

Trabajo 1-2.

Reconstrucción 3D

Proyectiva y Euclídea

TÉCNICAS AVANZADAS DE VISIÓN POR COMPUTADOR

Carlos Prados Sesmero | 70830147Z | M18129

Índice

Introducción.....	2
1. Desarrollo de la práctica	3
1.1. Detección de características visuales	3
1.2. Establecimiento de correspondencia entre SIFT.....	4
1.3. Cálculo de la matriz fundamental.....	5
Filtrado de falsas correspondencias (RANSAC)	5
Filtrado de falsas correspondencias (Error residual)	13
Comparativa de los métodos de filtrado	15
1.4. Reconstrucción Proyectiva 3D	16
Referencias	20

Introducción

En la asignatura Técnicas Avanzadas de Visión por Computador se ha planteado el desarrollo de un trabajo práctico que consiste en reconstruir en 3D la geometría proyectiva y euclídea.

El objetivo del trabajo es realizar una reconstrucción tridimensional a partir de la información de una escena suministrada por la adquisición de imágenes desde distintos puntos de vista. Todo ello implica detectar características visuales, filtrar dichas características, evitando falsas detecciones, estimar la Matriz Fundamental y realizar una reconstrucción tridimensional.

Se trabajará con dos diferentes escenarios (*Figura 1 y 2*), cada uno de los cuales está formado por tres planos con impresiones de fotografías que contienen gran cantidad de información. Cada fotografía es una imagen localizada en diferente punto de vista respecto al resto, por lo que lo denominaremos Escenario1_Imagen_1, Escenario1_Imagen2, ..., Escenario2_Imagen 3.



Fig. 1. Escenario 1



Fig. 2. Escenario 2

El desarrollo de la práctica se va a realizar en Matlab, gracias a una serie de programas encargados de encontrar puntos característicos en las imágenes (shift) y encargados de casar estos puntos característicos entre varias imágenes (match).

1. Desarrollo de la práctica

A continuación se van a desarrollar una serie de etapas, de manera incremental, en cada una de ellas se procede a desarrollar una actividad de la práctica.

1.1. DETECCIÓN DE CARACTERÍSTICAS VISUALES

El algoritmo SIFT es un algoritmo de visión artificial, publicado originalmente por David Lowe en 1999, que se encarga de extraer características distintivas de las imágenes en escala de grises. Mediante estas, es posible luego reconocer dicha imagen dentro de una base de datos o incluso dentro de otra imagen mayor con otra cantidad de elementos en desorden. Estas características son invariantes a factores de escala, traslación, rotación y parcialmente invariantes a cambios de iluminación y afinidades [1].

En esta práctica, ejecutando la función “match.m” es posible detectar puntos característicos en las imágenes (mediante la función “sift.m”) y casarlos entre varias imágenes.

Así por ejemplo, para el escenario 1, se encuentran 10976 puntos característicos para la primera imagen, 8993 para la segunda y 11132 para la tercera.

1.2. ESTABLECIMIENTO DE CORRESPONDENCIA ENTRE SIFT

El establecimiento de correspondencias se realiza mediante la función “match.m”, que a su vez se basa en el cálculo de los descriptores SIFT. Un ejemplo es el mostrado en la *Figura 3*, donde se observan las correspondencias (sin filtrar) de las imágenes 1 y 2 del escenario 1. En este caso se han encontrado 1990 correspondencias entre ambas imágenes.

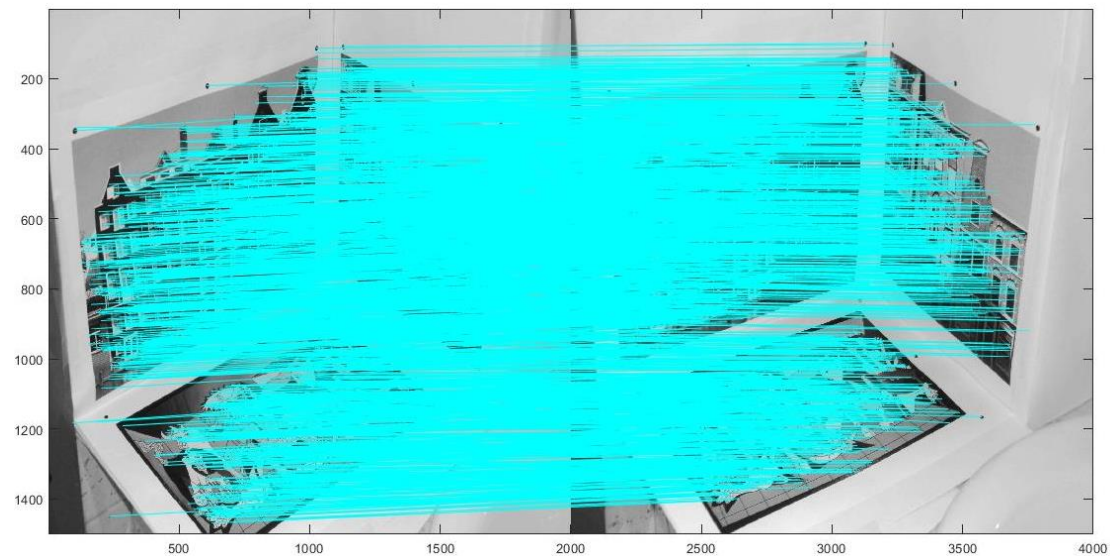


Fig. 3. Puntos casados (correspondencias) entre las imágenes 1 y 2 del escenario 1. Se muestran las correspondencias correctas y los puntos mal casados.

Si representamos la disparidad de los puntos casados entre la imagen 1 y 2 del primer escenario podemos observar el resultado mostrado en la *Figura 4*, donde se pueden observar líneas que siguen un patrón en forma de remolino. Hay ciertas líneas cuyo casamiento es erróneo, mientras que el casamiento de la mayoría de los puntos casados parece correcto a priori.

En la *Figura 5* podemos observar destacados en rojo una serie de disparidades cuya longitud es demasiado elevada comparada con el resto.

En ambas figuras se pueden localizar 9 errores visibles. Puesto que se han casado 1990 puntos, el porcentaje de puntos incorrectamente casados ha sido de 0.4523%. Este valor es muy reducido y nos servirá para posteriores cálculos de parámetros en la aplicación de RANSAC.

El umbral empleado se indicará más adelante.

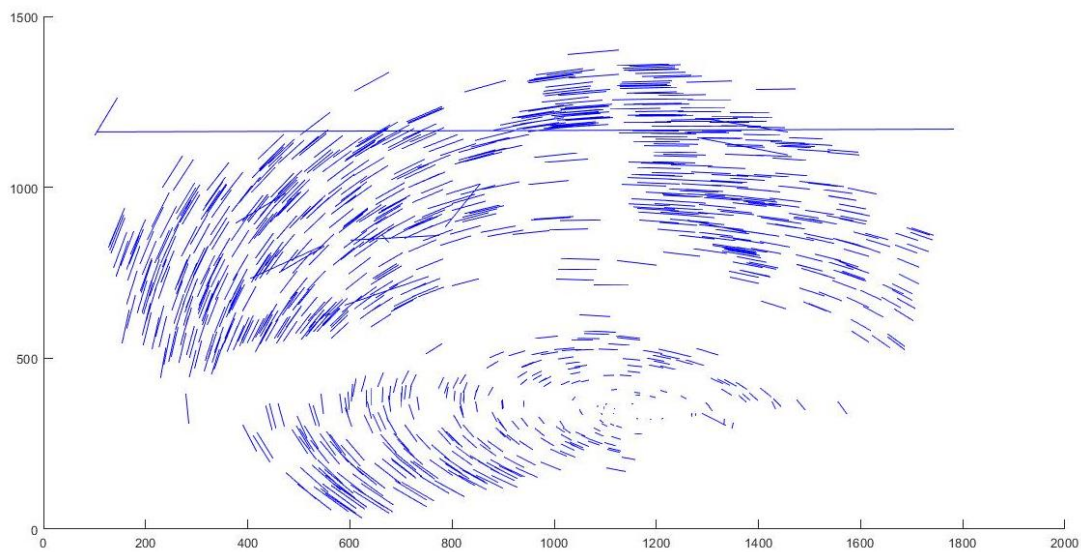


Fig. 4. Disparidad de los diferentes puntos casados. Patrón en remolino. Existen varios cuyo tamaño o dirección es incorrecta dentro del patrón común.

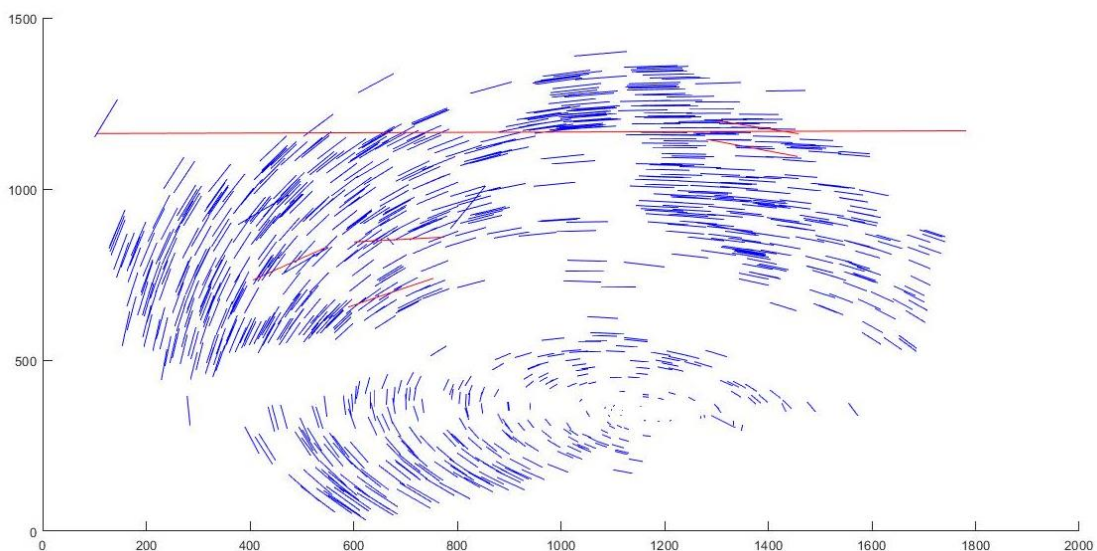


Fig. 5. Disparidad de los diferentes puntos casados. Las disparidades evidentemente incorrectas (con una longitud demasiado elevada) se muestran en rojo.

1.3. CÁLCULO DE LA MATRIZ FUNDAMENTAL

La Matriz Fundamental recoge la información geométrica de la disposición binocular, salvo un escalado en la traslación. En esta práctica, se va a emplear el método robusto RANSAC para no tener en cuenta los datos atípicos, es decir, las falsas correspondencias.

Filtrado de falsas correspondencias (RANSAC)

La existencia de falsas correspondencias puede ocasionar un elevado error en los cálculos posteriores (Matriz Fundamental, Reconstrucción Proyectiva). Para eliminarlas se han barajado diferentes opciones.

En primer lugar se va a aplicar el método RANSAC, el cual permite eliminar valores atípicos en una muestra. Para ello se ha realizado un estudio, para la determinación adecuada de los parámetros del método.

En la *Figura 6* podemos observar la repercusión del número de iteraciones del método en el porcentaje de “inliers” (estudio realizado con el resto de los parámetros constante).

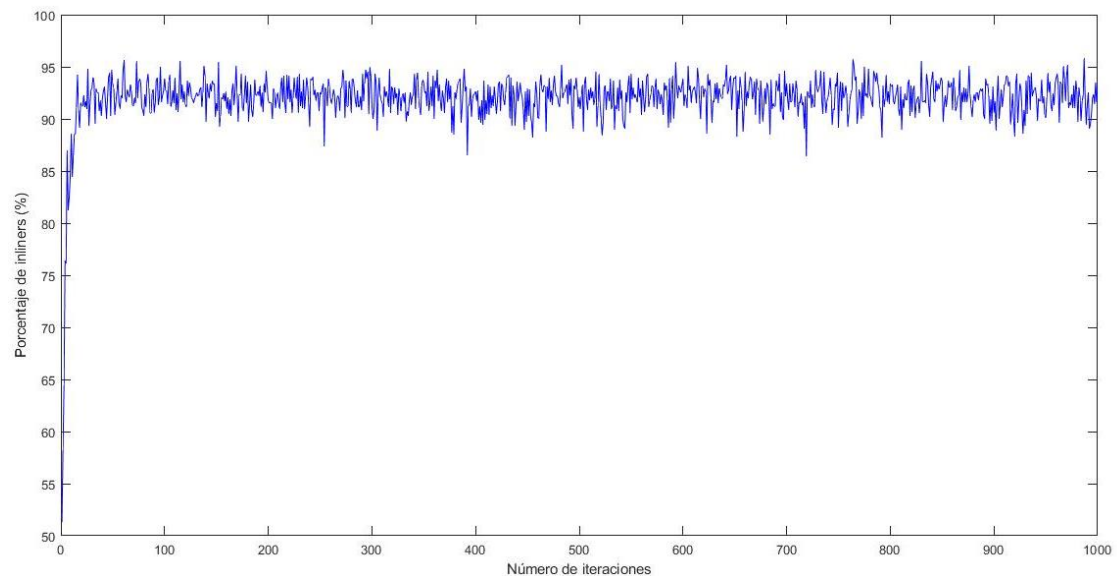


Fig. 6. Porcentaje de “inliers” respecto al número de iteraciones del método. Se observa que el porcentaje de “inliers” se estabiliza con un número reducido de iteraciones.

Si se focaliza el estudio en un rango inferior, el resultado obtenido es el mostrado en la *Figura 7*.

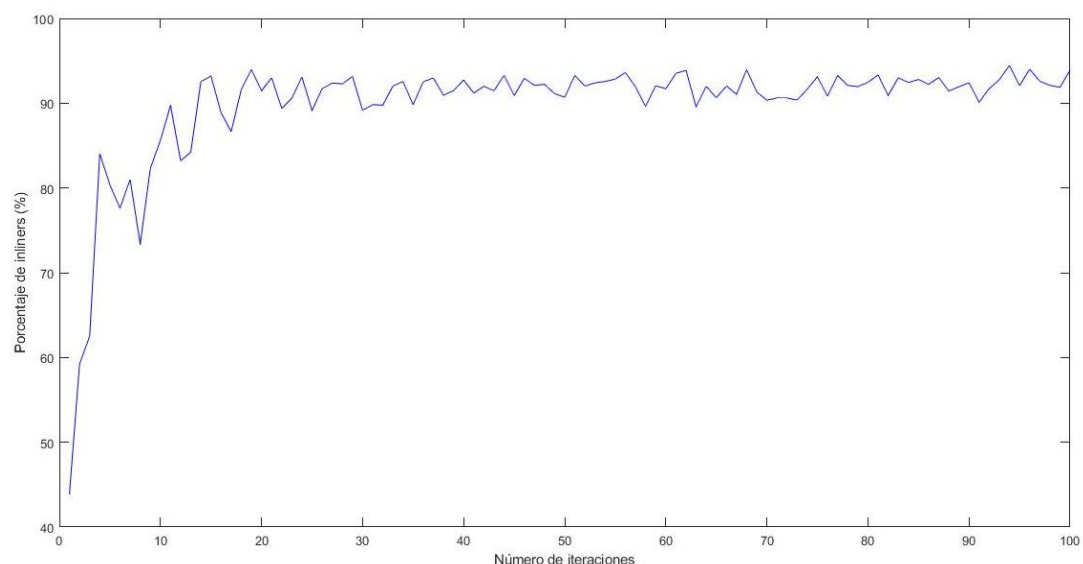


Fig. 7. Porcentaje de “inliers” respecto al número de iteraciones del método. Se observa que a partir de 20 iteraciones aproximadamente, el porcentaje de “inliers” se estabiliza.

Si se realiza un estudio del porcentaje de “inliers” en función del valor de “confidence”, se obtiene el resultado mostrado en la *Figura 8*.

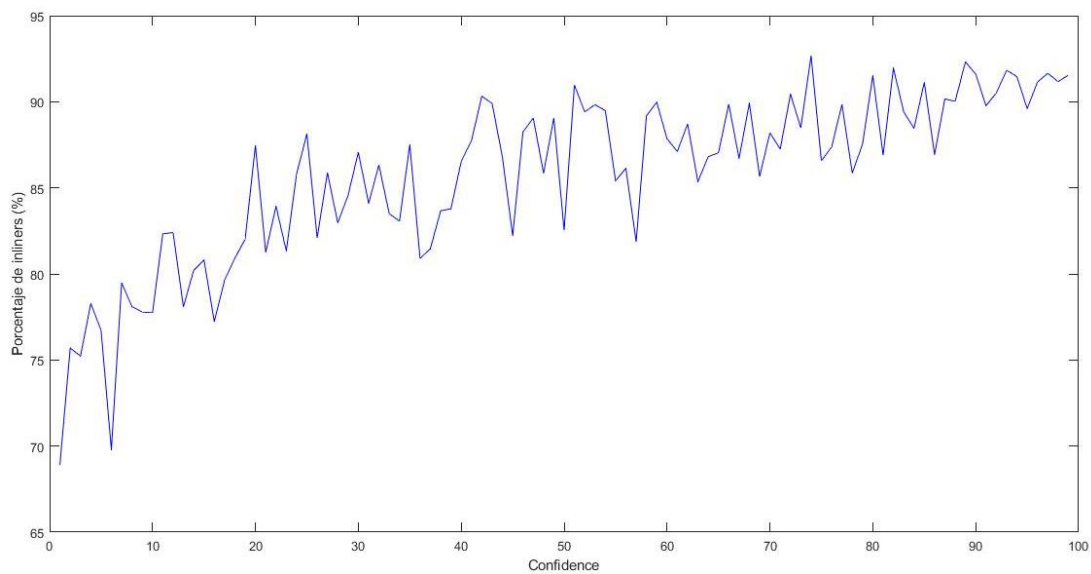


Fig. 8. Porcentaje de “inliers” en función del valor de “confidence”.

Puesto que se ha observado en estudios previos que hay pocos datos que tengan mala correspondencia, vamos a coger como valor de “confidence” igual a 95%, mientras que el número de iteraciones será de 50 (suficientes iteraciones sin comprometer el tiempo de computo).

Por último, con el número de iteraciones y con el valor de “confidence” establecido, se ha realizado un estudio del valor de “distance treshold”, el cual se muestra en la *Figura 9*.

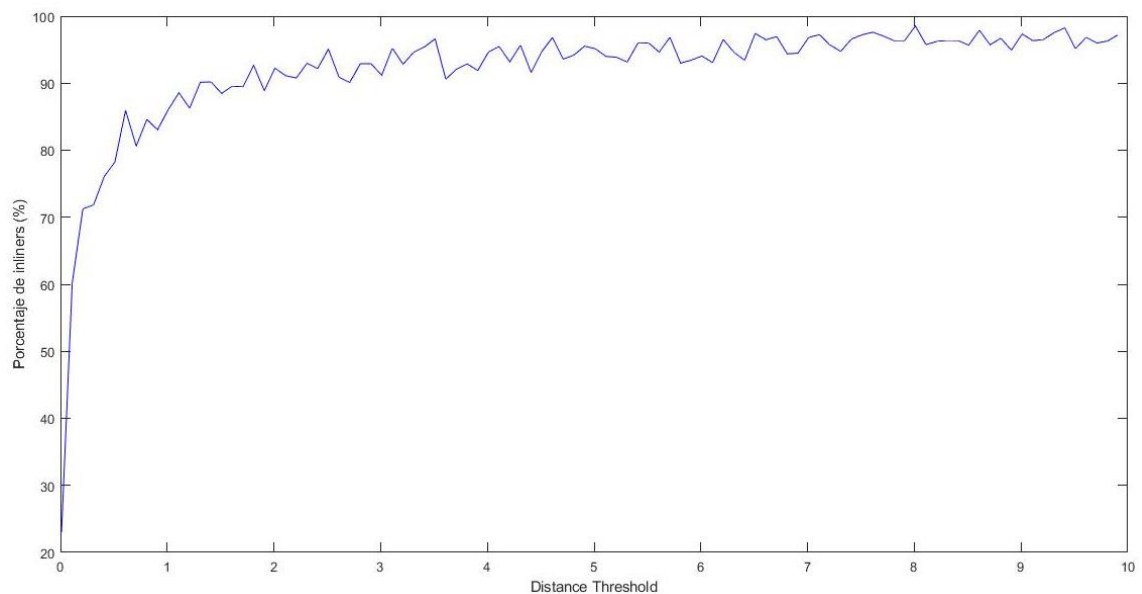


Fig. 9. Porcentaje de “inliers” en función del umbral de distancia.

En este caso, vamos a tomar como umbral de distancia el valor 2, de tal manera que haya un 90% de “inliers”.

Por último, es necesario comprobar el resultado obtenido. En primer lugar, se ha obtenido que el número medio de “inliers” obtenido es de aproximadamente el 90.75% (valor cauteloso pero suficiente para la correcta obtención de la matriz fundamental). En segundo lugar, después de un gran número de pruebas se ha obtenido que la media del error (calculado partiendo de que $m_2^T \cdot F \cdot m_1^T = 0$) es 0.0026 y que la varianza del error es del orden de 10^{-5} , es decir, un valor aceptable. Un ejemplo donde se muestra el error para todos los puntos casados considerados como buenos casamiento se muestra en la *Figura 10*.

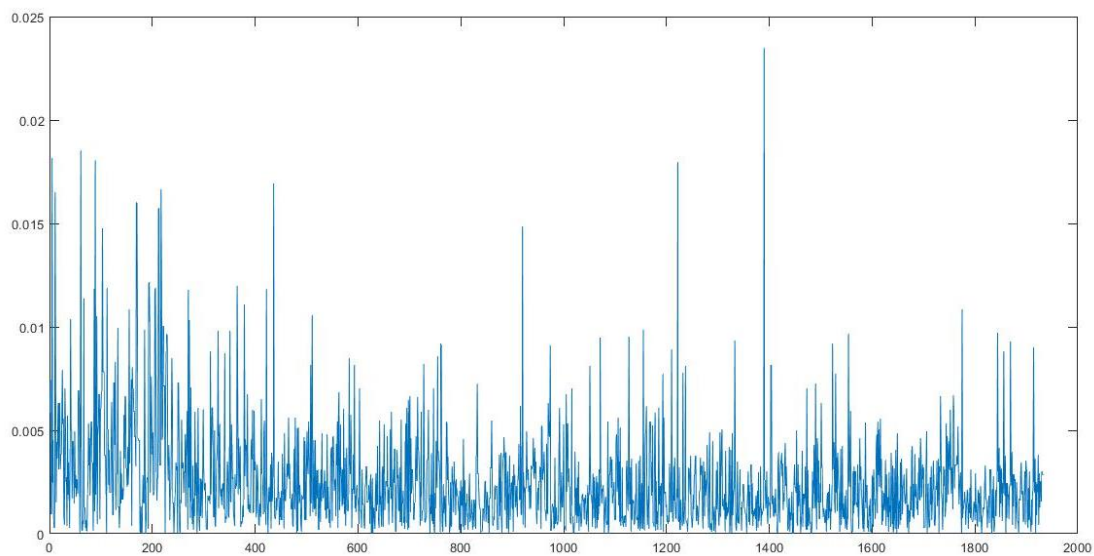


Fig. 10. Error para los diferentes casamientos entre dos imágenes. Existen valores cuyo error es superior (llegan a 0.03 en algunos casos).

En la *Figura 11* se representa la disparidad para los puntos casados (excluyendo los eliminados con el método RANSAC).

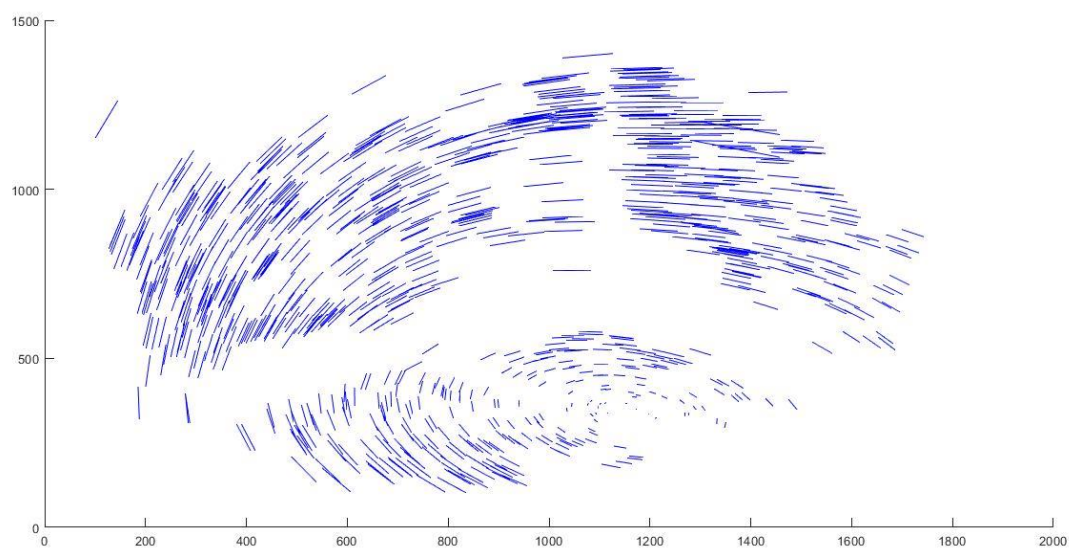


Fig. 11. Disparidad de los diferentes puntos casados, excluyendo los puntos eliminados por el método RANSAC (umbral de 0.5). Patrón en remolino. Existen un número inferior de puntos al mostrado en la Figura 5.

En la *Figura 12* podemos observar una comparativa de la disparidad para diferentes tipos de umbrales introducidos en la función “match”. A medida que aumenta el umbral, el número de correspondencias aumenta.

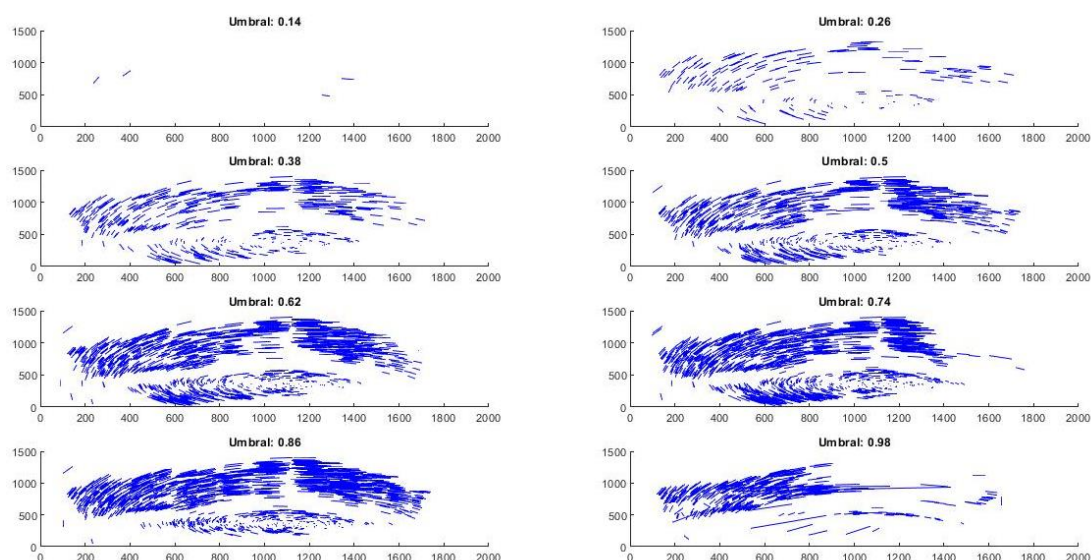


Fig. 12. Disparidad de los puntos entre la Imagen 1 y 2 del escenario 1, para diferentes valores de umbral de la función “match”. A medida que aumenta el umbral aumenta el número de características casadas (las gráficas comprenden un rango de umbral donde la primera imagen tienen un valor de 0.14, aumentando en 0.12 en cada fotografía, hasta llegar a 0.98).

Para la *Figura 12*, se puede observar que para el umbral mayor, existen disparidades incoherentes e incorrectas. Además, se han obtenido una serie de características representativas, las cuales se muestran en la *Tabla 1*. En esta tabla podemos observar diferentes aspectos:

- Cuando el umbral es muy pequeño ($\text{umbral} < 0.2$), se seleccionan pocas características en las imágenes, por lo que será una búsqueda rigurosa donde se procurará garantizar que los puntos casados son correctos. A pesar de ello, con un umbral muy pequeño el número de puntos casados es muy reducido, consiguiendo una matriz fundamental muy dependiente de los puntos casados.
- Cuando el umbral es pequeño ($0.2 < \text{umbral} < 0.55$), se selecciona un número medio-bajo de características, con un porcentaje de buenas correspondencias (inliners) muy elevado (entre el 87.5 y el 95%). La matriz fundamental en este rango se mantiene constante en todos los cálculos, siendo el error y la varianza muy reducidos.
- Cuando el umbral es moderado ($0.55 < \text{umbral} < 0.7$), se selecciona un número alto de características, con un porcentaje de buenas características alto. La matriz

fundamental se desvía un poco de la calculada anteriormente, siendo el error y la varianza ligeramente superior.

- Cuando el umbral es alto ($0.7 < \text{umbral} < 0.9$), se seleccionan un número muy alto de características, con un porcentaje de buenas características moderado (cerca del 70%). La matriz fundamental se empieza a desviar de la teóricamente correcta y el error y la varianza aumentan en más del 50%.
- Cuando el umbral es muy elevado ($0.9 < \text{umbral}$), se selecciona un número muy elevado de características, muchas de ellas con correspondencias erróneas. El número de buenas correspondencias se estima por RANSAC como un valor inferior al 50% (número muy malo que empeora a medida que aumenta el umbral). La matriz fundamental se desvía en mucha diferencia al resto.

Umbral	Matriz			Casados	Inliners	Error	Varianza
0,14	0	0	-0,004	10	80	9,00E-03	1,71E-02
	0	0	0,0051				
	0,0022	-0,0062	1				
0,26	0	0	-0,0038	380	92,89	9,10E-03	1,71E-02
	0	0	0,0066				
	0,0032	-0,0067	0,9999				
0,38	0	0	-0,0038	1142	87,74	9,30E-03	1,67E-02
	0	0	0,0066				
	0,0032	-0,0067	0,9999				
0,5	0	0	-0,0038	1990	94,82	9,10E-03	1,70E-02
	0	0	0,0066				
	0,0032	-0,0067	0,9999				
0,62	0	0	-0,0038	2860	91,82	9,10E-03	1,69E-02
	0	0	0,0067				
	0,0032	-0,0068	0,9999				
0,74	0	0	-0,0039	3780	77,28	9,10E-03	1,71E-02
	0	0	0,0065				
	0,0033	-0,0067	0,9999				
0,86	0	0	-0,0037	5161	67,12	9,90E-03	1,64E-02
	0	0	0,0068				
	0,0031	-0,0069	0,9999				
0,98	0	0	-0,0025	9280	13,38	9,10E-03	1,70E-02
	0	0	0,0055				
	0,0013	-0,0062	1				

Tabla 1: Resultados obtenidos para el Escenario 1 con los diferentes parámetros para un rango de umbrales. La columna “Matriz” indica la matriz fundamental, la columna “Casados” indica las características casadas en ambas imágenes y la columna “Inliners” indica el porcentaje (en %) de casamientos tomados para calcular la matriz fundamental mediante el método robusto RANSAC. Las columnas “Error” y “Varianza” se han estimado mediante la matriz fundamental sobre todos los puntos casados en ambas imágenes.

Como conclusión se obtiene que un buen valor de umbral oscila entre 0.2 y 0.55 (los errores globales son prácticamente iguales, por lo que se ha estimado ese rango debido a la similitud en la matriz fundamental y al porcentaje de inliers). Personalmente, un valor que se considera con unos resultados muy correctos es el valor de umbral igual a 0.45. Este valor de umbral se utilizará para conocer la matriz fundamental, la cual queda reflejada en la *Tablas 2*, para las diferentes imágenes.

Escenario 1	Imagen 1			Imagen 2			Imagen 3		
Imagen 1	0	-0,0005	0,4181	0	0	-0,0038	0	0	-0,001
	0	0	0,9084	0	0	0,0066	0	0	0,0047
	-0,4181	-0,9084	0	0,0032	-0,0068	0,9999	0,0001	-0,005	1
Imagen 2	0	0	0,0032	0	0,0007	-0,572	0	0	0,0011
	0	0	-0,0068	0	0	0,8203	0	0	0,0099
	-0,0038	0,0067	0,9999	0,572	-0,8203	0	-0,0022	-0,0103	0,9999
Imagen 3	0	0	0	0	0	-0,002	0	-0,001	0,5992
	0	0	-0,005	0	0	-0,0103	0	0	-0,8006
	-0,0009	0,0047	1	0,0009	0,0099	0,9999	-0,5992	0,8006	0

Tabla 2: Matrices fundamentales entre las diferentes imágenes para el Escenario 1.

Podemos sacar las siguientes conclusiones:

- Entre imágenes iguales se obtiene como matriz fundamental una matriz antisimétrica, es decir, su transpuesta es igual a su negativa (tiene la diagonal principal nula).
- Entre la imagen 'x' y la imagen 'y' se obtiene una matriz fundamental que es la matriz fundamental transpuesta entre la imagen 'y' y la imagen 'x'.
- Entre imágenes iguales, las correspondencias son perfectas, es decir, se obtiene un 100% de "inliers" mediante RANSAC.

De la misma manera podemos hacer el estudio para el Escenario 2, obteniendo los resultados de la *Figura 13*, cuyas características se resumen en la *Tabla 3*, y cuyas matrices fundamentales se encuentran en la *Tabla 4*.

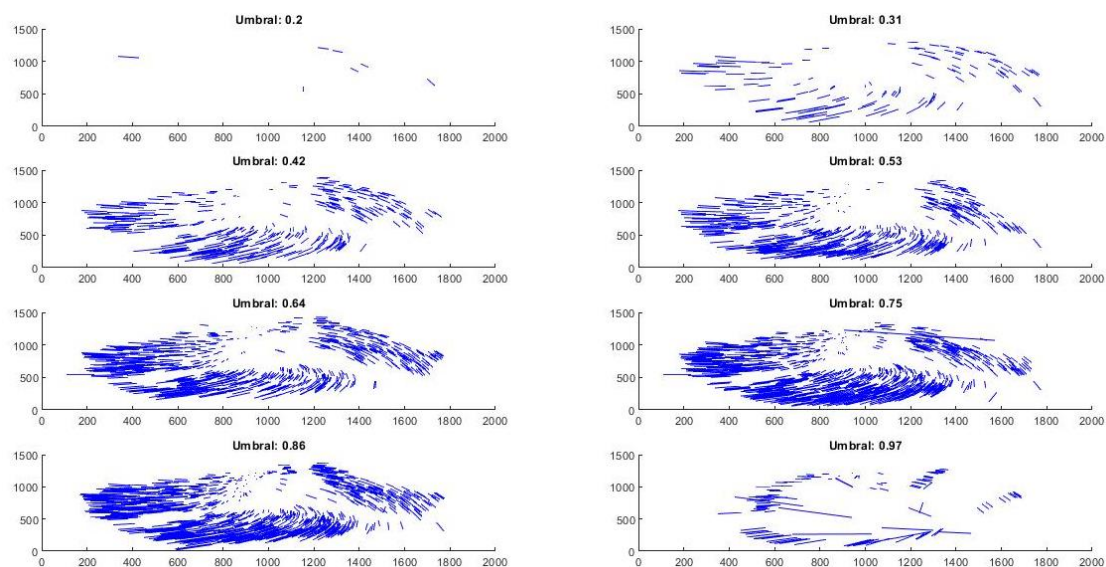


Fig. 13. Disparidad de los puntos entre la Imagen 1 y 2 del escenario 2, para diferentes valores de umbral de la función “match”. A medida que aumenta el umbral aumenta el número de características casadas (las gráficas comprenden un rango de umbral donde la primera imagen tienen un valor de 0.2, aumentando en 0.11 en cada fotografía, hasta llegar a 0.97).

Al igual que con el Escenario 1, el umbral mayor representa disparidades incorrectas.

Umbral	Matriz			Casados	Inliners	Error	Varianza
0,2	0	0	-0,0035	16	93,75	1,46E-02	9,80E-03
	0	0	0,0174				
	0,0026	-0,0183	0,9997				
0,31	0	0	-0,0029	271	92,25	1,33E-02	9,20E-03
	0	0	0,017				
	0,002	-0,0179	0,9997				
0,42	0	0	-0,0029	792	91,16	1,43E-02	7,30E-03
	0	0	0,0173				
	0,002	-0,0181	0,9997				
0,53	0	0	-0,0038	1439	75,26	1,35E-02	9,20E-03
	0	0	0,0184				
	0,0028	-0,0193	0,9996				
0,64	0	0	-0,0025	2003	75,29	1,35E-02	8,90E-03
	0	0	0,0167				
	0,0016	-0,0175	0,9997				
0,75	0	0	-0,0048	2715	72,3	1,40E-02	9,70E-03
	0	0	0,0189				
	0,0039	-0,02	0,9996				
0,86	0	0	-0,004	3959	57,36	1,37E-02	8,70E-03
	0	0	0,0166				
	0,003	-0,0177	0,9997				
0,97	0	0	-0,0025	7963	3,33	1,33E-02	9,20E-03
	0	0	0,0055				
	0,0013	-0,0062	1				

Tabla 3. Resultados obtenidos para el Escenario 2 con los diferentes parámetros para un rango de umbrales. El significado de las diferentes columnas queda explicado en el pie de la Tabla 2.

Gracias a la Tabla 3, podemos concluir que el error para los diferentes umbrales es similar, sin embargo, el número de inliners disminuye al superar el valor de 0.5 y disminuye más drásticamente aún cuando se supera el valor de 0.8. De esta manera, un valor válido y que será utilizado para calcular la matriz fundamental será el de un umbral de 0.45, al igual que en el Escenario 1.

Escenario 2	Imagen 1			Imagen 2			Imagen 3		
Imagen 1	0	0,0008	-0,6454	0	0	-0,0028	0	0	0,0002
	-0,0008	0	0,7639	0	0	0,0173	0	0	0,0074
	0,6454	-0,7639	0	0,0019	-0,0181	0,9997	-0,001	-0,008	0,9999
Imagen 2	0	0	0,0016	0	-0,0008	0,6242	0	0	0,0174
	0	0	-0,0175	0,0008	0	-0,7812	0	0	0,0367
	-0,0024	0,0166	0,9997	-0,6242	0,7812	0	-0,0186	-0,0371	0,9983
Imagen 3	0	0	-0,001	0	0	-0,0186	0	0,0004	-0,9229
	0	0	-0,0085	0	0	-0,0369	-0,0004	0	0,3849
	0,0002	0,008	0,9999	0,0174	0,0365	0,9983	0,9229	-0,3849	0

Tabla 4: Matrices fundamentales entre las diferentes imágenes para el Escenario 2.

Filtrado de falsas correspondencias (Error residual)

Otra opción para asegurar la correcta elección de los puntos casados es la eliminación de los puntos con un elevado error residual en el cálculo de la Matriz Fundamental.

Una primera opción se había planteado de tal manera que, una vez obtenida la media del error de todos los puntos correctamente casados mediante RANSAC, se rechazarán aquellos puntos cuyo valor absoluto del error supere en 3 veces la media del error. El resultado hace que se pase del error obtenido en la *Figura 10* al obtenido en la *Figura 14*, donde se puede observar que los valores de error altos han sido eliminados, homogeneizándose la media del error.

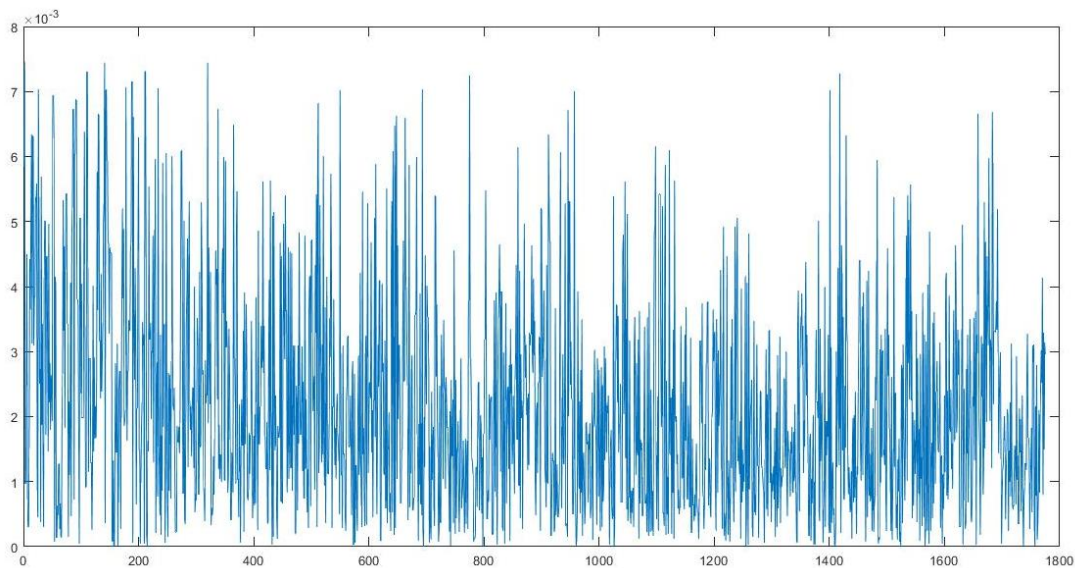


Fig. 14. Error para los diferentes casamientos entre dos imágenes. El valor máximo de error es de 0.0078 después de aplicar un filtro donde únicamente han pasado los valores inferiores a 3 veces la media.

Gracias a ello, la media del error se ha reducido, mientras que el número de datos tomados para el cálculo de la matriz fundamental se ha reducido. Además, el valor de la Matriz fundamental se ha modificado de la siguiente manera:

$$F_{ini} = \begin{pmatrix} 0 & 0 & -0.0038 \\ 0 & 0 & 0.0066 \\ 0.0032 & -0.0067 & 0.9999 \end{pmatrix} \rightarrow F_{fin} = \begin{pmatrix} 0 & 0 & -0.0039 \\ 0 & 0 & 0.0065 \\ 0.0032 & -0.0067 & 0.9999 \end{pmatrix}$$

Como se puede observar, la variación en la Matriz Fundamental es mínima. En la *Figura 15* podemos observar la disparidad una vez eliminados los valores cuyo error es superior a tres veces la media.

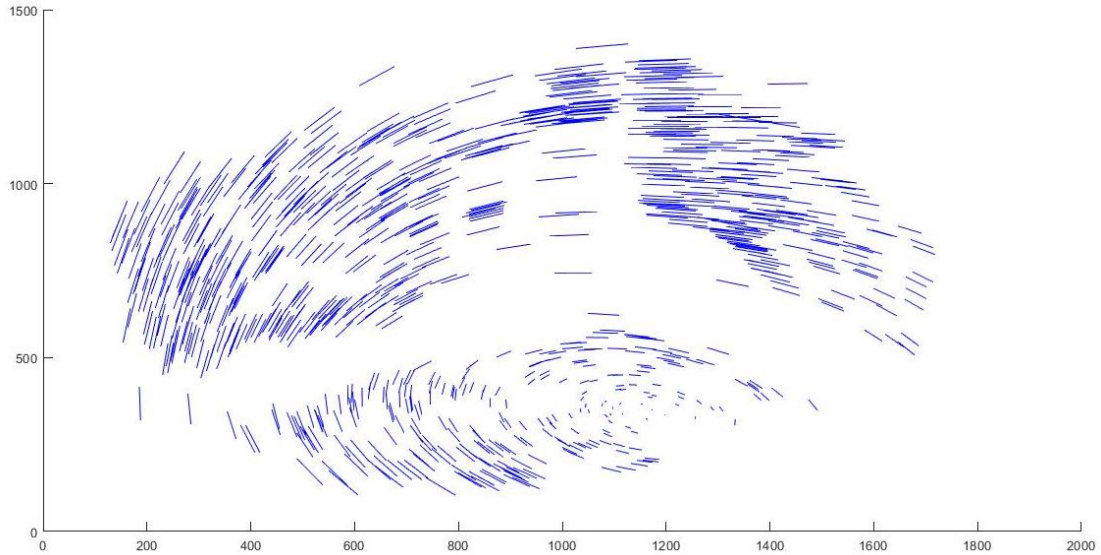


Fig. 15. Disparidad de los diferentes puntos casados, excluyendo los puntos eliminados por el método RANSAC y por el filtro del error. Patrón en remolino. Existen un número inferior de puntos al mostrado en la Figura 11.

Sin embargo, aplicar un filtro sobre las correspondencias ya filtradas mediante RANSAC no es muy útil. De esta manera se ha procedido a filtrar las correspondencias no filtradas, obteniendo los resultados mostrados en la *Figura 16*.

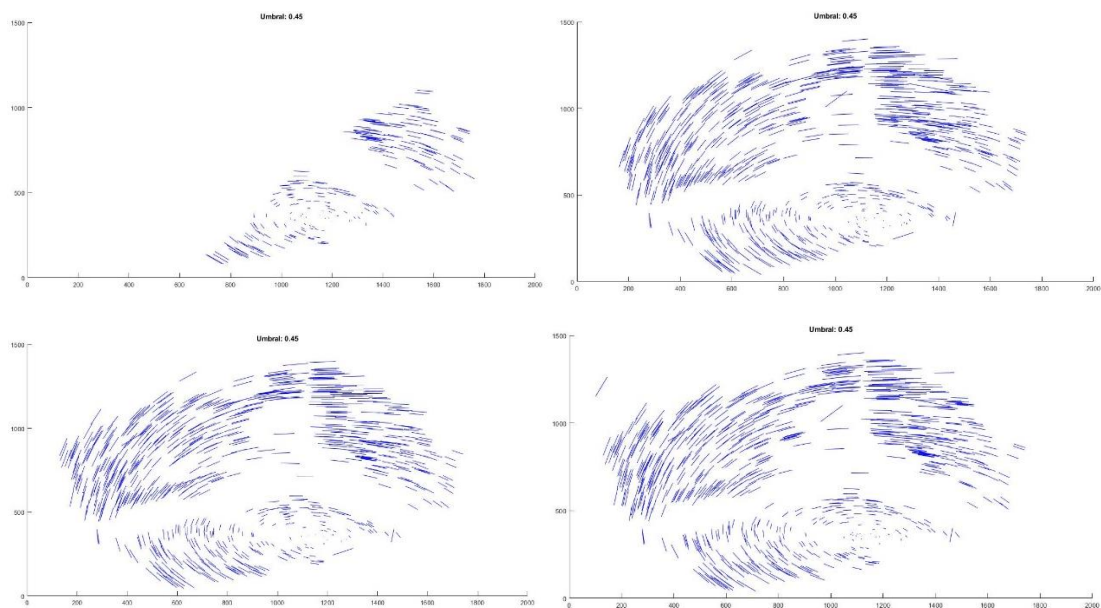


Fig. 16. Comparativa de los resultados de la disparidad para un filtro basado en el error para diferentes valores de paso. En la primera imagen se eliminan los errores superiores a 2 veces la media, en la segunda imagen los de 2.25, en la tercera los de 2.5 y en la cuarta los errores superiores a 3 veces la media.

En dicha figura se puede observar que los resultados no son del todo correctos, sin embargo, la matriz fundamental es idéntica a la obtenida por medio del método robusto RANSAC. Por último, el mejor valor para el filtro se ha determinado que es de 2.5 veces el error medio.

Comparativa de los métodos de filtrado

Ambos métodos (RANSAC y filtro por error) nos aportan idénticas matrices fundamentales. Sin embargo, nos quedaremos con el método de RANSAC debido a que es un método robusto. En la *Figura 17* se puede observar el resultado obtenido para ambos métodos (disparidad).

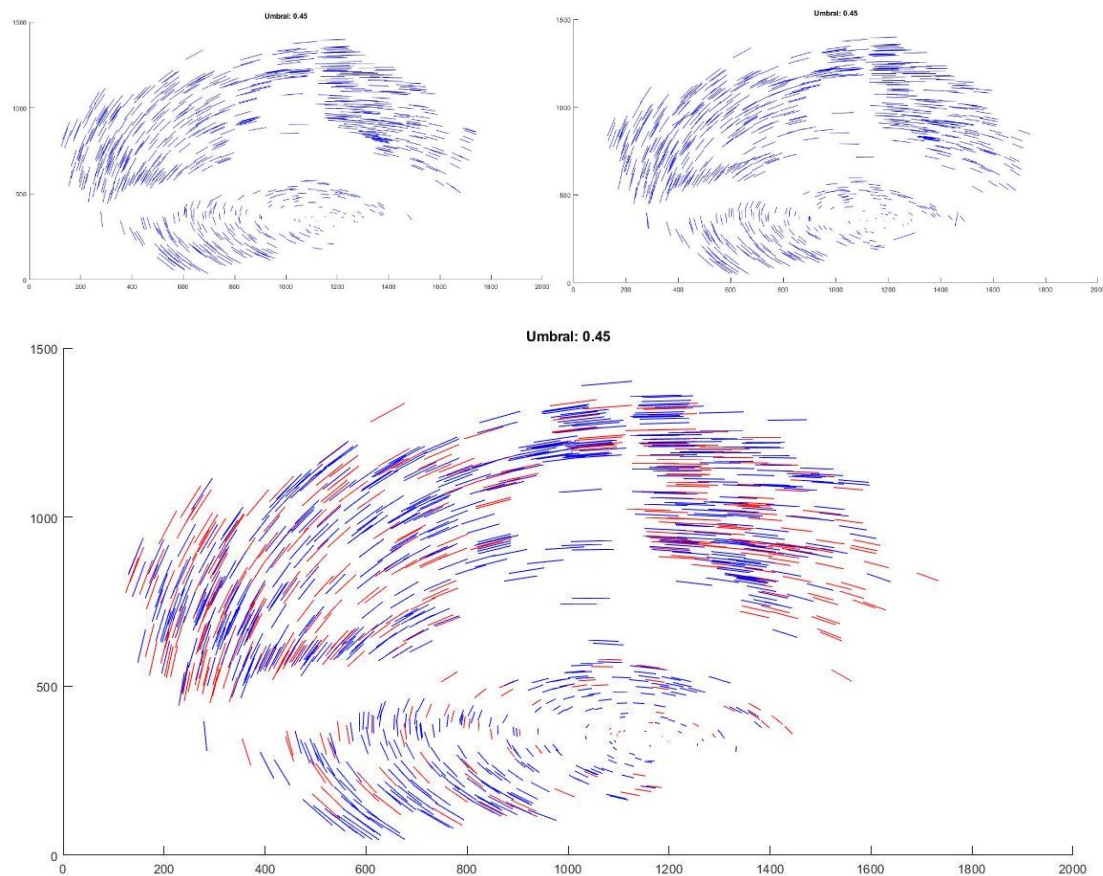


Fig. 17. Comparativa de ambos métodos para filtrar las malas correspondencias. Arriba a la izquierda se muestra el método RANSAC. Arriba a la derecha el método basado en el error. Abajo se muestran ambas juntas, siendo la parte azul la parte del RANSAC y la roja la del filtro por el error.

1.4. RECONSTRUCCIÓN PROYECTIVA 3D

La Reconstrucción Proyectiva 3D de la escena permite incorporar información de profundidad detectando los puntos más cercanos y los puntos más lejanos (información relativa).

Basándonos en las *Tablas 1 y 3*, vamos a tomar como umbral un valor de 0.65, ya que el porcentaje de “inliners” sigue siendo elevado y el mapa sería bastante denso.

El primer paso para calcular la profundidad de cada punto es el cálculo del epipolo en la segunda cámara. Para ello, con la matriz fundamental calculada, se hará uso de la siguiente ecuación:

$$\tilde{F}^T \tilde{e}' = 0$$

De esta manera, calcularemos la posición en ‘x’ e ‘y’ del epipolo en la segunda cámara. Con este valor, calcularemos la posición de cada punto en coordenadas del mundo mediante reconstrucción proyectiva de la siguiente manera:

$$X_p = x_l \quad ; \quad Y_p = y_l \quad ; \quad Z_p = \sqrt{(e_{x2} + d_x)^2 + (e_{y2} + d_y)^2} \quad \text{con} \quad d = \begin{bmatrix} d_x & d_y \end{bmatrix} = \begin{bmatrix} x_2 - x_l & y_2 - y_l \end{bmatrix}$$

Para conseguir una mayor precisión, se han tomado las imágenes 1 y 3 de ambos escenarios, quedando representado en falso color la reconstrucción proyectiva según se muestra en la *Figura 18 y 19* para el Escenario 1 y 2 respectivamente. Para ambas figuras, se ha realizado una purga de falsas correspondencias a través de RANSAC, por lo que está garantizado que todos los puntos existentes en las imágenes son correctos.

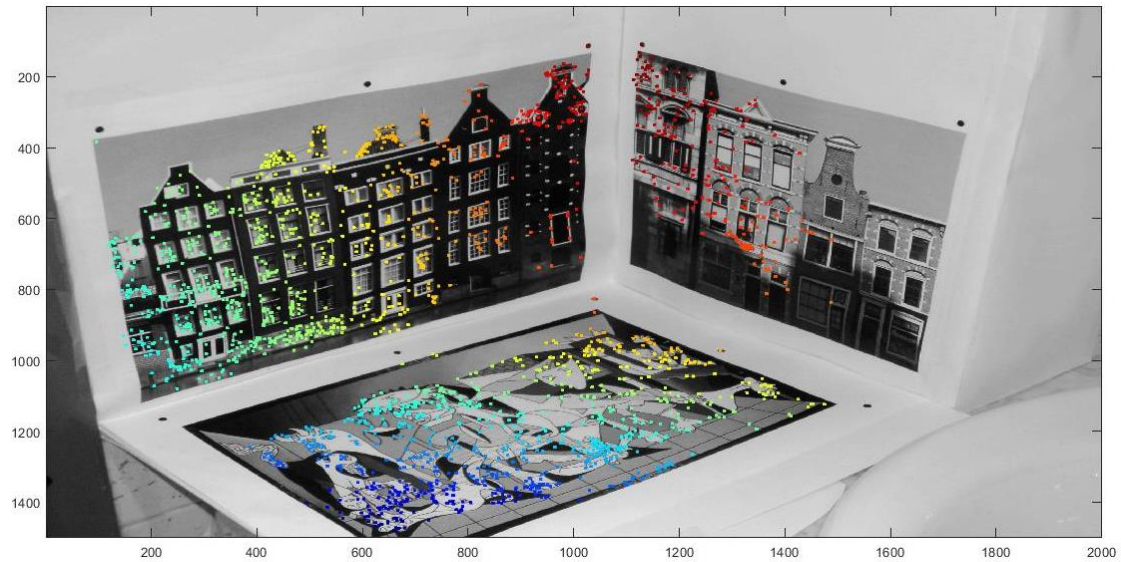


Fig. 18. Reconstrucción Proyectiva entre la imagen 1 y 3 del Escenario 1, con falso color.

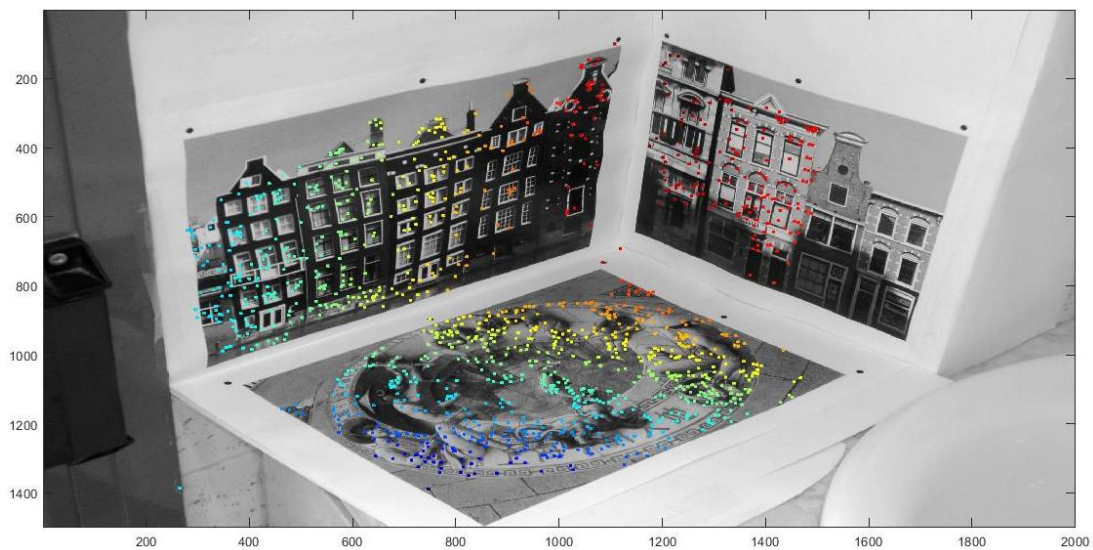


Fig. 19. Reconstrucción Proyectiva entre la imagen 1 y 3 del Escenario 2, con falso color.

El falso color sigue el patrón mostrado en la *Figura 20*, donde el color azul indica punto más cercano, y punto rojo indica punto más lejano.



Fig. 20. Rango de colores utilizado.

En la *Figura 21 (a-c)* se representa una serie de imágenes de la Reconstrucción Proyectiva obtenida en el Escenario 1 (en el archivo .zip adjunto se puede cargar la imagen en archivo Matlab para poder observarla de una manera más cómoda, llamada "Proyectiva_E1"), mientras que en la *Figura 22 (a-c)* se representa el Escenario 2 ("Proyectiva_E2").

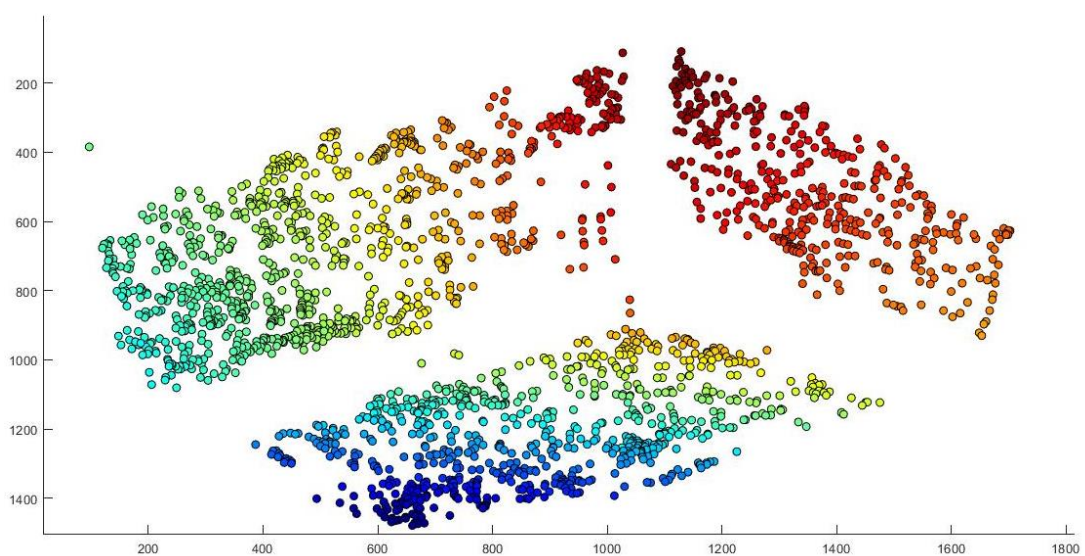


Fig. 21.a.

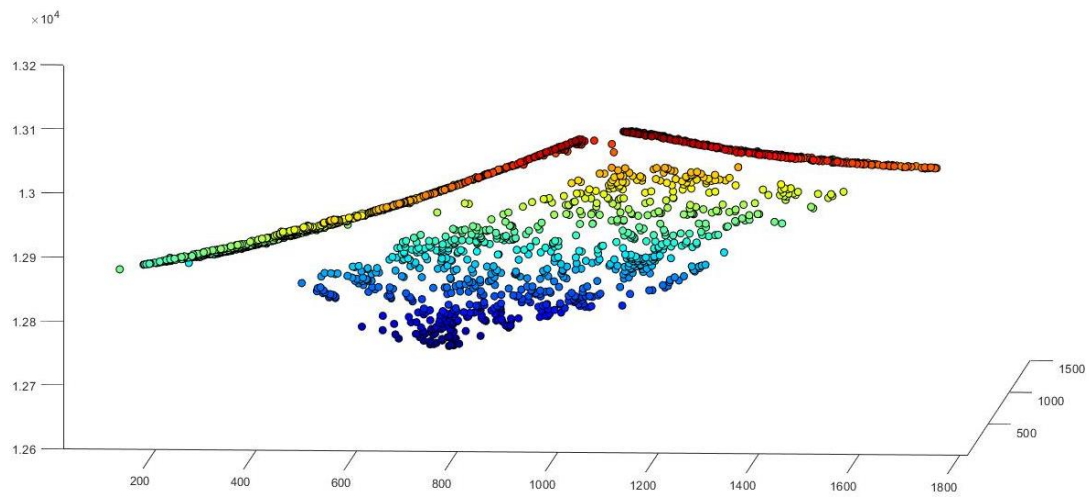


Fig. 21.b.

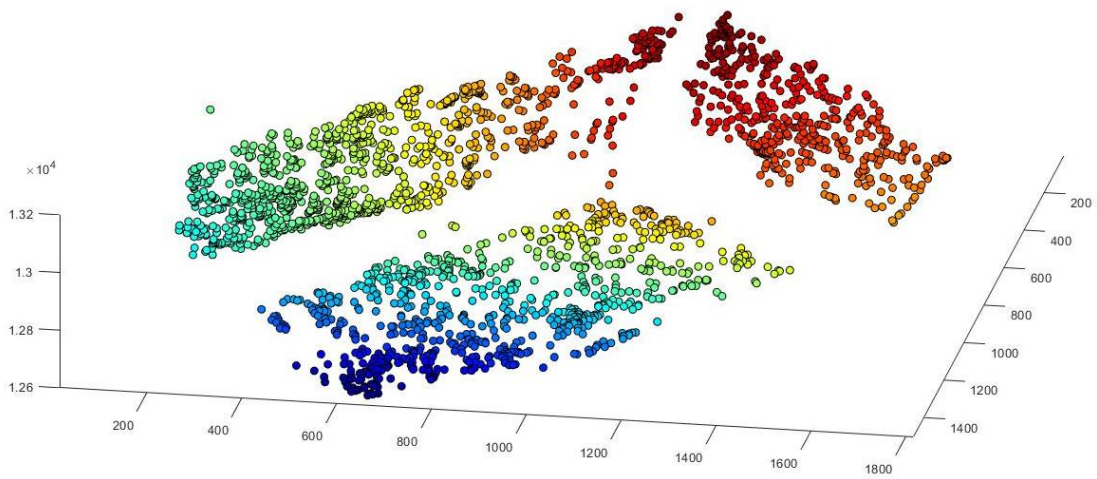


Fig. 21.c.

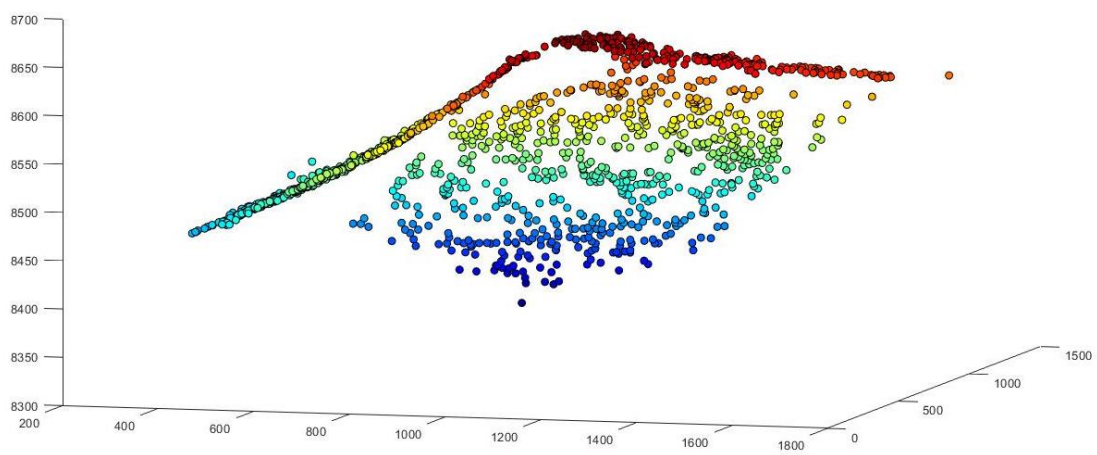


Fig. 22.a.

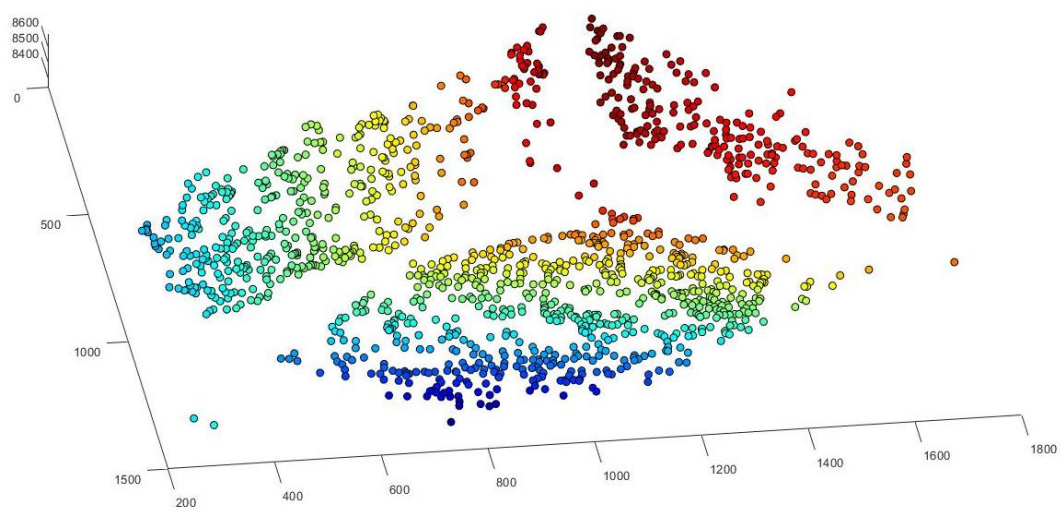


Fig. 22.b.

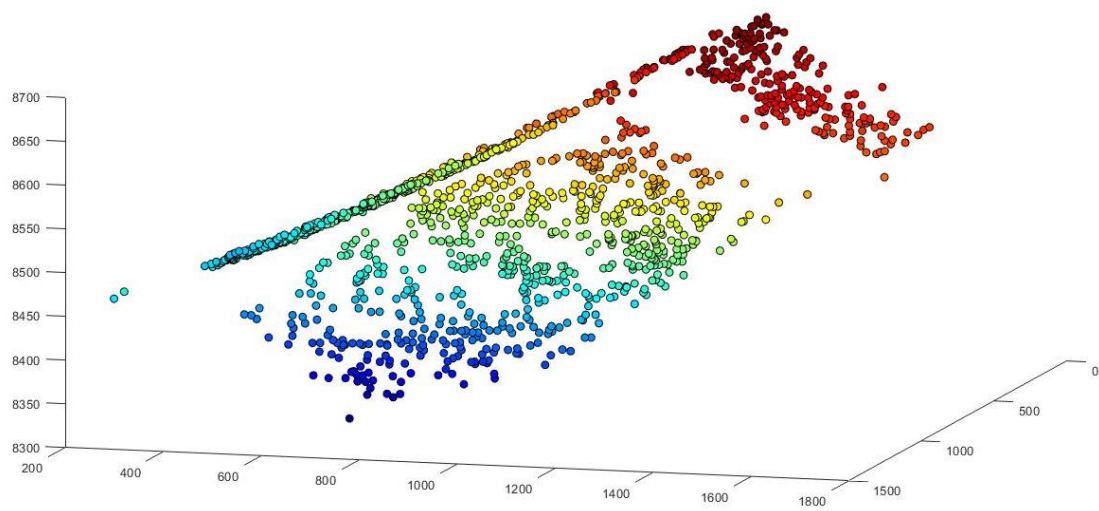


Fig. 22.c.

Referencias

- [1] J. B. Pablo Flores, «Algoritmo SIFT: fundamento teórico,» p. 5, 2019.