

Estimación de la cubierta nubosa en imágenes de cielo mediante el algoritmo de clasificación KNN.

A. Cazorla, F.J. Olmo y L. Alados-Arboledas

Grupo de Física de la Atmósfera. Dep. Física Aplicada. Universidad de Granada. Avda. Fuentenueva s/n, 18071 Granada. cazorla@ugr.es, fjolmo@ugr.es.

Resumen

La necesidad de una buena determinación de la cubierta nubosa obtenida a partir de imágenes de cielo está justificada por el hecho de la gran influencia que ejercen las nubes sobre la radiación recibida en superficie, máxime en condiciones de cielo parcialmente cubierto. Tradicionalmente se ha aplicado un umbral a la razón de los canales rojo y azul para determinar la presencia de nubes. Este procedimiento no es adecuado en las proximidades del Sol y cerca del horizonte. En este trabajo estudiamos un algoritmo de clasificación sencillo que puede ayudar a resolver el problema desde el punto de vista de la correcta estimación del número de octas de cielo cubierto.

1. Introducción

Comprender el efecto que ejercen las nubes sobre la radiación recibida en la superficie terrestre es de vital importancia para estudios basados en el balance de energía en superficie [1]. Por ejemplo, citemos la importancia del análisis de los cielos parcialmente cubiertos para la estimación de la radiación UV en superficie. Si bien la observación visual que se realiza habitualmente en las estaciones meteorológicas nos da una información básica para el análisis y el estudio del clima, su resolución es limitada en el tiempo y es sensible a los errores del observador. Los sistemas de cámara de cielo automáticos nos brindan una información de cubierta nubosa con gran resolución espacio temporal. Además, las técnicas de procesamiento de imágenes pueden ayudar en la detección de las propiedades de las nubes [2-3].

Para estimar la fracción de cielo cubierto por nubes, y su distribución espacial, Shield et al. [1] han utilizado un algoritmo de decisión que proporciona buenos resultados para nubes opacas y finas. Este algoritmo se basa en la razón de los canales rojo y azul (lo que llamamos razón R/B) proporcionados por las imágenes. Feister et al. [4] utilizan un algoritmo para la detección de nubes opacas, que también se basa en la razón R/B, con imágenes rojas y azules tomadas muy próximas en

el tiempo. El resultado final es una imagen que nos permite decidir sobre el campo de nubes. En los dos casos el algoritmo se ha implementado para el mismo sistema: el Whole Sky Imager (WSI).

En este trabajo hemos implementado un sistema de clasificación basado en el algoritmo KNN –K Nearest Neighbor–. Compararemos los resultados con los que se obtienen cuando se utiliza la razón R/B.

2. La cámara de cielo

2.1. La cámara digital

El dispositivo experimental utilizado es una cámara digital RETIGA 1300C de QImaging™, cuyas características fundamentales son:

- CCD de 2/3” en color (Mosaico de Bayer).
- 1280x1024 píxeles de 6.7 μm x 6.7 μm . Este tamaño de píxel proporciona una alta razón señal-ruido y un alto rango dinámico.
- Ruido de salida de 8e- y corriente oscura de 0.15e-/pix/s
- Eficiencia cuántica de 30% en 400nm, 43% en 500nm y 30% en 600nm.
- Salida digital de 12 bits. Lo que permite obtener imágenes de 36 bits de color (12 por canal) frente a los tradicionales 24 bits de color (8 por canal). Esto hace que el rango dinámico sea mayor.
- La cámara está refrigerada con una peltier a 25° C. Este sistema minimiza el ruido de la imagen, especialmente en condiciones de poca luz y, por tanto, tiempo de exposición alto.
- Interfaz con el PC mediante FireWire IEEE 1394.

El objetivo de la cámara es de tipo “ojo de pez” con un campo de visión (FOV) de 185°.

2.2. Montaje de la cámara de cielo

La cámara está instalada en una carcasa impermeable con un sistema de calefacción y de refrigeración. En la parte superior tiene una cúpula de cristal para proteger el objetivo. Este montaje permite una operación continua en el exterior.

Es imprescindible que el sistema tape la luz directa del Sol para evitar que el CCD se sature y se dañe. Además, al cubrir el disco solar también incrementamos el contraste en la imagen. Para realizar el sombreado, el sistema (cámara y carcasa) se ha instalado en un seguidor solar que dispone de una bola de sombra (ZAP Tracker de Kipp & Zonen™). Con esta configuración el sistema sigue al Sol y siempre queda sombreado.

2.3. Software de la cámara

La cámara incluye un software propio y una API –Interfaz de Programación de Aplicaciones– en C++ para su uso a medida. Como el software disponible es limitado hemos creado un nuevo software adecuado para nuestras necesidades. Este software controla los parámetros típicos de una cámara digital: tiempo de exposición, ganancia, offset, etc. También permite la autonomía temporal en la captura y archivo de fotos, lo que nos brinda la posibilidad de su adaptación al intervalo temporal que se considere adecuado. También se archiva un registro que contiene los parámetros de la cámara en el momento de la captura de la foto, así como la fecha y hora.

3. Análisis de las imágenes

A continuación describimos los procedimientos seguidos en el análisis de las imágenes de cielo obtenidas con la cámara.

3.1. Umbral sobre la razón R/B

Una forma sencilla de determinar la cubierta nubosa es mediante el empleo de umbrales. Es fácil pensar que las nubes, que son más o menos blancas, tendrán un valor de intensidad más alto que el cielo que es azul. Basándonos en esto, para cada imagen determinamos un umbral en escala de grises. De este modo, en principio, todos los píxeles que superan el umbral corresponden a nubes y el resto a cielo despejado. En imágenes en color disponemos de tres canales, lo que nos permite desarrollar procedimientos más complejos con los canales: sumar, restar, multiplicar o dividir. Como ya anotamos anteriormente la razón R/B ha sido históricamente el método más utilizado. Este método proporciona buenos resultados con

imágenes de zonas de cielo o en situaciones concretas, como puede ser el caso de cielos parcialmente nubosos con el Sol obstruido. Cuando lo aplicamos a imágenes de toda la bóveda celeste se pueden presentar diversos problemas: por un lado, la zona circumsolar es muy brillante y prácticamente siempre se determina como nube, y por otro lado, el horizonte brillante hace que también se sobreestime la cantidad de nubes en esta zona.

En la figura 1 se muestra un ejemplo, obtenido de nuestra base de imágenes, donde se aplica el algoritmo basado en la razón R/B. La figura 1b muestra una representación en 3D de la luminancia de la imagen en escala de grises. La figura 1c representa en 3D la razón R/B y, finalmente, la figura 1d representa la imagen de decisión de nubes para la razón R/B. Como podemos observar en las representaciones 3D, al ser la luminosidad muy alta en la zona circumsolar, el algoritmo la clasifica como zona ocupada por las nubes.

3.2. Clasificación con KNN

Podemos ver el problema de la determinación de la cubierta nubosa desde otro punto de vista. En realidad lo que queremos es clasificar cada píxel de la imagen, determinando así si se trata de cielo o nube. Abordarlo a nivel de píxel no es muy adecuado porque no tenemos información suficiente, pero podemos definir una vecindad y hacer un reconocimiento de patrones a nivel de ventanas de, por ejemplo 4x4 píxeles, lo que nos da no sólo información sobre el color, sino que también nos da una idea de la estructura del campo, lo cual, intuitivamente, es importante para una correcta determinación de la cubierta nubosa.

El algoritmo KNN permite abordar el problema de clasificación teniendo en cuenta que los patrones son regiones de imágenes. El primer paso es convertir esa región en un vector de características. Se han extraído hasta 14 características de los píxeles. A este vector de características lo llamaremos prototipo o patrón. Manualmente podemos clasificar los prototipos y definir una clase para ellos. Hemos considerado dos clases: cielo y nube.

El proceso seguido por el algoritmo KNN es el siguiente:

- Se almacena en una tabla los prototipos clasificados manualmente. A esto lo llamaremos conjunto de referencia (de tamaño N).

- Ante un nuevo patrón a clasificar se calcula su distancia euclídea a los N prototipos del conjunto de referencia y se consideran los K mas cercanos.
- Se contabilizan las clases a las que pertenecen esos K prototipos y el nuevo patrón se clasifica con la clase mayoritaria.

El caso más sencillo es cuando K toma el valor 1, es decir, el nuevo patrón se clasifica como perteneciente a la clase del prototipo más cercano a él en el conjunto de referencia.

El proceso que nosotros hemos seguido tiene las siguientes características:

- Dividimos las imágenes en ventanas de un tamaño adecuado.
- Extraemos las características (hasta 14) para obtener un vector que represente esa porción de imagen.
- Definimos un conjunto de imágenes que clasificamos manualmente para usar sus patrones como conjunto de referencia.
- Ante una nueva imagen obtenemos sus patrones y se los pasamos al clasificador KNN.
- Reconstruimos una imagen de decisión de nube a partir de los resultados del KNN con los patrones de la imagen que muestra los puntos (patrones) que se han considerado nube y los que se han considerado cielo. Esta imagen tendrá valor blanco para los puntos que se consideren nube y gris para los puntos que se consideren cielo. En negro representaremos la sombra del Sol y los bordes.
- Para evaluar, se calcula el porcentaje de cielo cubierto en la imagen de decisión de nube, se calcula el número de octas y se comprueba con la clasificación en octas de cielo cubierto que ha hecho visualmente un experto sobre esas mismas imágenes.

3.3. Resultados

Hemos elegido de nuestra base de datos un conjunto de 113 imágenes con diferentes situaciones de cubierta nubosa.

Se han utilizado cinco imágenes para crear el conjunto de referencia, dos de ellas de cielo completamente cubierto y tres de cielo despejado con diferentes condiciones de luminosidad. Cada imagen genera más de 300000 patrones, lo cual es

inviabile para establecerlo como conjunto de referencia. Para hacerlo viable desde un punto de vista computacional hemos reducido este conjunto hasta tener del orden de 600 patrones para cada clase, lo cual, aún así, sigue siendo un número bastante grande.

Se calcula el porcentaje de nubes de las imágenes usando el umbral de la razón R/B para comparar y se clasifican después con KNN para distintos valores de K (de 1, 3 y 5). La tabla 1 muestra los resultados de la media del error absoluto y la desviación estándar en la predicción del número de octas para todo el conjunto de imágenes y para las diferentes técnicas usadas. La figura 2 muestra dos ejemplos de imágenes originales y sus correspondientes imágenes de decisión de nube para los dos métodos (razón R/B y KNN).

Tabla 1: Comparación de resultados

	Error absoluto	Desv. Std.
1NN	1.15	0.94
3NN	1.19	0.95
5NN	1.19	0.95
R/B	1.35	1.43

4. Conclusiones

Hemos aplicado la técnica de umbrales sobre la razón R/B así como el método KNN con un total de 14 parámetros para clasificar los píxeles de una imagen de la bóveda celeste en condiciones de cielo cubierto o despejado. Nuestros resultados indican que el método KNN produce mejores resultados.

Con la razón R/B perdemos la información de lo que pasa alrededor del Sol y, en muchas ocasiones, de lo que pasa en el horizonte. La estimación en octas tiene un error grande y, lo que es más significativo, una desviación estándar elevada. El clasificador KNN subestima el número de octas, pero siempre mantiene la estructura que hay en el cielo, es decir, que las zonas con nubes serán nube en la imagen de decisión de nubes y las zonas de cielo serán cielo en la imagen de decisión. De todos modos no debemos olvidar que la comparación la estamos realizando con el valor de octas estimado de modo visual por un observador experto que, en ocasiones, no descuenta las pequeñas zonas de cielo despejado que puedan encontrarse entre las zonas de nubes. El error más apreciable lo encontramos en la zona circunsolar, que suele clasificarse como cielo en casi todas las imágenes. De todos modos, la estructura del algoritmo KNN que hemos desarrollado hasta el momento, aunque prometedora, aún no es capaz de

clasificar correctamente cierto tipo de nubes dispersas y zonas de formación.

Estos errores se podrían solventar eligiendo imágenes adecuadas para construir el conjunto de referencia. También podríamos intentar clasificar más número de clases, incluyendo varios tipos de nubes.

La ventaja del KNN radica en que se presenta como un sistema modular muy flexible. El siguiente paso lógico sería tratar de extraer más información de la imagen para obtener más características y, lo más importante, aplicar alguna técnica de "minería de datos" para optimizar el conjunto de características de los patrones. Un número elevado de características (14 es un número bastante grande) hace que los datos estén muy dispersos en el espacio euclídeo y el cálculo de las distancias no es tan discriminante como lo puede ser un espacio euclídeo con menos dimensiones.

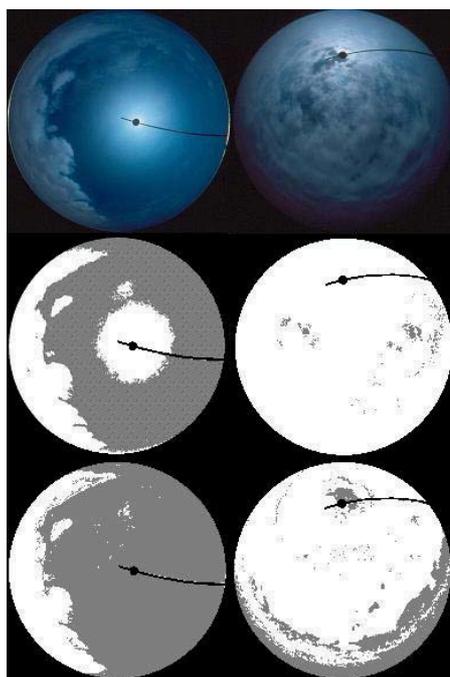


Figura 2: La primera fila muestra dos imágenes originales. La segunda fila muestra las imágenes de decisión de nube aplicando la razón R/B. La tercera fila muestra las imágenes de decisión de nube aplicando el KNN.

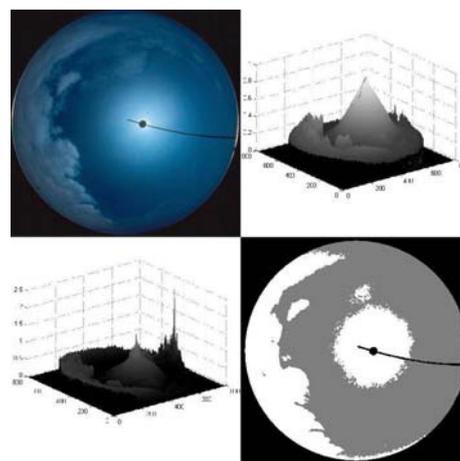


Figura 1: (a) Imagen de muestra. (b) Representación 3D de luminancia. (c) Representación 3D de la razón R/B. (d) Imagen de decisión de nubes con la razón R/B

5. Agradecimientos

Este trabajo ha sido financiado por la CICYT a través de los proyectos REN2003-03175 y CGL2004-05984-C07 03.

6. Referencias

- [1] J.E. Shields, M.E. Karr and T.P. Tooman, *The whole sky imager – A year of progress*, in Proceedings of the Eighth Atmospheric Radiation Measurement (ARM) Science Team Meeting (Tucson, Arizona, 1998), pp. 677-685.
- [2] J. Sabburg and F. Wong, *Evaluation of a Sky Camera for use in radiation measurements*, J. of Atmos. Ocean Tech., 16, 752-759 (1999).
- [3] B. Davis, D.J. Griggs and G.D. Sullivan, *Automatic estimation of clouds amount using computer vision*, J. of Atmos. Ocean Tech., 9, 81-85 (1992).
- [4] U. Feister, J. Shields, M. Karr, R. Johnson, K. Dehne, M. Woldt, *Ground-based cloud images and sky radiances in the visible and near infrared region from whole sky imager measurements*, Proceedings of climate monitoring – satellite application facility training workshop, Dresden 2000.