

Multispectral Registration, Undistortion and Tree Detection for Precision Agriculture

Alfonso López¹, Juan M. Jurado¹, Carlos J. Ogayar¹, Francisco R. Feito¹

¹Department of Computer Science, University of Jaén

Abstract

Multi-lens multispectral cameras allow us to record multispectral information for a whole area of terrain, even though we may only need the vegetation data. Based on the intensity of each multispectral image we can retrieve the contours of the trees that appear on the recorded terrain. However, multispectral cameras use a physically different lens for each range of wavelengths and misregistration effects could appear due to the different viewing positions. As these types of lenses are dedicated to capture larger areas of terrain, their focal distance is lower and because of this we get what is called a fisheye distortion. Therefore if we want to retrieve the shape of each tree and its multispectral data we need to process the channels so them all are represented as undistorted images under a same reference system.

Keywords: Image registration, Multispectral image, Distortion removal, Image segmentation

1. Introducción

La información multispectral es una de las principales herramientas en el área de la agricultura de precisión, ya que nos permite monitorizar una plantación y analizar las posibles enfermedades presentes a través de múltiples índices derivados de las combinaciones de imágenes multispectrales.

Las cámaras multispectrales nos permiten trabajar, de forma general, con al menos cuatro bandas: Red (RED), Green (GRE), Rededge (REG) y Near Infrared (NIR), aunque podrían incluir otras como Blue (BLU). Cada imagen capturada aporta información diferente ya que cada una trabaja con un determinado rango de longitudes de onda. Se puede establecer así una estructura de imágenes multicapa de tal forma que nos permita consultar el valor de intensidad de un píxel para cada imagen y calcular índices para un punto de la escena que podría pertenecer a la vegetación de interés, lo que posibilitaría hacer un seguimiento de la plantación.

La captura de cada banda multispectral se produce en condiciones ligeramente diferentes: por ejemplo, la posición desde la que se captura la escena varía para cada una de las imágenes debido a la estructura del dispositivo. Tampoco se puede suponer que los planos de las imágenes capturadas por las cuatro lentes son paralelos, dado que cada una tiene asociada un eje óptico (y por lo tanto un punto principal (c_x, c_y)). A raíz de esta situación de captura se produce una desalineación entre imágenes que debe necesariamente eliminarse si se desea obtener información fiable en diferentes capas para la vegetación que se está tratando.

Por otro lado, las características físicas de las lentes provocan

una distorsión *fish-eye* en los extremos de las imágenes que también debería revertirse antes de buscar la vegetación de interés, especialmente pensando en hacer un seguimiento de polígonos (vegetación) a través de diferentes imágenes que pueden capturar un mismo área de terreno.

El objetivo de este trabajo es proponer un método cuya salida sean imágenes multispectrales sin distorsión y que representen exactamente el mismo área del terreno. Una vez que se logra revertir todos los problemas citados se podrán extraer los contornos de la vegetación que buscamos, de tal forma que se limitan las zonas hacia las que dirigir nuestro esfuerzo de monitorización de una plantación.

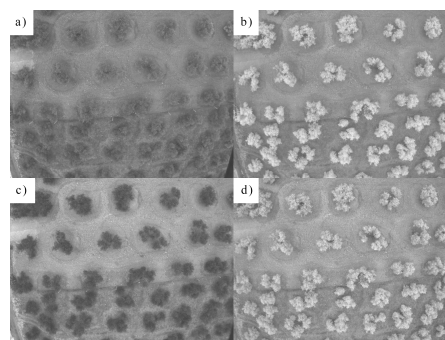


Figure 1: Imágenes multispectrales, capturadas por el dispositivo Parrot Sequoia, que forman el modelo multicapa: a) GRE b) NIR c) RED d) REG.

1.1. Trabajos anteriores

Múltiples métodos han sido propuestos para solucionar el problema del registro de las imágenes multispectrales, y normalmente podemos dividirlos en dos grandes corrientes. Una primera corriente se basa en la obtención de características comunes entre imágenes [SXZJ14], con el inconveniente de que la intensidad de cada capa es distinta del resto (representan longitudes de onda diferentes) y por lo tanto, es posible que no se puedan alinear todos los pares de imágenes posibles. Una segunda corriente se basa en el uso de los parámetros contenidos en cada una de las imágenes [JRH17], lo cual presenta también el inconveniente de que pueden existir errores de calibración en la cámara que influirían directamente en dicha información.

Una vez se registran las imágenes el objetivo es monitorizar la plantación. El análisis más común consiste en calcular índices a partir de la combinación de imágenes multispectrales para detectar situaciones como las enfermedades de los árboles. Sin embargo, no todo lo que se captura en las imágenes es útil para el análisis. Una solución bastante frecuente es definir polígonos que engloban todas las zonas de interés a grandes rasgos, en lugar de distinguir de forma exacta la vegetación [CRDG*15], sin embargo, si se quiere hacer un seguimiento de cada árbol a lo largo del tiempo es necesaria una distinción de cada individuo.

2. Metodología

En este apartado se presentan las soluciones que se proponen para obtener los contornos de la vegetación de interés partiendo de imágenes multispectrales obtenidas a través del vuelo de un dron.

2.1. Eliminación de la distorsión de ojo de pez

Este tipo de distorsión se produce en lentes con una distancia focal muy pequeña que permite abarcar una mayor superficie de terreno. En la Figura 2 se puede observar claramente esta situación, especialmente en los extremos de la imagen, y con mayor intensidad aún en las esquinas de la misma. Se puede establecer así un modelo donde el píxel $p_{x,y}$ de la imagen resultante, de mismas dimensiones que la imagen original, toma el valor de intensidad de la posición (i, j) de la imagen distorsionada, donde $i, j \in \mathbb{R}$.

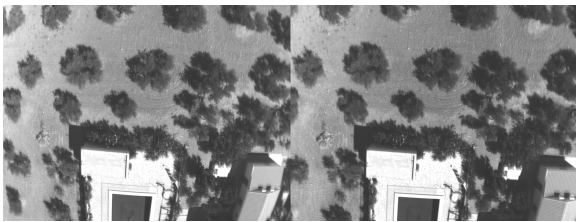


Figure 2: Imagen multispectral RED antes y después de rectificar la distorsión de ojo de pez.

Para llevar a cabo esta rectificación serán necesarias dos funciones, $f(x)$ y $f(y)$, donde $f(x)$ nos permite acceder a i y $f(y)$ a j . A partir de la Figura 3 se puede comprender con mayor facilidad la ecuación 1, que nos permite obtener el valor α , el cual representa

el ángulo normalizado en el intervalo $[0, 1]$ (el ángulo original se encuentra en $[0, \pi/2]$) entre el eje óptico y un segmento $c_0 p_{x,y}$, sea c_0 el centro óptico y c_p el punto principal de la imagen.

$$\alpha = \frac{2}{\pi} * \tan^{-1} \left(\frac{\sqrt{(x - c_{p_x})^2 + (y - c_{p_y})^2}}{f} \right) = \frac{2}{\pi} * \tan^{-1} \left(\frac{r_1}{f} \right) \quad (1)$$

A partir de los cuatro coeficientes de distorsión contenidos en los metadatos de la imagen (k_0, k_1, k_2, k_3), se puede obtener un nuevo valor β (2) que representa el ángulo existente entre el eje óptico y el segmento $c_0 p_{i,j}$.

$$\beta = k_1 + k_2 * \alpha + k_3 * \alpha^2 + k_4 * \alpha^3 \quad (2)$$

Si observamos la Figura 3, se puede establecer una relación entre triángulos similares, T1 ($p_{x,y}$, Proyección de $p_{x,y}$ en el eje horizontal, c_p) y T2 ($p_{i,j}$, Proyección de $p_{i,j}$ en el eje horizontal, c_p). Teniendo en cuenta que se desea conocer el valor de i y j , se pueden obtener dichos términos de 3, aplicar 4 y desnormalizar β , lo que nos conduce a 5,6:

$$\frac{x - c_{p_x}}{r_1} = \frac{i - c_{p_x}}{r_2}, \frac{y - c_{p_y}}{r_1} = \frac{j - c_{p_y}}{r_2} \quad (3)$$

$$\tan(\beta) = \frac{r_2}{f} \quad (4)$$

$$f(x) = \frac{\pi * f * \tan(\beta) * (x - c_{p_x})}{r_1 * 2} + c_{p_x} = i \quad (5)$$

$$f(y) = \frac{\pi * f * \tan(\beta) * (y - c_{p_y})}{r_1 * 2} + c_{p_y} = j \quad (6)$$

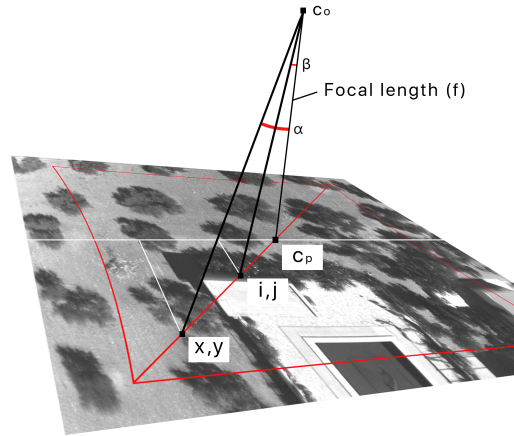


Figure 3: Modelo para la rectificación de la distorsión de ojo de pez planteado sobre una imagen multispectral. El área de terreno contenido en la imagen resultante se representa mediante una forma de color rojo (a modo de ejemplo). Asimismo, se muestra la relación de un píxel de la imagen sin distorsión, $p_{x,y}$, con una posición (i, j) de la imagen original.

Conocidas las dimensiones de la imagen, se puede aplicar $f(x)$

y $f(y)$ a todos los píxeles de la misma, obteniéndose así para cada uno dos valores, i y j , que se corresponden con una posición cualquiera dentro de un píxel de la imagen distorsionada. Por lo tanto, parece necesario considerar un algoritmo de interpolación, como el bilineal, para obtener el valor de intensidad final de un píxel, como se describe en 7. Una vez se aplique esta operación se obtendrá una imagen sin distorsión como la de la Figura 2.

$$dst(x,y) = Bilinear(src, (f(x), f(y))) \quad (7)$$

2.2. Registro de imágenes

El problema de la falta de alineación entre imágenes se puede observar en la Figura 4, donde se produce un efecto de *image ghosting* derivado de la distancia física entre lentes y su orientación al combinar todas las imágenes a través del canal *alpha*. Esto se produce ya que el sistema de referencia para cada imagen es distinto del resto.

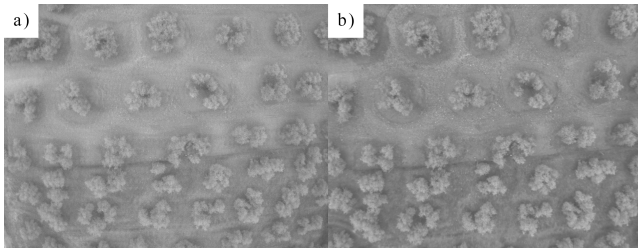


Figure 4: a) Efecto de *image ghosting* mediante la superposición de las bandas iniciales empleando el canal *alpha*. b) Superposición de imágenes corregidas utilizando el canal *alpha*.

Conocidos los parámetros de captura de la imagen, como los coeficientes internos de la cámara, se podría optar por un registro basado en la rectificación de las bandas a partir de dichos parámetros, con la desventaja de que pueden existir errores en este proceso derivados de la calibración del propio dispositivo [JRHC17]. Estos errores se podrían solucionar a su vez mediante algún proceso de corrección o estimación de parámetros como RANSAC. Sin embargo, en este trabajo se propone un enfoque más genérico e independiente de los metadatos de la imagen mediante una detección de características comunes entre las diferentes bandas.

En la solución planteada se utiliza el algoritmo Enhanced Correlation Coefficient [EP08], que utiliza un esquema iterativo de orden lineal basado en un análisis de correlación entre imágenes, con la ventaja de que se puede adaptar la precisión del algoritmo a cualquier situación, trabaja a nivel de subpíxel y es invariante respecto a distorsiones fotométricas.

Como sucede en cualquier proceso del registro de imágenes, se debe establecer una imagen de referencia que podría ser cualquiera de las cuatro imágenes multiespectrales disponibles, aunque en nuestro algoritmo se utiliza la banda GRE, tal y como establece el dispositivo de captura de imágenes multiespectrales Parrot Sequoia.

A pesar de las características deseables del algoritmo ECC, es necesario buscar aquellos pares de bandas más afines en cuanto a

intensidad [JRHC17] [YES*18] con el fin de aumentar la fiabilidad del proceso, de tal forma que la operación de registro podría constar de varios niveles, como se propone a continuación:

- NIR → REG → GRE
- REG → GRE
- RED → GRE

Hasta dos parámetros son necesarios para definir el comportamiento del algoritmo: precisión que se desea alcanzar (aunque es posible que no converja a ese nivel) y número de iteraciones máximo, obteniéndose una matriz de transformación, M_i , que nos permitirá pasar de una banda B_i a la imagen de referencia GRE en última instancia.

Dado que los planos de las imágenes obtenidas no son paralelos, dicha matriz M_i debe representar al menos una transformación afín, estimándose así 6 parámetros (tamaño 2x3). De forma más genérica, si se desconociera la relación entre planos podría calcularse una transformación de homografía donde habrá que estimar hasta 8 parámetros (y por lo tanto el proceso será más costoso).

Llegados a este punto nuestras bandas se hallan en el mismo sistema de referencia, sin embargo, no todas han sufrido la misma transformación. Se propone así un ejemplo donde B_1 sufre una traslación hacia la derecha y lo contrario sucede para B_2 . Es evidente que cierta parte de la imagen para cada banda quedará completamente de color negro, y en todas ellas será diferente dada la posición física de las lentes en el dispositivo. Se debe establecer así un área sobre el modelo multicapa que no será más que la intersección de todas aquellas posiciones de intensidad no nula en las cuatro imágenes multiespectrales.

$$\min_x = \max_x(M_i * [0, 0, 1]^T, M_i * [0, h - 1, 1]^T, \min_x) \quad \forall i \in [0, 3] \quad (8)$$

$$\min_y = \max_y(M_i * [0, 0, 1]^T, M_i * [w - 1, 0, 1]^T, \min_y) \quad \forall i \in [0, 3] \quad (9)$$

$$\max_x = \min_x(M_i * [w - 1, 0, 1]^T, M_i * [w - 1, h - 1, 1]^T, \max_x) \quad \forall i \in [0, 3] \quad (10)$$

$$\max_y = \min_y(M_i * [0, h - 1, 1]^T, M_i * [w - 1, h - 1, 1]^T, \max_y) \quad \forall i \in [0, 3] \quad (11)$$

Nótese como M_i será la matriz de transformación aplicada a la banda B_i , que podría ser una composición de matrices en el caso de que B_i no se alinee de forma inmediata con la imagen de referencia. Por otro lado, la transformación podría incluir una rotación, y por lo tanto, no sería válido comprobar sólo dos esquinas no adyacentes para hallar el área común de las cuatro imágenes.

2.3. Extracción de vegetación

Una vez se encuentran todas las capas representadas en un mismo sistema de referencia se puede extraer información de la imagen con mayor precisión. En este trabajo se propone la extracción de árboles, como individuos, del área capturada por las imágenes multiespectrales.

De forma básica, se podrían extraer los contornos necesarios en una imagen de alto contraste como la banda RED, sin embargo, esta no nos permite distinguir un árbol y su sombra. Tampoco podríamos recurrir a la banda NIR exclusivamente (aunque esta sí diferencia sombra) ya que el nivel de intensidad del suelo y la vegetación es similar. Nótese como el objetivo de este algoritmo en

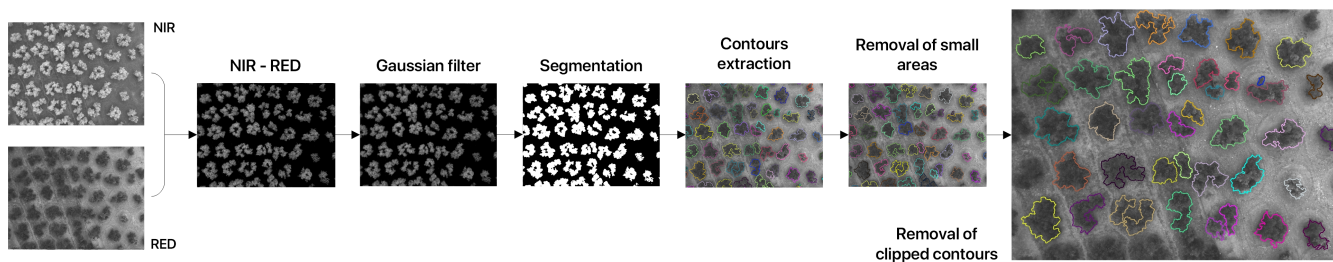


Figure 5: Procesamiento de imágenes multispectrales para la obtención de contornos de árboles presentes en el terreno capturado.

concreto, al igual que los anteriores, es que no necesite supervisión alguna y sea más bien automático. No sería ese el caso si se utiliza sólo esta última banda, dado que se debería manipular el umbral de intensidad a utilizar (dejando a un lado los posibles errores que se puedan producir al detectar los contornos debido precisamente a esa intensidad tan similar).

Dada la función de reflectancia de la vegetación a lo largo del espectro, $f(\lambda)$, se debe tener en cuenta que dicha función alcanza un máximo local en la imagen NIR, y un mínimo local en la imagen RED. En ambas se muestra el color del suelo y el de las sombras con una intensidad similar, mientras que los árboles se representan con una intensidad cercana al blanco y al negro respectivamente. Una simple operación de diferencia de estas bandas nos permite hallar un resultado, que aunque contiene cierto ruido, puede servir como una primera aproximación.

El proceso que se requiere hasta llegar a la imagen final de la Figura 5 implica a su vez dos tipos de tareas: una primera parte de construcción de los contornos, representada en las cuatro primeras operaciones de dicha figura, y una última sección dedicada al post-procesamiento de los resultados de etapas anteriores. Dado que el objetivo último que se persigue es facilitar el almacenamiento de información asociada a los árboles, se deben desechar aquellos contornos de los que sólo se dispone de información parcial debido a que el polígono resultante se encuentra seccionado.

Una última operación contempla la eliminación de aquellos contornos que no pertenecen necesariamente a un árbol sino a pequeños rastros de vegetación en el terreno, para lo cual se pueden adoptar dos puntos de vista diferentes:

- Filtrado de vegetación a partir del área del polígono y un umbral de área. Es la solución que se implementa en este algoritmo. Para proponer un umbral de área se deben tener en cuenta las dimensiones de la imagen.
- Operaciones morfológicas, como la erosión, con la desventaja de que deforman los contornos originales y no siempre eliminan rastros de vegetación de mayor tamaño. Si no se busca un contorno muy riguroso y el núcleo de la operación no es excesivamente grande, podría ser una opción.

3. Conclusiones y trabajo futuro

En este trabajo se ha propuesto un método de registro y ajuste de las imágenes multispectrales desde un punto de vista genérico y

lo más preciso posible, con el fin de que se pueda extraer información de ellas más tarde. Además de la precisión, una característica muy deseable encontrada en las primeras transformaciones de adecuación es que no necesitan supervisión. En el caso de la extracción de contornos, el resultado es aceptable con los parámetros correctos, pero no es automático (se pueden controlar hasta 3 parámetros: tamaño de núcleo de filtro gaussiano, umbral de área y umbral de segmentación, donde los dos primeros sí se pueden configurar como estáticos en la gran mayoría de situaciones).

Por lo tanto, quedarían dos tareas pendientes que extienden el algoritmo aquí propuesto:

- Adaptación del registro de imágenes para aquellos casos en los que el algoritmo no converja, como podría suceder cuando el dispositivo de vuelo captura imágenes a una altura muy baja. También se podría optimizar el algoritmo de registro a través del uso de imágenes de menor tamaño que permitan una convergencia más temprana.
- Automatización del proceso de extracción de contornos, en base a propiedades como la luminancia o la distancia entre los árboles de la plantación que se va a tratar.

References

- [CRDG*15] CANDIAGO S., REMONDINO F., DE GIGLIO M., DUBBINI M., GATTELLI M.: Evaluating multispectral images and vegetation indices for precision farming applications from uav images. *Remote Sensing* 7, 4 (2015), 4026–4047. URL: <http://www.mdpi.com/2072-4292/7/4/4026>, doi:10.3390/rs70404026. 2
- [EP08] EVANGELIDIS G. D., PSARAKIS E. Z.: Parametric image alignment using enhanced correlation coefficient maximization. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 30, 10 (Oct 2008), 1858–1865. doi:10.1109/TPAMI.2008.113. 3
- [JRHC17] JHAN J., RAU J., HAALA N., CRAMER M.: Investigation of parallax issues for multi-lens multispectral camera band co-registration. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives* 42, 2W6 (8 2017), 157–163. doi:10.5194/isprs-archives-XLII-2-W6-157-2017. 2, 3
- [SXZJ14] SHEN X., XU L., ZHANG Q., JIA J.: Multi-modal and multi-spectral registration for natural images. In *Computer Vision – ECCV 2014* (2014), Springer International Publishing, pp. 309–324. 2
- [YES*18] YASIR R., ERAMIAN M., STAVNESS I., SHIRTLIFFE S., DUDDU H.: Data-driven multispectral image registration. In *2018 15th Conference on Computer and Robot Vision (CRV)* (May 2018), pp. 230–237. doi:10.1109/CRV.2018.00040. 3