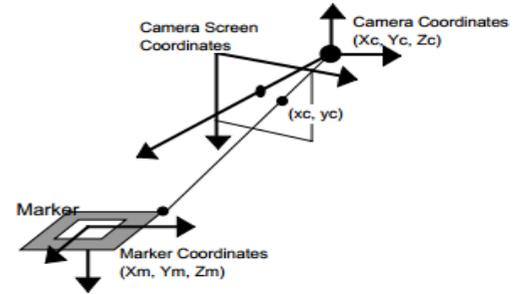


Fig. 5. Relación entre coordenadas de la marca y las de la cámara.

La matriz de transformación (T_{cm}) del sistema de coordenadas de la marca hacia las coordenadas de la cámara está representada en la ecuación 1.

$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix} = \begin{bmatrix} V_{11} & V_{12} & V_{13} & W_x \\ V_{21} & V_{22} & V_{23} & W_y \\ V_{31} & V_{32} & V_{33} & W_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X_m \\ Y_m \\ Z_m \\ 1 \end{bmatrix} \quad \text{Ec (1)}$$

$$= \begin{bmatrix} \mathbf{V}_{3 \times 3} & \mathbf{W}_{3 \times 1} \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X_m \\ Y_m \\ Z_m \\ 1 \end{bmatrix} = \mathbf{T}_{cm} \begin{bmatrix} X_m \\ Y_m \\ Z_m \\ 1 \end{bmatrix}$$

Para realizar la proyección de la marca detectada se utiliza una transformación de perspectiva que viene dada por la ecuación.

$$\mathbf{P} = \begin{bmatrix} P_{11} & P_{12} & P_{13} & 0 \\ 0 & P_{22} & P_{23} & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad \begin{bmatrix} hx_c \\ hy_c \\ h \\ 1 \end{bmatrix} = \mathbf{P} \begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix} \quad \text{Ec (2)}$$

Donde los valores de la matriz \mathbf{P} son completados por los parámetros de la cámara utilizada.

Además cuando dos lados paralelos de una marca son proyectados en la imagen, las ecuaciones de estas líneas en las coordenadas de la cámara son las siguientes:

$$a_1x + b_1y + c_1 = 0, \quad a_2x + b_2y + c_2 = 0 \quad \text{Ec (3)}$$

Cada parámetro de estas ecuaciones se calcula en el proceso de detección de contornos de las marcas. Dada la matriz P que se calcula mediante la calibración de la cámara las ecuaciones de los planos que contienen los dos lados paralelos son las siguientes:

$$\begin{aligned} a_1P_{11}X_c + (a_1P_{12} + b_1P_{22})Y_c + (a_1P_{13} + b_1P_{23} + c_1)Z_c &= 0 \\ a_2P_{11}X_c + (a_2P_{12} + b_2P_{22})Y_c + (a_2P_{13} + b_2P_{23} + c_2)Z_c &= 0 \end{aligned} \quad \text{Ec (4)}$$

Sean n_1 y n_2 los vectores normales de estos planos respectivamente, el vector director de los dos lados paralelos es dado por el producto tensorial de n_1 y n_2 . Dado esto se pueden obtener los vectores u_1 y u_2 de los dos conjuntos de lados paralelos de la marca fiducial.

Estos vectores deberían ser perpendiculares, sin embargo errores derivados del procesamiento de imágenes hacen que estos no sean exactamente perpendiculares. Para compensar este error se definen los vectores v_1 y v_2 en el plano que incluye a u_1 y u_2 como muestra la figura 5.

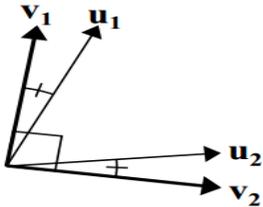


Fig. 6. Los vectores v_1 y v_2 se calculan a partir de u_1 y u_2 .

A partir de esto se puede definir un vector v_3 perpendicular a v_1 y v_2 . Con estos tres vectores se puede definir el componente de rotación $V_{3 \times 3}$ de la matriz de transformación de la ecuación 1 como $[V_1^t V_2^t V_3^t]$.

Una vez se ha obtenido $V_{3 \times 3}$, utilizando la ecuación 1, los cuatro vértices de la marca fiducial en el sistemas de coordenadas de esta y con estos mismo vértices pero en coordenadas de la cámara se obtienen ocho ecuaciones que incluyen el componente de traslación $W_x W_y W_z$.

Al resolver estas ecuaciones en la matriz de transformación se tiene como esta rotada y trasladada la cámara con respecto a la marca. [9][10]

3 METODOLOGIA APLICADA

3.1 Obtención de las imágenes

Como se ha adelantado en la introducción del documento las imágenes utilizadas para el desarrollo del sistema han sido obtenidas de un dataset. Concretamente el dataset es *City of Sights* [11] [12] y fue desarrollado para aplicaciones de realidad aumentada.

Este dataset se caracteriza por tener secuencias de imágenes de diferentes escenarios, algunos de ellos donde se encuentran edificios o monumentos y en otros como el que se ha escogido como base para el sistema a implementar solo tiene una región plana con una marca fiducial en el centro.

El principal motivo por el cual se ha decidido utilizar esta base de datos es la precisión que tienen los datos de referencia que se dan de cada secuencia de video. En particular las utilizadas por este sistema son las secuencias FARO_02 que tienen un margen de error de 0.1mm en sus datos de referencia.

Y otro motivo no menos importante a la hora de elegir utilizar este dataset y no crear uno propio con las imágenes capturadas por el quadrotor, es el costo que implicaría realizarlo a nivel de horas de trabajo para lograr secuencias de imágenes con un margen de error aceptable en los datos de referencia.

Las secuencias de imágenes son tomadas por una cámara en movimiento instalada en el brazo de un robot que graba un plano que contiene el centro una marca fiducial de unos 40mm de lado. En la figura 7 puede observarse dicho plano.



Fig. 7. Plano de las secuencias de imágenes de referencias

Los datos de referencia que esta base de datos brinda están compuestos de una posición x, y, z y una rotación dada por una quaternion A, B, C, D para cada *frame* de las secuencias de imágenes.

Dado que no en todos los frames de la secuencia elegida se percibe la marca fiducial una de la tareas realizadas ha sido seleccionar los *frames* donde la marca fiducial se

observara de forma clara. Con estos *frames* se ha creado un video de referencia y se selecciono los datos de referencia para estas imágenes.

3.2 Desarrollo de aplicación de estimación

Para evaluar el sistema de estimacion mediante marcas fiduciales se implemento una aplicación que utiliza el flujo de trabajo de ARToolKitPlus. Dado que esta biblioteca viene por defecto para ser utilizada con las imágenes capturadas a través de una cámara en tiempo de ejecución, se ha tenido que desarrollar unas pequeñas funcionalidades que envían cada frame de los videos del dataset de referencia guardados en archivos.

Por lo tanto la aplicación consiste en coger cada imagen que contiene un video y pasarlas a través del flujo de trabajo de ARToolKitPlus explicado en apartados anteriores.

Este flujo devuelve la matriz de transformación de la ecuación 1 de donde se obtienen los datos para ser comparados con los datos de referencia brindados por el dataset.

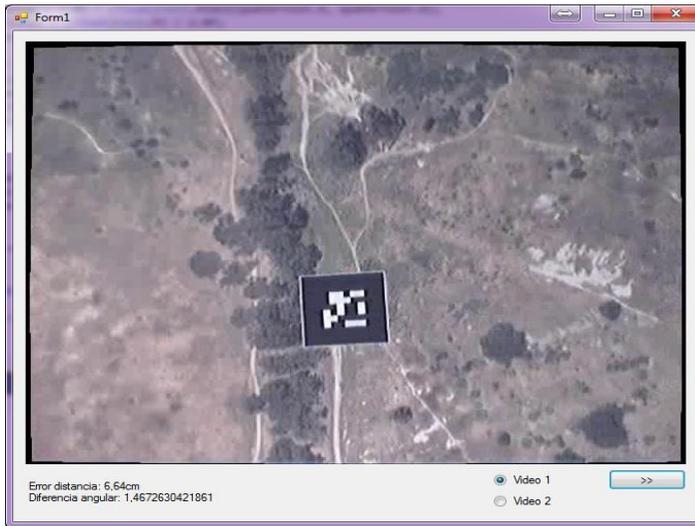


Fig. 8. Interfaz de usuario de la aplicación desarrollada

En cuanto a los aspectos técnicos de la aplicación esta fue implementada mediante el lenguaje Microsoft Visual C# y cuenta con una interfaz de usuario que se muestra en la figura 8.

Esta interfaz da la posibilidad de elegir entre dos videos de referencia a evaluar y con el botón de la derecha pasar a través de los frames de dichos videos. En la esquina inferior izquierda un texto va mostrando el error de cada imagen del video.

3.4 Métricas y formulas aplicadas

Una de las decisiones más importantes a la hora de evaluar el sistema ha sido elegir métricas que se utilizarán para llevar a cabo esta acción. Estas métricas evaluarán la precisión en la estimación de la posición y orientación.

Dado los datos que nos proporciona el dataset de referencia elegido y teniendo en cuenta los datos que brinda la estimación realizada por el sistema se ha elegido como métrica para la estimación de la posición la distancia euclídea. Esta distancia mide la diferencia entre la posición estimada y la posición de referencia. Y se ha calculado aplicando la siguiente formula:

$$d_E(P, Q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \dots + (p_n - q_n)^2} = \sqrt{\sum_{i=1}^n (p_i - q_i)^2} \quad \text{Ec (5)}$$

Donde para este caso n es igual a tres y p_n corresponde a una coordenada dada por el sistema y q_n es una coordenada de referencia.

Para la evaluación de la estimación en la orientación se ha escogido la diferencia angular que exista entre el cuaternión que representa la rotación de la cámara según el sistemas y el cuaternión que representa la rotación de la cámara según los datos de referencia.

Para calcular esta diferencia, sean q_1 y q_2 los cuaterniones dados por el sistema y el de referencia respectivamente, la formula utilizada es la siguiente:

$$\alpha = \text{acos} ([q_1, q_2^*]) \quad \text{Ec (6)}$$

Donde q_2^* es el inverso del cuaternión q_2 y $[q_1, q_2^*]$ es el producto vectorial entre los dos cuaterniones.

4 RESULTADOS OBTENIDOS

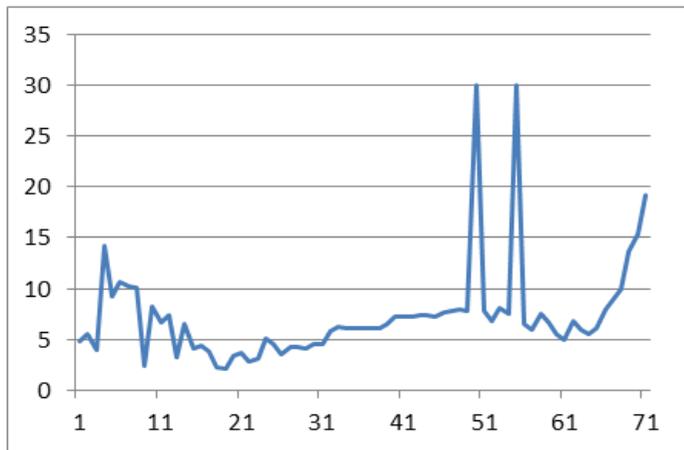
La evaluación del sistema se ha realizado sobre dos secuencias de video en las cuales los movimientos que realiza la cámara difieren mucho de un video a otro. Se puede apreciar claramente que en video 1 las imágenes se han tomado a una distancia menor que el video 2.

En las graficas que se muestran a continuación para cada video además de las correspondientes a cada métrica a evaluar se tiene una grafica que muestra la distancia de la cámara respecto a la marca fiducial.

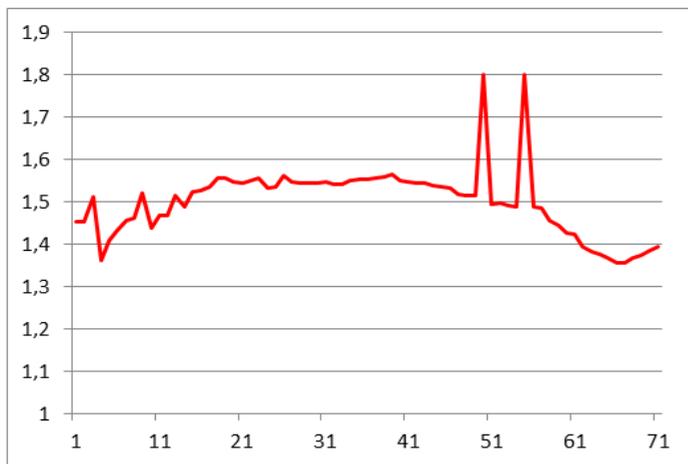
Las desviaciones abruptas de las graficas se corresponden a frames de cada video donde la marca fiducial no se ha podido detectar.

4.1 Resultados video 1

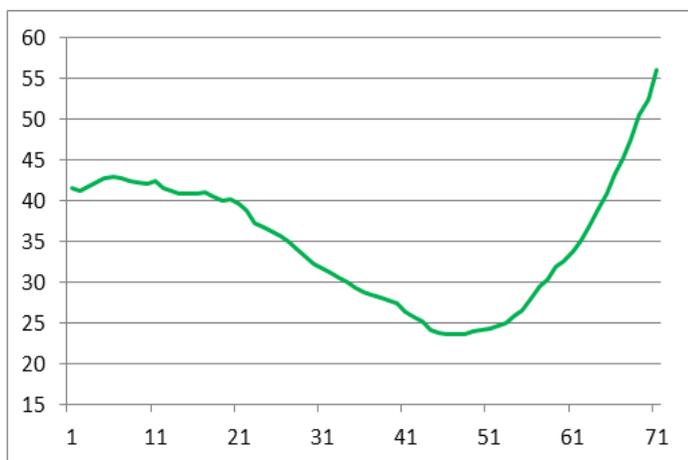
En esta secuencia de video la cámara se acerca a la marca fiducial a una distancia de 20cm para volver a alejarse al final de la secuencia.



Gráfica 1: Distancia euclídea en cm. entre posición de referencia y posición estimada por el sistema



Gráfica 2: Diferencia angular entre cuaterniones de la orientación de referencia y orientación estimada por el sistema

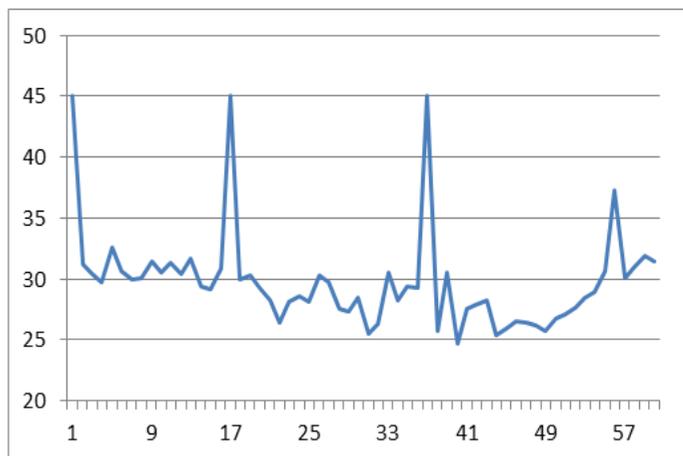


Gráfica 3: Distancia euclídea en cm entre la posición de la cámara según datos de referencia y posición de la marca.

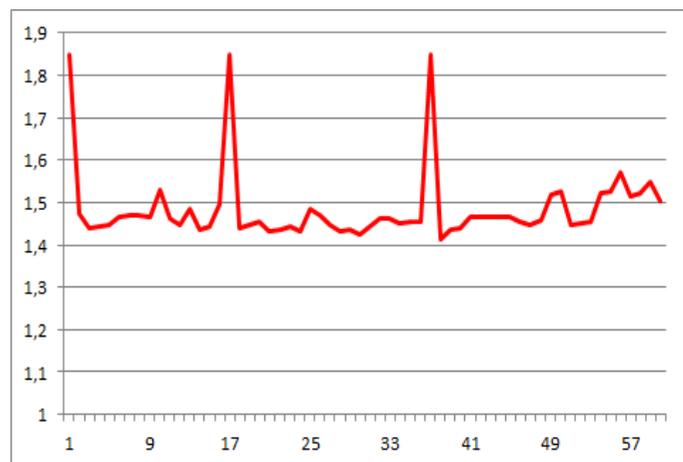
4.2 Resultados video 2

La secuencia de video 2 tiene como característica que las imágenes son tomadas con la cámara mas alejada que en la secuencia anterior.

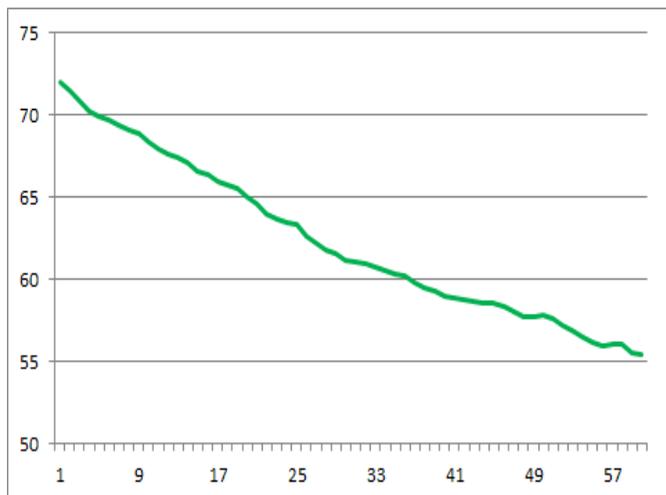
En la secuencia de imágenes del video 2 la cámara se encuentra alejada de la marca fiducial y se va acercando a esta de forma progresiva.



Gráfica 4: Distancia euclídea en cm. entre posición de referencia y posición estimada por el sistema



Gráfica 5: Diferencia angular entre cuaterniones de la orientación de referencia y orientación estimada por el sistema



Gráfica 6: Distancia euclídea en cm entre la posición de la cámara según datos de referencia y posición de la marca.

4.3 Analisis de resultados

Viendo los resultados obtenidos para las dos secuencias de video se puede afirmar que el aumento de la distancia que se aprecia entre cada video no afecta a la detección de las marcas fiduciales, las cuales son detectadas casi en la totalidad de los frames analizados.

Por otra parte si que puede verse como el aumento de la distancia de la cámara con respecto a la marca fiducial hace que el error en la estimación de la posición aumente. Un ejemplo claro de esto es el final del video 1 y la totalidad del video 2 donde la cámara se acerca y el error en la estimación de la posición hace lo propio.

En cuanto al error correspondiente a orientación de la cámara puede observarse que no se ve afectado por la distancia de la cámara y que para las dos secuencias de video se mantiene en el mismo rango. Por lo que este error puede deberse a factores que no tienen que ver con el movimiento de la cámara sino con la calibración de la cámara o la calidad de la imagen capturada.

5 CONCLUSIÓN

Una vez acabado el proyecto se puede decir que se ha logrado cumplir el objetivo de crear un sistema de estimación y orientación para un quadrotor utilizando marcas fiduciales. Quedando como trabajo futuro la implementación de realizar la estimación a través de la asunción de planaridad de la región.

En cuanto a los resultados obtenidos se puede concluir que el sistema realiza una estimación de la posición más precisa cuando la cámara está más cerca de la marca fiducial. Lo cual viene determinado porque la proyección de la marca fiducial en la imagen es más pequeña cuando la cámara se aleja. Por lo tanto el tamaño de la marca fiducial del

entorno donde se aplicará el sistema deberá aumentarse cuando se quiera una estimación precisa cuando la cámara se aleja de la marca fiducial.

En la introducción se comentaba que era de interés desarrollar este sistema para cuando los sensores del quadrotor no estuvieran disponibles. Y dado que la detección de las marcas se realiza casi en su totalidad este sistema es una buena alternativa para cuando estos sensores fallen.

También es preciso mencionar que como mejora futura se podría implementar una funcionalidad que interpolara las posiciones detectadas con anterioridad, para cuando una marca fiducial no es detectada, se pueda dar una posición estimada en función de la interpolación.

AGRADECIMIENTOS

El trabajo realizado en este proyecto ha resultado ser más complejo y extenso del que inicialmente estipulaba y he podido llevarlo a cabo con la ayuda de varias personas que me han dado su apoyo y consejos.

En particular quiero agradecer a Daniel Ponsa, tutor del proyecto, que me ha ayudado con referencias y consejos útiles a la hora de escribir este artículo.

También a Lukas Gruber de Graz University of Technology y autor del dataset de referencia por su ayuda para comprender su trabajo.

Y por último a todas las personas que me han apoyado durante este trabajo con su paciencia y consejos.

BIBLIOGRAFIA

- [1] Fernando Muñoz, "Los usos más increíbles de los 'drones' ", [Diario ABC](#), Junio, 2013
- [2] Parrot, [AR Drone 2.0](#), Accedido febrero 2014
- [3] R. Silva, J. C. Oliveira, G. A. Girdali, Introduction to Augmented Reality, National Laboratory for Scientific Computation, Brazil
- [4] N. Cirera, O. Ferhat, L. Gómez, H. Gao and X. Hu, Planar Surface 3D Tracking, Universitat Autònoma de Barcelona, 2012
- [5] Dana Cobzas, Martin Jagersand, 3D SSD Tracking from Uncalibrated Video, University of Alberta, Canada
- [6] Georg Klein, David Murray, Parallel Tracking and Mapping for Small AR Workspaces, Department of Engineering Science, University of Oxford
- [7] Daniel Wagner, Dieter Schmalstieg, ARToolKitPlus for Pose Tracking on Mobile Devices, Institute for Computer Graphics and Vision, Graz University of Technology
- [8] Mark Fiala, ARTag, a fiducial marker system using digital techniques, National Research Council of Canada
- [9] Hirokazu Kato, Mark Billinghurst, Marker Tracking and HMD Cali-

bration for a Video-based Augmented Reality Conferencing System, Hiroshima City University

- [10] Sanni Siltanen, Theory and applications of marker-based augmented reality, VTT SCIENCE 3, ISBN 978-951-38-7449-0
- [11] Lukas Gruber, Steffen Gauglitz, Jonathan Ventura, Stefanie Zollmann, Manuel Huber, Michael Schlegel, Gudrun Klinker, Dieter Schmalstieg, Tobias Hollerer, The City of Sights: Design, Construction, and Measurement of an Augmented Reality Stage Set
- [12] Sitio web [THE CITY OF SIGHTS: AN AUGMENTED REALITY STAGESET](#), Accedido febrero 2014